Kilday, John-Paul (2011) Genomic and epigenetic characterisation of childhood ependymoma. PhD thesis, University of Nottingham.

GENOMIC AND EPIGENETIC CHARACTERISATION OF CHILDHOOD EPENDYMOMA

John-Paul Kilday, MBChB MRCPCH

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

December 2011

This thesis is dedicated to the children and their families whose lives have been affected by brain tumours

ABSTRACT

Paediatric ependymomas remain a clinical management challenge, with a relatively poor prognosis when compared to other childhood tumours of the central nervous system. An improved understanding of underlying ependymoma biology may identify new correlates of outcome and potential therapeutic targets. To address this, Affymetrix^{TM} 500K SNP arrays were used to establish the nature and range of genomic imbalances in 63 paediatric ependymomas (42 primary and 21 recurrent). Over 80 % of tumours were analysed against patient-matched constitutional DNA. In addition, the Illumina$^{®}$ GoldenGate$^{®}$ Cancer Panel I array was used to identify differences in methylation profile across 98 paediatric ependymomas (73 primary and 25 recurrent). While collective assessment revealed the most common anomalies, specific aberrations were characteristic of certain ependymoma subgroups, particularly those relating to tumour location, patient age, disease recurrence and patient prognosis. The genomic imbalance of 15 selected candidate genes (NSL1, DNAJC25, NAV1, CDKN2A, CHI3L1, HOXA5, TXN, BNIPL, and PRUNE) were confirmed by quantitative Polymerase Chain Reaction. Genomic gain involving regions of chromosome 1q were associated with an unfavourable patient outcome, such as the focal locus on 1q21.3 encompassing PRUNE. The genomic gain of PRUNE correlated with an increased encoded protein expression, as assessed by immunohistochemistry (IHC).

This adverse prognostic association with 1q was upheld in the subsequent part of this work. Fluorescent in situ hybridisation and IHC were used to evaluate a panel of six putative prognostic markers (1q25 gain, PRUNE, Tenascin-C, Nucleolin, Ki-67 and NAV1 expression) across a paediatric intracranial ependymoma tissue microarray cohort of 107 primary tumours treated within the confines of two aged defined clinical trials (UK CCLG 1992 04 and SIOP 1999 04). Within the younger UK CCLG 1992 04 cohort, copy number gain of chromosome 1q25 and PRUNE overexpression were independently associated with an increased risk of disease progression, while strong PRUNE expression was also an independent marker of worse overall survival. In addition, increased Tenascin-C expression correlated with a reduced overall survival on univariate analysis. For older children in the SIOP 1999 04 cohort, strong PRUNE expression in ependymomas was again identified as an adverse prognostic marker, correlating with increased mortality on univariate assessment.

ACKNOWLEGDEMENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS (AND NOTE)

aCGH – Array comparative genomic hybridisation

AEIOP – Associazione Italiana Ematologia Oncologia Pediatrica

ALL – Acute lymphoblastic leukaemia

AMP – Adenosine monophosphate

ANOVA – Analysis of variance

ASO – Allele-specific oligonucleotide

ATP – Adenosine triphosphate

aUPD – Acquired uniparental disomy

BAC – Bacterial artificial chromosome

BASH – Beadarray Subversion of Harshlight

bHLH – Basic helix-loop-helix

BMP – Bone morphogenetic proteins

BTP2 – Brain tumour polyposis 2

BRLMM – Bayesian Robust Linear Model with Malanobis distance classifier

C – Cytosines

CGH – Comparative genomic hybridisation

CHTN – Cooperative Human Tissue Network

CI – Confidence interval

CML – Chronic myeloid leukaemia

CNS – Central Nervous System

CNAG – Copy Number Analyser for Genechip

CNV – Copy number variation

CNVRs – Copy number variable regions

CSF – Cerebrospinal fluid

DAPI – 4',6-diamidino-2-phenylindole

DHH family – Phosphoesterases with signature motif Asp-His-His

DNA – Deoxyribonucleic acid

DPX – Depex-polystyrene dissolved in Xylene

EDTA – Ethylenediaminetetraacetic acid

EFS – Event free survival

EGF – Epidermal growth factor

EGFR – Epidermal growth factor receptor

FISH – Fluorescent in situ hybridisation

FFPE – Formalin fixed paraffin embedded

G – Guanine

GCOS – Genechip Operating Software

GISTIC – Genomic Identification of Significant Targets In Cancer

GFAP – Glial Fibrillary Acidic Protein

GTPase – Guanosine triphosphatase

GTYPE – Genechip Genotyping Analysing Software

HDACs – Histone methyltransferases

hTERT – Human telomerase reverse transcriptase

IHC – Immunohistochemistry

LI – Labelling index

LOH – Loss of heterozygosity

LSO – Locus specific oligonucleotide

mRNA – Messenger ribonucleic acid

MBD – Methyl-binding domain

MEN1 – Multiple Endocrine Neoplasia type 1

MM – Mismatch probes

NOD-SCID – Non-obese diabetic, severe combined immunodeficient

NSC – Neural stem cell

OCR – Oligonucleotide Control Reagent

Oligo – Oligonucleotide array

OS – Overall survival

PAC – P1-derived artificial chromosome

PBS – Phosphate buffered saline

PCA – Principal component analysis

PCR – Polymerase chain reaction

PLG – Phase lock gel

PM – Perfect match probes

PNETs – Primitive neuroectodermal tumours

QC – Quality control

qPCR – Quantitative polymerase chain reaction

RFU – Relative fluorescence units

RNA - Ribonucleic acid

RNAi - Ribonucleic acid interference

RTK – Receptor tyrosine kinase

SAM – Sentrix-Array Matrix

SAPE – Streptavidin Phycoerythrin

SDS – Sodium dodecyl sulphate

SFOP – **Société Française** d'Oncologie **Pédiatrique**

SIOP – International Society of Paediatric Oncology

SNP – Single nucleotide polymorphism

SRCC – Spearman's rank correlation coefficient

SSC – Saline-sodium citrate

T – Thymine

TAE – Tris-acetate-EDTA

Tm – Annealing temperature

TMA – Tissue microarray

TS2 – Turcot syndrome type 2

U – Uracils

UK CCLG – United Kingdom Children's Cancer and Leukaemia Group

UPD – Uniparental disomy

UPMGA – Unweighted pair group method with arithmetic mean

VEGF – Vascular endothelial growth factor

WHO – Word Health Organisation

YAC – Yeast artificial chromosome

Note: Throughout the text of this work, gene names are presented in italics whilst protein names are presented in normal font.

CHAPTER 1

INTRODUCTION

1.1 Paediatric brain tumours

1.1.1 Background

Tumours of the brain and spinal cord are collectively the most common solid malignancies in children, accounting for approximately 25 % of all childhood cancers (Stiller 2004). The reported incidence of paediatric central nervous system (CNS) tumours in the United Kingdom has increased by almost a quarter within the last three decades, a presumed consequence of advanced diagnostic and reporting techniques (Stiller 2004). However, despite the development of neuroimaging and therapeutics, improvements in survival for paediatric brain tumour patients have been modest when compared to most other children's cancers, such that they remain the leading cause of cancer deaths in this age group (Wong, Tsang et al. 2006). This implies a need to improve understanding of the underlying biology of children's brain tumours, as has been the case with cancers such as paediatric haematological malignancies. Indeed, over 80 % of children with leukaemia in industrialised countries can now be cured of disease, compared to only approximately 50 – 60 % of those with a CNS tumour (Stiller 2004). Such unfavourable cure rates for paediatric brain tumour patients are compounded by enduring or progressive neuro-cognitive and psychological deficits experienced by a substantial proportion of survivors, acquired as a consequence of the tumour itself or therapeutic interventions.

The histological diversity of paediatric CNS tumours contributes to the difficulty of tumour classification which underpins all current treatment programmes, potentially reflecting biological disparity between tumour subgroups (Stiller 2004). Although several classification systems have been proposed, the most frequently used is the World Health Organization (WHO) 2007 grading system (Louis, Ohgaki et al. 2007) (Table 1.1). It is hoped that an improved knowledge of paediatric brain tumour biology would refine these classification systems in use, encouraging novel prognostic stratifications and allowing more targeted therapeutic strategies.

Table 1.1: WHO grading of tumours of the central nervous system.
(Reproduced from Louis, Ohgaki et al. 2007)

	I	II	III	IV
Astrocytic tumours				
Subependymal giant cell astrocytoma	•			
Pilocytic astrocytoma	•			
Pilomyxoid astrocytoma		•		
Diffuse astrocytoma		•		
Pleomorphic xanthoastrocytoma		•		
Anaplastic astrocytoma			•	
Glioblastoma				•
Giant cell glioblastoma				•
Gliosarcoma				•
Oligodendroglial tumours				
Oligodendroglioma		•		
Anaplastic oligodendroglioma			•	
Oligoastrocytic tumours				
Oligoastrocytoma		•		
Anaplastic oligoastrocytoma			•	
Ependymal tumours				
Subependymoma	•			
Myxopapillary ependymoma	•			
Ependymoma		•		
Anaplastic ependymoma			•	
Choroid plexus tumours				
Choroid plexus papilloma	•			
Atypical choroid plexus papilloma		•		
Choroid plexus carcinoma			•	
Other neuroepithelial tumours				
Angiocentric glioma	•			
Chordoid glioma of the third ventricle		•		
Neuronal and mixed neuronal-glial tumours				
Gangliocytoma	•			
Ganglioglioma	•			
Anaplastic ganglioglioma			•	
Desmoplastic infantile astrocytoma and ganglioglioma	•			
Dysembryoplastic neuroepithelial tumour	•			

	I	II	III	IV
Central neurocytoma		•		
Extraventricular neurocytoma		•		
Cerebellar liponeurocytoma		•		
Paraganglioma of the spinal cord	•			
Papillary glioneuronal tumour	•			
Rosette-forming glioneuronal tumour of the fourth ventricle	•			
Pineal tumours				
Pineocytoma	•			
Pineal parenchymal tumour of intermediate differentiation		•	•	
Pineoblastoma				•
Papillary tumour of the pineal region		•	•	
Embryonal tumours				
Medulloblastoma				•
CNS primitive neuroectodermal tumour (PNET)				•
Atypical teratoid / rhabdoid tumour				•
Tumours of the cranial and paraspinal nerves				
Schwannoma	•			
Neurofibroma	•			
Perineurioma	•	•	•	
Malignant peripheral nerve sheath tumour (MPNST)		•	•	•
Meningeal tumours				
Meningioma	•			
Atypical meningioma		•		
Anaplastic / malignant meningioma			•	
Haemangiopericytoma		•		
Anaplastic haemangiopericytoma			•	
Haemangioblastoma	•			
Tumours of the sellar region				
Craniopharyngioma	•			
Granular cell tumour of the neurohypophysis	•			
Pituicytoma	•			
Spindle cell oncocytoma of the adenohypophysis	•			

1.1.2 Neural development

The term 'embryonal' is often attributed to paediatric brain tumours as they are thought to arise in immature tissue environments as a consequence of deviation from normal neuro-developmental processes (Scotting, Walker et al. 2005).

The central nervous system (CNS), comprising the brain and spinal cord, is a complex three-dimensional structure that originates from the ectodermal neuroepithelium of the embryo. During early CNS development, the ectoderm thickens to form the neural plate during a process termed neural induction. Dorsal fusion of the rolled neural plate in turn creates the neural tube. A subsequent spatio-temporal mediated sequence of proliferation, lateral migration from the neural tube and differentiation then ensues to define CNS sub-regions demonstrating constitutive cellular diversity.

Initially, neuroepithelial cells are thought to proliferate and give rise to radial glial cells in the embryonic ventricular zone (Noctor, Martinez-Cerdeno et al. 2007), such that radial glia are considered the principal neuronal progenitors of the primitive CNS (Mori, Buffo et al. 2005). As development continues after birth, radial glial cells either mature directly into ependymal cells which comprise the ventricular lining of the brain and are implicated in CSF circulation, or transform into postnatal/adult multipotent neural stem cells (NSCs) (Spassky, Merkle et al. 2005). These NSCs express the astroglial marker Glial Fibrillary Acidic Protein (GFAP) and become localised to specialised germinative niches (Vescovi, Galli et al. 2006). Two such regions are the subventricular and subgranular zones of the postnatal and adult brain, where multipotent NSCs have been identified albeit in restricted numbers (Eriksson, Perfilieva et al. 1998; Bonfanti and Peretto 2007). The multipotent NSCs can be the source of further stem cell progeny or give rise to lineage restricted progenitor cells. In the subventricular zone, these latter cells are termed early transit amplifying progenitors which can generate either terminally differentiated glia of the mature CNS (astrocytes and oligodendrocytes) or Type A migrating neuroblasts committed to neuronal differentiation (Doetsch 2003). Similarly, NSCs in the subgranular zone can give rise to mature neurons via the formation of intermediate Type D precursor cells (Doetsch 2003). Since glial cells are the predominant CNS cell type and, unlike neurons, retain a capacity to divide, it is unsurprising that most paediatric brain tumours are glial in origin (Packer 1999).

The processes responsible for determining the timing and location of these stem cell developmental programmes are not fully understood, although they are thought to be modulated by a variety of signalling molecules and pathways that control cell proliferation, migration, maturation and specification. An improved appreciation of

these developmental regulatory mechanisms is imperative as aberrations within these processes may contribute to CNS tumour formation. For instance the Sonic Hedgehog, Wnt and Notch signalling pathways have been shown to be involved in a range of processes essential for embryonic CNS growth and patterning. These include stem/progenitor cell self-renewal, proliferation, migration and determination of cell fate (Morrison 2001; Hitoshi, Alexson et al. 2002; Lai, Kaspar et al. 2003; Machold, Hayashi et al. 2003; Reya, Duncan et al. 2003; Katoh and Katoh 2005; Nyfeler, Kirch et al. 2005). Dysregulation of these pathways have been reported in numerous malignancies including CNS tumours (Taipale and Beachy 2001; Pardal, Clarke et al. 2003; Reya and Clevers 2005). Indeed, the association of these signalling systems with ependymoma pathogenesis is discussed later in this chapter. Similarly, extrinsic growth factors such as epidermal growth factor (EGF) have been have been reported to promote the proliferation of multipotent neural stem cells (Gritti, Frolichsthal-Schoeller et al. 1999). Amplification of the EGF receptor (EGFR) is a frequent finding in adult glioblastoma multiforme (Ohgaki and Kleihues 2007) and has been associated with adverse prognosis in ependymoma (Mendrzyk, Korshunov et al. 2006). Bone morphogenetic proteins (BMPs) are other growth factors which have been reported to influence neural stem cell proliferation and terminal differentiation (Furuta, Piston et al. 1997; Li, Cogswell et al. 1998; Panchision, Pickel et al. 2001), while altered BMP signalling has been identified in intracranial ependymomas (Palm, Figarella-Branger et al. 2009) and high grade gliomas (Vandeputte, Troost et al. 2002).

Notwithstanding external signalling pathways, intrinsic stem cell regulatory mechanisms, such as those facilitating telomeric maintenance, have also been postulated to have a role in cancer when disrupted. Telomeres are nucleoproteins derived from DNA repeat sequences that can contribute to genomic stability by capping chromosomal termini (Ridley, Rahman et al. 2008). Telomere length can be maintained by the ribonucleoprotein telomerase which creates new repeat sequences, a process catalysed by its enzymatic subunit hTERT. Murine studies have shown that telomerase expression in different regions of the developing CNS correlates with the degree of neural progenitor cell proliferation (Prowse and Greider 1995; Ostenfeld, Caldwell et al. 2000; Klapper, Shin et al. 2001), thereby suggesting alterations in telomeric homeostasis play a role in neural development (Rahman, Heath et al. 2009). While telomere erosion occurs with sequential mitotic cycles in normal somatic cells,

telomerase-mediated telomere maintenance is present in almost all types of malignant cells, (Shay and Bacchetti 1997) and overexpression of hTERT has been associated with adverse prognosis in paediatric intracranial ependymomas. This is further discussed later in this thesis.

1.1.3 Brain tumour stem cells

The term 'stem cell' is reserved for those with the ability for self-renewal and multipotency. The identification of cells with these stem-like characteristics facilitating neurogenesis in the adult brain led to the hypothesis that brain tumours were not necessarily restricted to originating from terminally differentiated CNS tissue, but could result from the malignant transformation of these proliferating neural stem cells into cancer inducing stem cells (Vescovi, Galli et al. 2006; Rahman, Heath et al. 2009).

This cancer stem cell hypothesis has underpinned the identification of tumourigenic cell subsets that share the properties of corresponding normal tissue stem cells in several human malignancies, including leukaemia, breast and colon cancer (Lapidot, Sirard et al. 1994; Bonnet and Dick 1997; Al-Hajj, Wicha et al. 2003; O'Brien, Pollett et al. 2007; Ricci-Vitiani, Lombardi et al. 2007). The first putative brain tumour stem cells were identified within a minority population of cells expressing the cell surface antigen CD133, isolated from glioblastoma multiforme and medulloblastoma (Singh, Clarke et al. 2003). These cells possessed a marked capacity for self-renewal (as evidenced by the formation of floating aggregates termed neurospheres), proliferation and differentiation (Singh, Clarke et al. 2003). Moreover, the CD133+ cells initiated tumour formation upon xenograft transplantation in non-obese diabetic, severe combined immunodeficient (NOD-SCID) mouse brain (Singh, Hawkins et al. 2004). In a subsequent study, cells demonstrating a radial glial immunophenotype (CD133+/Nestin+/ RC2+/BLBP+) were isolated from ependymoma and also provided evidence of self-renewal in vivo by forming orthotopic murine tumours (Taylor, Poppleton et al. 2005). In both cases, very few CD133+ cells were required to generate tumour formation. By contrast, cells lacking CD133 expression were not tumourigenic, even when large cell numbers were injected (Singh, Hawkins et al. 2004; Taylor, Poppleton et al. 2005).

This realisation has modified our appreciation of cancer development by suggesting that the majority of cells within paediatric brain tumours are not tumourigenic, with only a minority capable of tumour initiation. Currently, the term 'brain tumour stem cell' defines cells with the ability to proliferate, self renew and have a capacity to produce the different cell lineages comprising the tumour mass (Clarke, Dick et al. 2006). Immunophenotyping is not incorporated in this definition as subsequent studies have shown a tumourigenic capacity for cells lacking CD133 expression in glioblastoma (Wang, Sakariassen et al. 2008).

Alternative work has focussed on establishing the molecular mechanisms triggering malignant stem cell conversion in the CNS. While some studies have considered the influence from the cellular microenvironment (Calabrese, Poppleton et al. 2007), others have proposed a role for signalling pathways shown to be implicated in both normal stem cell regeneration and uncontrolled tumourigenic cell growth, such as the Sonic Hedgehog, Wnt and Notch pathways (Gaiano, Nye et al. 2000; Hitoshi, Alexson et al. 2002; Lai, Kaspar et al. 2003; Machold, Hayashi et al. 2003; Radtke and Raj 2003; Reya, Duncan et al. 2003; Nyfeler, Kirch et al. 2005; Reya and Clevers 2005; Romer and Curran 2005).

The nature of the neoplastic cell of origin for different CNS tumour groups, fundamental to the cancer stem cell hypothesis, is another area of ongoing research. While it has been shown that mutations in early neural progenitor cells are sufficient to promote tumourigenesis (Yang, Ellis et al. 2008; Alcantara Llaguno, Chen et al. 2009; Gibson, Tong et al. 2010; Johnson, Wright et al. 2010), it remains conceivable that a more lineage committed precursor or even a de-differentiated mature cell could acquire stem-like characteristics on malignant transformation.

1.2 Clinical and pathological aspects of paediatric ependymoma

1.2.1 Epidemiology

Ependymomas are the third most common paediatric CNS tumours after astrocytomas and PNETs, with an incidence of approximately 2.2 per million children (Kulkarni

7

2004). They account for 6 – 12 % of brain tumours in children and almost 2 % of all childhood cancers (Pollack 1994; Miller, Young et al. 1995; Bouffet, Perilongo et al. 1998). Although also reported in adults, over half of all cases occur in children under five years of age (Heidemann RL 1997; Duffner, Krischer et al. 1998; Grill, Le Deley et al. 2001). While some studies have shown that ependymomas occur predominantly in males (Goldwein, Leahy et al. 1990; Horn, Heideman et al. 1999), others have observed equal incidence rates between boys and girls (Rousseau, Habrand et al. 1994; Robertson, Zeltzer et al. 1998).

1.2.2 Tumour location

Almost 90 % of paediatric ependymomas are intracranial in origin with two-thirds arising in the posterior fossa, presumably from the ependymal cell surface of the fourth ventricle (Heidemann RL 1997; Grill, Le Deley et al. 2001; Kulkarni 2004). Nevertheless, paediatric ependymomas are capable of occurring anywhere within the CNS. This includes the parenchyma of the cerebral hemispheres away from the ependymal lining and, rarely, the spinal cord (Merchant and Fouladi 2005). This contrasts markedly with ependymoma location in adults, where approximately 75 % of tumours arise in the spine (Ebert, von Haken et al. 1999). Ependymomas also share certain clinical characteristics with germ cell tumours and have been reported outside the CNS in the sacrococcygeal area, mediastinum (Aktug, Hakguder et al. 2000) and ovary (Hirahara, Yamanaka et al. 1997), suggesting variable aberrant cell migration and differentiation pathways during ependymoma tumourigenesis.

1.2.3 Aetiology and predisposing syndromes

The aetiology of ependymoma remains relatively unknown, with cranial exposure to therapeutic doses of ionising radiation being the only established environmental risk factor (Stiller 2004). While simian virus 40 (SV40) DNA sequences have been found in a proportion of tumours (Bergsagel, Finegold et al. 1992), subsequent studies have not supported a causal association between SV40 exposure and ependymoma development (Krainer, Schenk et al. 1995; Weggen, Bayer et al. 2000; Reuther, Lohler et al. 2001). Genetic predisposition has also been implicated as a cause in certain cases. For instance, Neurofibromatosis type 2 is an autosomal dominant condition caused by germline

mutations in the NF2 gene (22q12.2). This disorder is particularly associated with the formation of vestibular schwannomas, although meningiomas and spinal ependymomas are also relatively common findings (Ebert, von Haken et al. 1999; Lamszus, Lachenmayer et al. 2001; Evans 2009). Spinal ependymomas have also been reported in adults with Multiple Endocrine Neoplasia type 1 (MEN1) syndrome, characterised by numerous endocrine tumours (Kato, Uchimura et al. 1996), albeit infrequently. Intracranial ependymomas have been described as a rare feature of Turcot syndrome type 2, otherwise known as brain tumour polyposis 2 (TS2/BTP2) (Onilude, Lusher et al. 2006). This familial cancer syndrome is caused by inherited mutations of the APC gene (5q22.2) and is characterised by the development of intestinal polyps, yet is more commonly associated with medulloblastoma formation (Hamilton, Liu et al. 1995). Indeed, somatic mutations of APC have not been identified in ependymoma (Onilude, Lusher et al. 2006). While familial ependymomas have been described in the absence of such familial syndromes (Dimopoulos, Fountas et al. 2006), these reports are uncommon and the majority of paediatric tumours appear to arise without a known inherited predisposition.

1.2.4 Histology and ependymoma classification

Ependymomas are neuroepithelial tumours of variable morphological appearance. The histology of this tumour group remains an extremely controversial correlate of prognosis. The WHO 2007 grading system identifies four major types of ependymoma: subependymoma and myxopapillary ependymoma (grade 1), classic or low grade (grade II) and anaplastic (grade III) (Louis, Ohgaki et al. 2007) (Figure 1.1 and Table 1.2). The latter two grades are the principal variants found in children. Classic ependymomas often demonstrate dense cellularity and perivascular pseudorosettes (Goldwein, Leahy et al. 1990) while a diagnosis of anaplasia requires the presence of necrosis, calcification, increased microvascular proliferation, mitotic activity and cell density (Bouffet, Perilongo et al. 1998).

Table 1.2: Current WHO classifications of ependymomas

(Reproduced from Louis, Ohgaki et al. 2007).

WHO Grading System	Subgroups	Histopathology
I	Subependymoma	Isomorphic nuclei embedded in a dense fibrillary matrix of glial cell processes with frequent microcystic change. Mitoses rare or absent.
	Myxopapillary ependymoma	GFAP-expressing cuboidal to elongated tumour cells arranged in a papillary manner around vascular stromal cores.
II (Classic)	Cellular ependymoma Papillary ependymoma Clear cell ependymoma Tanycytic ependymoma	Monomorphic nuclear morphology. Mitoses rare or absent. Perivascular pseudorosettes and ependymal rosettes.
III (Anaplastic)		High mitotic activity. Palisading necrosis. Microvascular proliferation. Perivascular pseudorosettes.

Although specific grades have been recognized, the reality of assigning such classifications to an ependymoma remains subjective (Kulkarni 2004). Benign, classic and anaplastic morphology probably represent different points along a continuous pathological spectrum, while tumour heterogeneity may result in islands of anaplasia occurring within classic histological regions (Figure 1.1C), causing uncertainty regarding the degree of focal anaplasia that represents true anaplasia. This is highlighted by the proportion of anaplastic ependymomas in reported series ranging from $7 - 89$ % (Bouffet, Perilongo et al. 1998) and a study of 34 ependymomas where tumour grade was revised in almost a quarter of cases (Horn, Heideman et al. 1999). Furthermore, the current criteria defining anaplasia has met with criticism from some neuropathologists who do not believe endothelial proliferation or necrosis reflect this tumour grade (Schiffer, Chio et al. 1991).
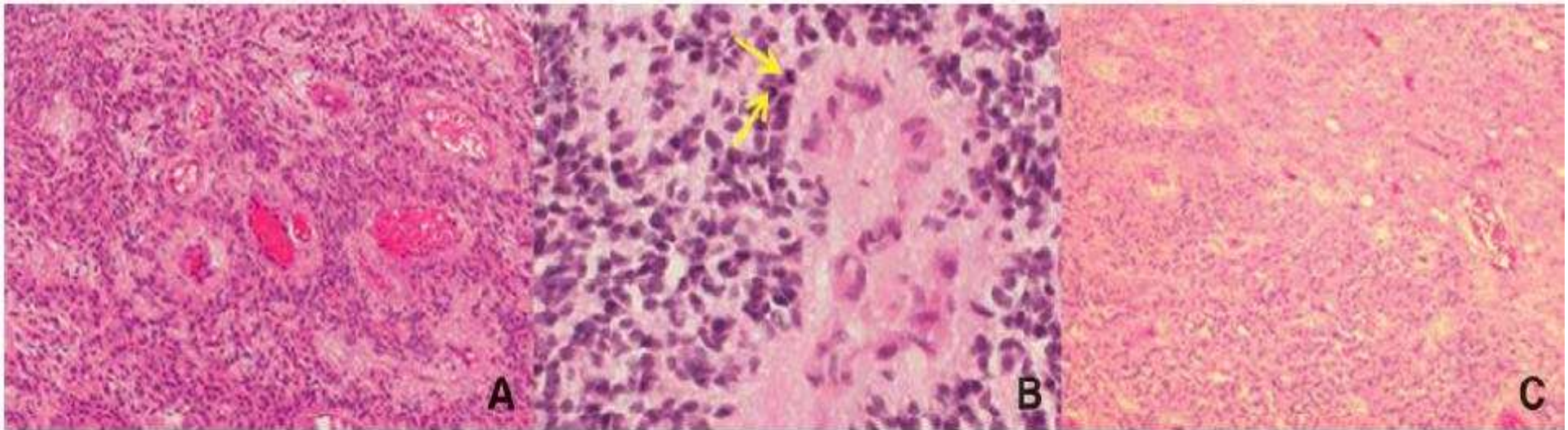
Figure 1.1: Histological appearances of paediatric ependymoma. A. Classical (grade II) ependymoma with perivascular pseudorosettes. B. Anaplastic (grade III) ependymoma with high mitotic activity (arrows) and vascular proliferation. C. Ependymoma displaying tumour heterogeneity with a grade II region (top right) and a grade III region (bottom left) in the same section. Objectives ×10, ×40 and ×10 respectively.

The inherent ambiguity present in the current histomorphological classification of ependymoma undoubtedly contributes to the contradictory results of studies investigating a prognostic role for tumour grading. Several studies have linked the morphological features of anaplasia to a poor outcome or suggested ependymoma grade affects survival post radiotherapy (Horn, Heideman et al. 1999; Figarella-Branger, Civatte et al. 2000; Grill, Le Deley et al. 2001; Oya, Shibamoto et al. 2002; Paulino, Wen et al. 2002; Korshunov, Golanov et al. 2004; Massimino, Gandola et al. 2004; Tihan, Zhou et al. 2008; Merchant, Li et al. 2009). However other studies refuted this or fail to establish an association between anaplasia and prognosis, including those adopting a WHO classification system (Goldwein, Leahy et al. 1990; Sutton, Goldwein et al. 1990; Pollack, Gerszten et al. 1995; Perilongo, Massimino et al. 1997; Duffner, Krischer et al. 1998; Robertson, Zeltzer et al. 1998). Moreover, studies analyzing the prognostic role of specific histological features such as necrosis, mitotic activity, cellularity, pleomorphism and vascular proliferation, have yielded conflicting results (Hamilton and Pollack 1997; Rickert and Paulus 2005).

Recent attempts to reduce the prognostic ambiguity of ependymoma histology have included the meta-analysis of increased numbers of ependymomas using co-operative neuropathology reviews to reduce subjective interpretation bias. Tihan and colleagues found that WHO tumour grading was an independent prognostic factor for event-free survival, but not overall survival, in 96 paediatric posterior fossa ependymomas where histology had been reviewed by three neuropathologists (Tihan, Zhou et al. 2008). In contrast, a panel of five neuropathologists devised a novel histological grading scheme following consensus review of 233 intracranial ependymomas from children enrolled in four European paediatric clinical trials (Ellison, Kocak et al. 2011). The trials were designed for younger patients receiving chemotherapy as the primary adjuvant therapy (BBSFOP (Grill, Le Deley et al. 2001) and UK CCLG 1992 04 (Grundy, Wilne et al. 2007)) or patients aged above three years treated primarily with post-operative radiotherapy (AEIOP ependymoma (Massimino, Gandola et al. 2004) and SIOP 1999 04). Using the new scheme, tumours were divided into two grades – II and III. Cell density, nodularity, mitotic activity and angiogenesis were considered important histological criteria for re-classification, whereas necrosis and cytological atypia were not (Figure 1.2).
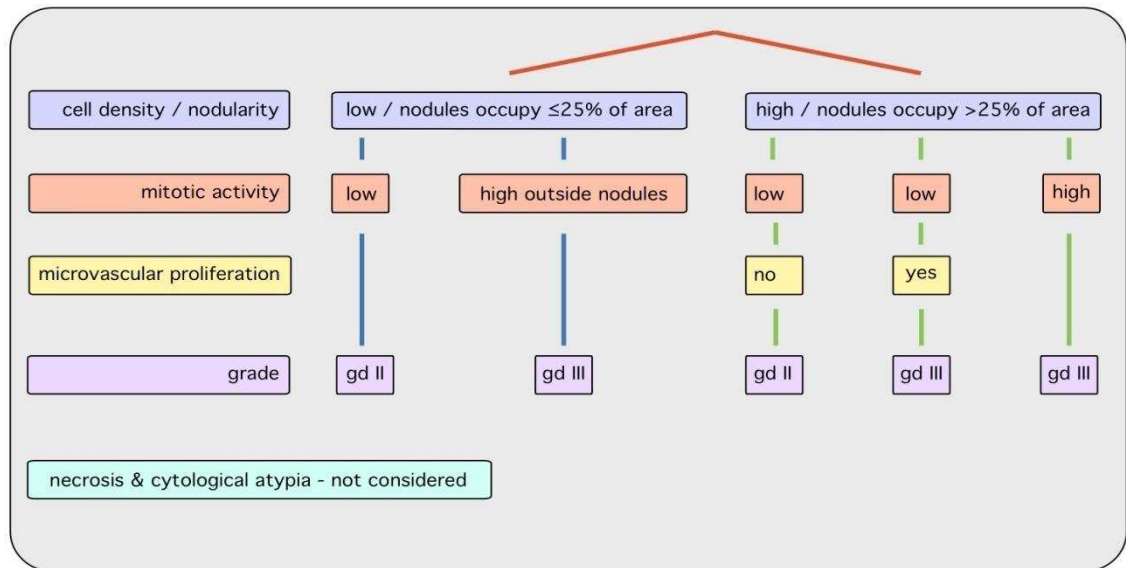
Figure 1.2: The Ellison classification scheme for paediatric ependymoma histology (Ellison, Kocak et al. 2011).

The group concluded that concordance on grading among the five pathologists improved significantly after devising this new classification scheme. Consensus regarding grade and the histopathological variables analysed was significantly associated with patient survival, but only in older children from the AEIOP clinical trial. Therefore while this scheme does not appear appropriate for ependymomas in younger children, it may be for ependymomas from older children. An explanation for the discrepant outcome results between children in the AIEOP and SIOP 1999 04 trials was not given, although there was a significant difference between the proportions of tumours with a gross total resection in the two studies.

These recent studies suggest that undertaking large scale histological reviews should be encouraged through national and international collaborations. Until consensus is reached on uniform classification criteria, potentially for different age categories within the paediatric population, the use of tumour grade as a prognostic marker in paediatric ependymoma will remain contentious.

1.2.5 Treatment of paediatric intracranial ependymoma

Historically, the success of treatment for paediatric brain tumours was measured according to survival alone. However, oncology therapy has evolved to adopt a multi-modal, holistic approach focussing more on the quality of such survival. Treatment

modalities for paediatric ependymoma include surgery and the adjuvant therapies of chemotherapy and radiotherapy.

Surgery remains the primary therapy for paediatric ependymoma, with complete tumour excision being the ultimate aim. However, the rate of complete resection achieved in this tumour group can vary from 31 – 85 %, reflecting diverse surgical techniques, inconsistent criteria for defining complete tumour excision and variable surgical accessibility of ependymomas from different CNS locations (Rousseau, Habrand et al. 1994; Bouffet, Perilongo et al. 1998; Merchant and Fouladi 2005). Historical analysis of the natural history of ependymoma in children has demonstrated that the post-operative prognosis for children with intracranial ependymomas remained poor with surgery alone but improved with the addition of radiotherapy (Mork and Loken 1977). In turn, this has led to the general consensus that surgery should be accompanied by adjuvant therapy.

The choice of adjuvant therapy used is influenced by the philosophies of different national institutions and groups. While radiotherapy is effective, concerns remain about its potential long-term toxicity to the immature central nervous system (Grundy, Wilne et al. 2007). Consequently, most European centres generally reserve local radiotherapy to the site of the primary tumour for children older than three years, following complete surgical resection or as a prelude to second-look surgery (Bouffet, Perilongo et al. 1998; Merchant and Fouladi 2005). Furthermore, European groups have focussed on adopting chemotherapy regimens for young children, with the aim of avoiding or delaying radiotherapy (Grundy, Wilne et al. 2007). Results from the United Kingdom's CCLG CNS 1992 04 clinical trial has suggested that such a deferral strategy could benefit up to 42 % of patients aged younger than three years (Grundy, Wilne et al. 2007), although results from other groups attempting to delay or avoid radiotherapy in this age range have been less encouraging. The French Society of Paediatric Oncology group, using post-operative chemotherapy alone reported a four year progression-free survival of only 22 % (Grill, Le Deley et al. 2001), while a five year progression free survival rate of 27 % was obtained by the Pediatric Oncology Group using either 12 or 24 months of chemotherapy prior to radiotherapy (Duffner, Krischer et al. 1998).

Chemotherapy is also used in several current European studies for older children following incomplete surgical excision to aid the removal of residual disease, either alone or in conjunction with further surgery. However, definitive conclusions on the benefit of chemotherapy in paediatric ependymoma remain elusive since several phase II trials of single agent chemotherapy in ependymoma have proved disappointing, while a paucity of radiological data on tumour response to chemotherapy and follow-up data on the long-term neuro-cognitive effects of prolonged chemotherapy exists (Duffner, Krischer et al. 1998; Bouffet, Tabori et al. 2007).

In contrast to European treatment strategies, most North American institutions support postoperative radiotherapy as the standard of care for localised intracranial ependymomas in children aged above 12 months, in favour of the deferral regimens that use chemotherapy. Radiotherapeutic advances have improved the precision of CNS irradiation, such that focal radiotherapy is now advocated, using doses of 59.4 Gy to the primary tumour site (Merchant and Fouladi 2005; Merchant, Li et al. 2009).

The reasoning for this fundamental therapeutic difference is underpinned by data from a study at St. Jude Children's Research Hospital, USA, in which 153 children with localised intracranial ependymoma were treated with this regime (Merchant, Li et al. 2009). At a median follow up period of 5.3 years, the event-free survival and overall survival were 69.1 % and 81 % respectively (Merchant, Li et al. 2009). Although these results are encouraging, approximately one-third of all treatment failures were in children below three years of age. In addition, almost nine percent of children treated with adjuvant radiotherapy alone developed distant relapse, which was substantially higher than that seen using the United Kingdom's CCLG CNS 1992 04 chemotherapy-based clinical trial (Grundy, Wilne et al. 2007). Adverse events attributed to radiotherapy were observed for patients in the St. Jude's study, including hearing impairment (Hua, Bass et al. 2008), protracted attention deficiency (Kiehna, Mulhern et al. 2006) and a decline in reading ability (Conklin, Li et al. 2008), while the follow-up period was unable to account for longer term neuro-cognitive impairment and the delayed effects from radiotherapy, such as neuro-endocrine sequelae and secondary malignancies (Grundy, Wilne et al. 2007).

Moreover, other studies exclusively using post-surgical adjuvant radiotherapy for paediatric ependymomas have failed to replicate such promising outcome results, with five year event-free and overall survival rates ranging from 41 – 58 % and 54 – 73 % respectively, although this may reflect advances made in current neurosurgical and radiotherapeutic approaches (Pollack, Gerszten et al. 1995; Perilongo, Massimino et al. 1997; Robertson, Zeltzer et al. 1998; Akyuz, Emir et al. 2000; Oya, Shibamoto et al. 2002; van Veelen-Vincent, Pierre-Kahn et al. 2002; Jaing, Wang et al. 2004; Mansur, Perry et al. 2005; Shu, Sall et al. 2007). A forthcoming study by the Children's Oncology Group of over 300 children with intracranial ependymoma (ACNS0121) will provide more information on the safety and effectiveness of conformal radiotherapy in young patients.

1.2.6 Survival and clinicopathological prognostic markers

The prognosis for paediatric ependymomas remains relatively poor when compared to other brain tumours in children, despite advances in neurosurgery, neuroimaging techniques and post-operative adjuvant therapy. The five year survival rate ranges from 24 – 75 %, with a five year progression free survival rate of 23 – 60 % (Messahel, Ashley et al. 2009; Wright and Gajjar 2009; Zacharoulis and Moreno 2009). In addition, late relapses up to fifteen years after initial treatment are not uncommon (Paulino, Wen et al. 2002). The most common site for tumour recurrence is at the site of the primary tumour with metastatic relapse occurring in approximately 20 – 25 % of cases (Messahel, Ashley et al. 2009; Wright and Gajjar 2009). The ability to predict patient outcome has been hampered by the heterogeneous clinical behaviour of ependymomas in children, insufficient recruitment into large prospective clinical trials and contradictory studies regarding the prognostic role of histology, as discussed previously, and other clinical markers (Hamilton and Pollack 1997; Tabori, Ma et al. 2006).

Ependymoma is considered by some to be a 'surgical' disease. However, whilst complete resection is the most consistently reported favourable clinical prognostic factor (reviewed by (Bouffet, Perilongo et al. 1998)); (Duffner, Krischer et al. 1998; Robertson, Zeltzer et al. 1998; Horn, Heideman et al. 1999; Grill, Le Deley et al. 2001; Merchant, Li et al. 2009), this is not a universal finding and some studies fail to

demonstrate this relationship for paediatric posterior fossa tumours (Goldwein, Leahy et al. 1990; Akyuz, Emir et al. 2000). Furthermore, despite complete excision, local tumour recurrence can develop in up to 50 % of cases, even following adjuvant radiotherapy (Hamilton and Pollack 1997; Tabori, Ma et al. 2006).

Patient age at diagnosis and tumour location have also been suggested as prognostic factors. Historically, children below three years of age and infratentorial ependymomas have been associated with a poor outcome (Sutton, Goldwein et al. 1990; Heidemann RL 1997; Perilongo, Massimino et al. 1997; Sala, Talacchi et al. 1998). It remains unclear whether this reflects tumour biology, the surgical inaccessibility of posterior fossa tumours which are more prevalent in younger children, or the avoidance of adjuvant radiotherapy in early life resulting from concerns regarding long term clinical sequelae (Rousseau, Habrand et al. 1994; Hamilton and Pollack 1997; Duffner, Krischer et al. 1998; Sala, Talacchi et al. 1998; Grundy, Wilne et al. 2007).

Poor outcome and the unpredictable behaviour of this tumour in children have turned attention to improving our knowledge of ependymoma biology.

1.3 Genetics of cancer

Cancer is a genetic disease of somatic cells characterised by disruption of cell cycle control and associated with increased cell proliferation, disturbed differentiation and evasion of apoptosis (Tamarin 2002). The two principal categories of genes implicated in carcinogenesis are oncogenes and tumour suppressor genes.

1.3.1 Oncogenes and tumour suppressor genes

Oncogenes are the activated form of normal cellular genes, termed proto-oncogenes, which contribute to the natural regulation of cell growth, survival and differentiation (Pelengaris 2006). Oncogenes are dominant at a cellular level and the majority can be classified according to function as growth factors, receptors, signal transducers or transcription factors (King 2000). Mechanisms facilitating oncogenic activation include translocation of a proto-oncogene to a region of increased transcription activity,

genomic amplification resulting in an increased gene copy number and activating gene mutations, the most frequent being missense point mutations (Tamarin 2002). In addition, epigenetic hypomethylation may increase oncogene expression without accompanying mutation, while oncoprotein overexpression can occur in the absence of oncogenic anomalies, for instance secondary to dysregulated upstream signalling (Pelengaris 2006).

Tumour suppressor genes exert an inhibitory effect on cell cycle regulation or promote apoptosis in order to protect cells from becoming neoplastic (Tamarin 2002). The functional inactivation of these genes by genetic mutation, deletion or epigenetic phenomena can result in tumourigenesis through an imbalance in these protective cellular mechanisms (Roussel 2006).

The concept of tumour suppressor genes was introduced by Knudson's proposed 'two hit' hypothesis following his assessment of patients with retinoblastoma, a paediatric ophthalmic tumour which can be inherited or sporadic (Knudson 1971). Knudson observed that familial cases were often bilateral and presented in very young patients. He subsequently hypothesised that familial retinoblastoma arose from the germline inheritance of a mutated allele of a gene, later defined as the Retinoblastoma (Rb) gene, followed by another somatic mutation or 'hit' of the remaining allele. The younger age at presentation was explained by DNA instability resulting from the first inherited mutation increasing the susceptibility to a second mutation. In contrast, the relatively delayed presentation of sporadic retinoblastoma was explained by both functioning copies of the functioning wild type Rb allele being present at birth. Inactivating them required two independent somatic mutations, the probability of which is low. Nevertheless, in either case, both copies of the gene have to be inactivated to initiate tumourigenesis.

While most tumour suppressor genes are thought to be recessive at a cellular level, thereby requiring such homozygous mutations or genomic deletions to initiate carcinogenesis, this is not universally the case. Haploinsufficient tumour suppressor genes have been recognised, including the cyclin dependent kinase inhibitors CDKN1B (12p13.1) and CDKN2C (1p32.3), where loss of only a single allele is sufficient to promote neoplastic susceptibility (Fero, Randel et al. 1998). Similarly, dominant

negative mutations in tumour suppressor genes such as TP53 (17p13.1) can generate an antagonistic protein that inhibits the function of the wild type product (Baker, Preisinger et al. 1990). Moreover, epigenetic gene silencing may render tumour suppressor genes inactive without disrupting their DNA sequence (Roussel 2006), as discussed later.

In reality, most cancers are complex phenomena, requiring simultaneous alterations in the activity of numerous oncogenes and tumour suppressor genes for initiation (King 2000). While the significance of these genes to carcinogenesis is well established, other potential contributory genomic factors are emerging. For instance, the functional importance of copy number variation within regions of the human genome has been highlighted as important in the study of human susceptibility to diseases (Redon, Ishikawa et al. 2006). Copy number variable regions (CNVRs) span approximately 12 % of the human genome, although often occurring at non-coding sites (Redon, Ishikawa et al. 2006). Nevertheless, hundreds of genes have been identified within CNVRs (Redon, Ishikawa et al. 2006), while recent reports suggest that CNVRs and encompassed genes may contribute to cancer development (Diskin, Hou et al. 2009).

1.3.2 Loss of heterozygosity

Loss of heterozygosity (LOH), as defined by Knudson's two hit hypothesis, represents the functional loss of one parental allele of a gene in which the residual allele harbours a recessive mutation, thereby producing a mutant phenotype. Four principal mechanisms of LOH are observed in cancer; mitotic recombination, gene conversion, gene deletion and whole chromosomal loss (Tamarin 2002).

Mitotic recombination and gene conversion (Figure 1.3) can result in acquired uniparental disomy (UPD), where LOH has occurred without loss of genomic material and both alleles of a gene are derived from one parent. This is also called copy neutral LOH and leads to the generation of daughter cells with reciprocal chromosomal products (Fitzgibbon, Smith et al. 2005). In contrast, other causes of acquired UPD exist which result in altered quantities of genomic material in the cell progeny, such as chromosomal duplication and non-disjunction (Tamarin 2002). Acquired UPD has already been identified in paediatric malignancies such as neuroblastoma and leukaemia

(Fitzgibbon, Smith et al. 2005; George, Attiyeh et al. 2007; Raghavan, Smith et al. 2008). Moreover, the homozygosity resulting from acquired UPD can identify recessive mutational gene targets in cancer. For instance, acquired UPD of FLT3 (13q12) has been implicated in the adverse prognosis of patients with leukaemia and lymphoma (Fitzgibbon, Iqbal et al. 2007; Gupta, Raghavan et al. 2008).
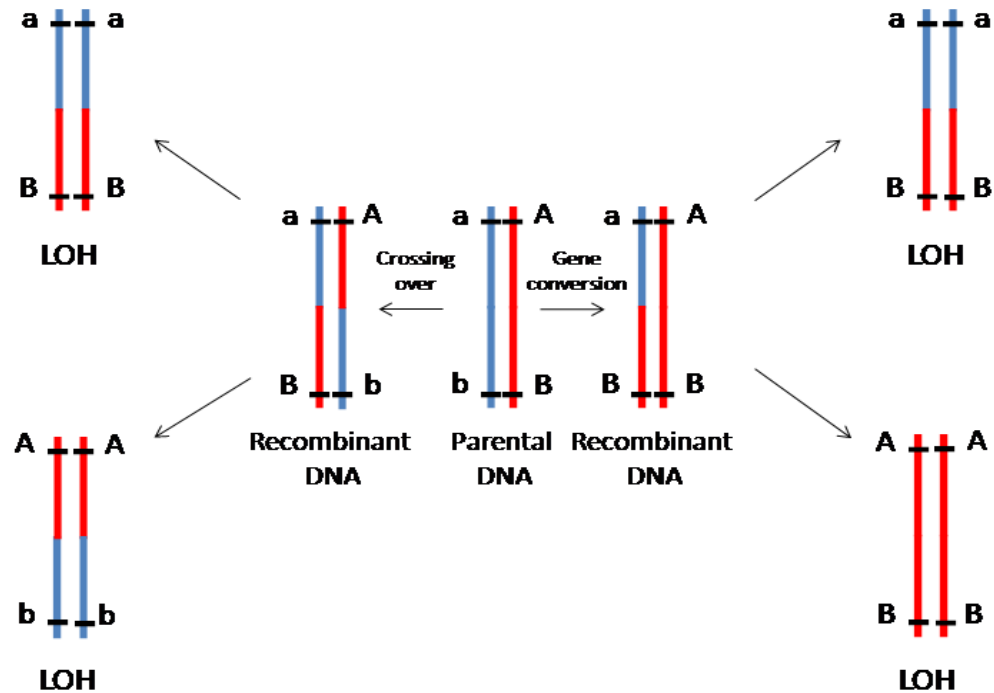
Figure 1.3: Loss of heterozygosity resulting from either mitotic recombination or gene conversion. Mitotic recombination occurs during cell division and involves genomic exchange between two sister chromatids of homologous chromosomes. Following a recombinational cross-over event and mitosis, LOH can result (Vrieling 2001). Gene conversion is a non-reciprocal DNA transfer between two strands of DNA at recombination, causing altered sequence information in the recipient strand. LOH can result from conversion of one gene allele to another (Zhang, Lindroos et al. 2006).

1.3.3 Epigenetics

In contrast to genetic regulation of gene function, epigenetic events can modify gene expression without alteration of the underlying DNA sequence (Roussel 2006). Two principal forms occurring in human cancer are DNA methylation and histone deacetylation, both exerting their effects through transcriptional regulation (Esteller 2008).

DNA methyltransferases are enzymes that can induce methylation of cytosine residues within CpG islands. This can directly prevent transcription factors binding to gene promoter regions, within which the majority of CpG islands cluster, and consequently impede gene expression. Alternatively, methylated DNA can attract methyl-binding

domain (MBD) proteins. These proteins interact with further enzymes, including histone deacetylases (HDACs), histone methyltransferases and other chromatin remodelling proteins in order to chemically modify histones such that chromatin is altered into a transcriptionally respressive state (Bird and Wolffe 1999).

Hypermethylation of gene promoters can silence tumour suppressor genes, thereby facilitating tumourigenesis (Esteller 2002; Herman and Baylin 2003). Examples of this include the hypermethylation of BRCA1 (17q21.31) in breast and ovarian cancer (Esteller, Silva et al. 2000), MLH1 (3p22.2) in gastric and colon cancers (Herman, Umar et al. 1998; Fleisher, Esteller et al. 1999) and CDKN2A (9p21.3) in a variety of cancers (Esteller, Corn et al. 2001). Indeed, the methylation status of several selected genes has been assessed in paediatric brain tumours including medulloblastomas (reviewed by (de Bont, Packer et al. 2008)) and ependymomas (detailed in section 1.5.4), identifying the frequent hypermethylation of candidate genes such as RASSF1A (3p21.3), CASP8 (2q33.1), CDKN2A and HIC1 (17p13.3) across both tumour groups. Global analysis of the medulloblastoma epigenome has also been performed, identifying tumour-specific hypermethylation of genes such as COL1A2 (7q21.3), S100A6 (1q21.3), HTATIP2 (11p15.1), CDH1 (16q22.1) and LXN (3q25.32) (Anderton, Lindsey et al. 2008). Hypomethylation is also thought to play an important role in cancer (Ehrlich 2002). It may enable the increased expression of putative proto-oncogenes genes that would be quiescent in normal cells due to inherent promoter methylation (Tao, Yang et al. 2000; Ehrlich 2002). Similarly, histone deacetylation has been implicated in oncogenesis for a variety of malignancies, including paediatric brain tumours (Rahman, Osteso-Ibanez et al. 2010).

Methylation also regulates the process of genomic imprinting where specific parental alleles are differentially expressed in the germline and maintained in the somatic cells of the offspring. Relaxation of such imprinting can encourage aberrant gene expression and contribute to carcinogenesis, as first reported in Wilms' tumours. In normal tissue, the growth factor IGF2 (11p15.5) is imprinted at the maternal allele which is silenced by methylation, with gene expression provided exclusively by the paternal allele. However in certain Wilms' tumours, loss of imprinting leads to the biallelic transcription and subsequent overexpression of IGF2, thereby conferring a cellular growth advantage (Brown, Power et al. 2008).

Since epigenetic modification appears to have an important role in human cancer, novel therapeutic strategies are being designed to counteract their mechanistic properties, including agents that reverse DNA methylation or inhibit histone deacetylation (Belinsky, Klinge et al. 2003; Winter-Vann, Baron et al. 2005; Egler, Korur et al. 2008). These are being combined with new technologies to rapidly screen the cancer genome for DNA methylation and histone acetylation patterns.

1.4 Increasing resolution of genomic profiling

Since the inception of cytogenetic analysis, advances in technology have refined coverage of the human genome in a single experiment, through the sequential development of comparative genomic hybridisation (CGH), array CGH and single nucleotide polymorphism (SNP) arrays (Table 1.3).

Table 1.3: Comparison of the resolution of different techniques used in genomic analysis

Technique	Platform Type	Number of Probes per Array	Genomic Resolution
Karyotypic analysis	Light microscopy		5 – 10Mb
CGH	Fluorescence microscopy		1 – 10Mb
Array CGH (BAC/PAC or Oligo)	Microarray	6000-240,000	1Mb – 100Kb
100K SNP array	Microarray	100,000	Mean inter-probe distance 23Kb
500K SNP array	Microarray	500,000	Mean inter-probe distance 5.8 Kb
SNP 6.0 array	Microarray	1.8 million	Mean inter-probe distance < 700 bases

1.4.1 Comparative genomic hybridisation

CGH was designed to identify regions of genomic gain and loss in DNA samples of interest (Kallioniemi, Kallioniemi et al. 1992). The procedure involves the simultaneous competitive hybridization of test and control DNA, each labelled with different fluorescent markers, to normal 'target' metaphase chromosomes which are visualised using the DNA counterstain DAPI (Figure 1.4). Test DNA samples are

labelled green, while control DNA samples are labelled red (Inazawa, Inoue et al. 2004). For tumour analysis, control DNA is extracted from normal comparative tissue or the constitutional samples of healthy subjects. Cot-1 DNA is included to prevent the hybridisation of repetitive genomic sequences. Differences in fluorescent intensities between the two markers along the length of each chromosome subsequently identify regions of genetic loss (red) and gain (green) within the tumour genome. Regions with no copy number difference between the test and the standard DNA appear yellow. Loci implicated in neoplastic transformation can thereby be established.

The advantages of CGH is that it allows whole-genome tumour analysis at a higher resolution than karyotypic studies (1 − 10Mb), using DNA isolated from fresh frozen or formalin fixed paraffin embedded (FFPE) archival material. However, relatively large amounts of tumour DNA are required and the improved resolution is still limited for detecting smaller copy number anomalies (Inazawa, Inoue et al. 2004). CGH is also unable to detect genomic aberrations that do not alter the net amount of chromosomal material, such as balanced translocations and copy number neutral LOH (Carvalho, Ouwerkerk et al. 2004; Fitzgibbon, Smith et al. 2005).
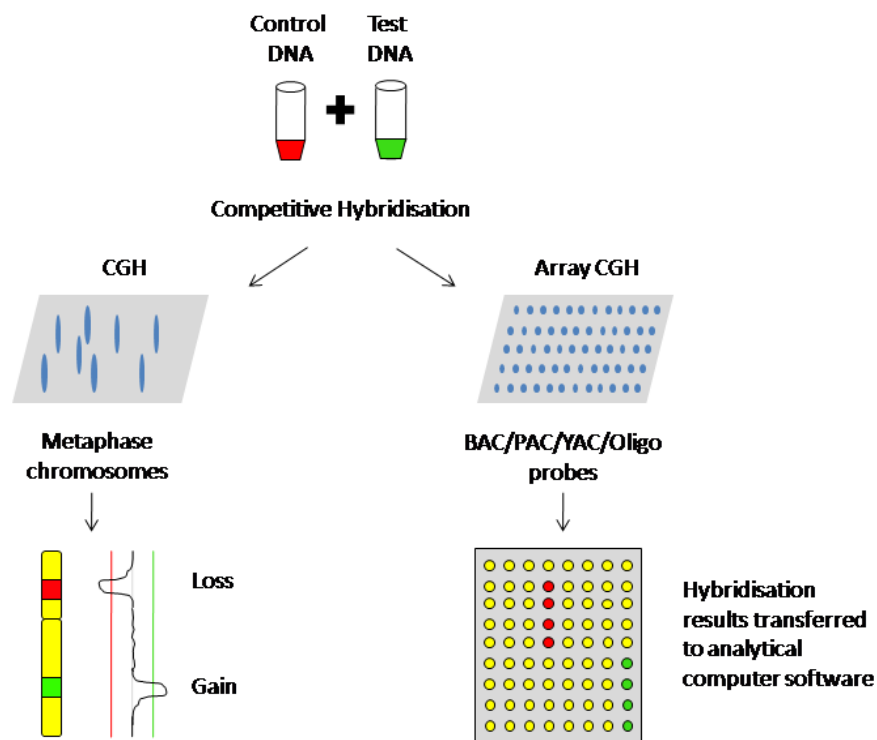
Figure 1.4: Graphical summary comparing conventional CGH with array CGH. CGH = comparative genomic hybridisation. BAC = Bacterial artificial chromosome, PAC = P1-derived artificial chromosome, YAC = yeast artificial chromosome, Oligo = oligonucleotide array.

1.4.2 Array comparative genomic hybridisation

Array CGH (aCGH) uses a succession of genomic clone probes, incorporated onto a glass slide, that span the genome with higher resolution than conventional CGH. Recent advances have enabled the probes to consist of short DNA sequences called oligonucleotides (oligo arrays) (Carvalho, Ouwerkerk et al. 2004). Alternatively, the probes can be derived from bacterial artificial chromosomes (BACs), P1-derived artificial chromosomes (PACs) and Yeast artificial chromosomes that can contain an insert between 100 – 300bp in size (BAC/PAC arrays) (Cowell, Matsui et al. 2004). In array CGH, the probes are the hybridization targets for the sample DNA. Both test and control DNA undergo fluorescent labeling, followed by competitive hybridization to the probes on the array and Cot-1 DNA addition. The ratio of red to green fluorescence represents differential copy number between the two samples (Warr, Ward et al. 2001) (Figure 1.4).

The resolution of aCGH is determined by probe size, inter-probe distance and the number of probes per given slide (Johnson, Hamoudi et al. 2006). While BAC arrays contain approximately 6000 clone probes and only have a genomic resolution of up to 1Mb, oligonucleotide (oligo) arrays are preferential as they can contain up to 240,000 probes, resulting in a resolution as high as 100kb. Consequently, oligo aCGH can detect single copy number gains and losses, making it more informative than other aCGH methods (Carvalho, Ouwerkerk et al. 2004; Cowell, Matsui et al. 2004). Moreover, the process of oligonucleotide probe design is also faster and less expensive than BAC clone production (Carvalho, Ouwerkerk et al. 2004).

As with conventional CGH, array CGH can be performed on FFPE tissue although DNA integrity can influence the quality of results generated (Johnson, Hamoudi et al. 2006). Despite an improved genomic resolution, array CGH also remains unable to detect genomic alterations in regions without probe coverage, or events occurring without a net chromosomal imbalance. Furthermore, artifacts in the copy number data generated can be introduced by cross-hybridisation of low copy number repeat sequences present in both the clone probes and the test DNA sample, not blocked by the addition of Cot-1DNA (Fellermann, Stange et al. 2006).

1.4.3 Single Nucleotide Polymorphism arrays

Single nucleotide polymorphisms (SNPs) are single base pair changes that occur in a DNA sequence and are the most common genetic variations between human beings. Indeed, the human genome project has enabled researchers to identify SNP locations throughout the genome and an estimated 10 million SNP loci are thought to exist (Dong, Wang et al. 2001; Lander, Linton et al. 2001).

Commercial biotechnology companies, such as Affymetrix®, have subsequently created SNP arrays in order to genotype a designated number of SNPs simultaneously in one experiment (Affymetrix 2006). Each Affymetrix® array contains several million 'features' (Figure 1.5A), consisting of more than a million copies of an oligonucleotide probe (Affymetrix 2006). Each probe is 25 base pairs in length and is of a defined sequence, harbouring a known SNP of interest. Each SNP is interrogated by 6- or 10-probe quartets where each probe quartet contains a pair of perfect match (PM) probes and a pair of mismatch (MM) probes used to identify non-specific hybridization for each allele (Figure 1.5B). Therefore, in total, there are 24 to 40 different oligonucleotide probes per SNP (Affymetrix 2006).
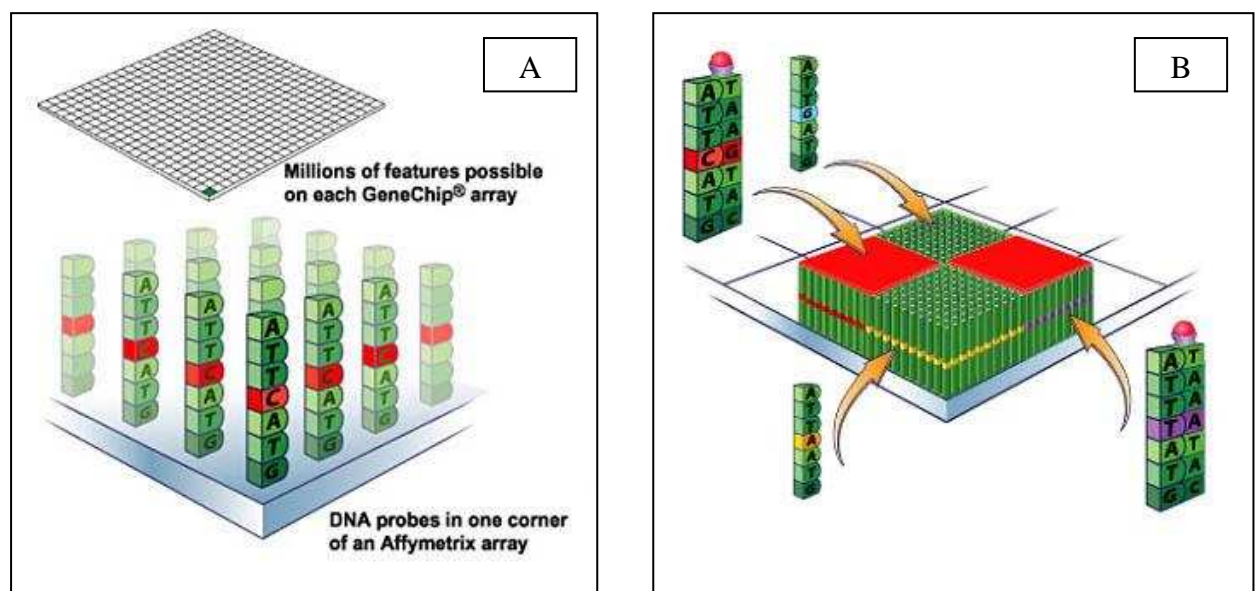
Figure 1.5: Visualisation of oligonucleotide probes on an Affymetrix® SNP array. A. Demonstration of an Affymetrix® SNP array 'feature'. B. Demonstration of PM and MM probes. Both figures obtained from the Genechip® Human Mapping 500K Set User Guide (Affymetrix 2006).

DNA from the sample of interest is digested, ligated, amplified and labelled with a fluorescent marker, before hybridisation to the oligonucleotide probes which produces signal intensity for each SNP. Following hybridisation, the array is washed to remove unbound DNA fragments, then stained and scanned for subsequent computerised data analysis.

When examining tumours for genomic aberrations using the SNP array, a control dataset for comparison must be included. Differential signal intensities between tumour and control DNA samples for consecutive SNPs within a region of interest can thereby lead to the identification of areas of tumour-specific copy number change and loss of heterozygosity. Patient-matched blood DNA samples are often used, as a means of differentiating tumour-specific events from the patient's own constitutional SNP variation. If unavailable, tumour signal intensity data can instead be normalised against constitutional samples derived from a population based pool. Since diversity in SNP variation has been reported between different ethnicities (Iafrate, Feuk et al. 2004; Sebat, Lakshmi et al. 2004; Redon, Ishikawa et al. 2006), this population should be selected to most accurately represent the patient's inherent genomic variation.

With improvements in technology, the 'density' of SNPs on these arrays has increased resulting in a progressively extensive coverage of the human genome (Table 1.3), such that SNP arrays are now a standard method for high resolution genomic analysis in cancer research. While facilitating the identification of smaller regions of genomic imbalance than would be feasible with array CGH technology, the ability to identify single base pair changes also enables genotyping SNP arrays to detect copy number neutral events such as LOH secondary to gene conversion, mitotic recombination and acquired uniparental disomy (section 1.3.2) (Zhao, Li et al. 2004; Paulsson, Cazier et al. 2008).

In addition to the improved genomic coverage conferred by the SNP array, only relatively small amounts of DNA (250 nanograms per chip) are required to process a given sample, an important consideration since brain tumour samples obtained at surgery are often extremely small. However, only DNA from freshly frozen tumours has an acceptable purity for SNP analysis, unlike CGH which can utilize DNA from paraffin embedded tissue, albeit with a reduced accuracy. The increased resolution of

the SNP array also generates vast amounts of data which takes time to process and analyse. Moreover, when using any of the higher resolution platforms for tumour genomic analysis, it is assumed that the control DNA sample has a universally diploid genomic profile. However, as stated previously, constitutional copy number variation exists among different populations (Iafrate, Feuk et al. 2004; Sebat, Lakshmi et al. 2004; Redon, Ishikawa et al. 2006) and may influence the results obtained.

1.5 Biological aspects of paediatric ependymoma

1.5.1 Cytogenetic studies – ependymoma karyotypes

A review of data from 21 karyotypic studies of 65 primary paediatric intracranial ependymomas revealed abnormal karyotypes in 57 (87.7 %) of cases, which are presented in Table 1.4. The abnormalities demonstrated a spectrum of complexity ranging from single rearrangements to structural and numerical aberrations. Whilst abnormalities involving chromosomes 22 and 1q were present in almost 30 % and 20 % of cases respectively, no cytogenetic abnormality was characteristic of paediatric ependymoma and the frequency of specific aberrations appeared similar between children and adults (Hamilton and Pollack 1997). This suggested that analysis of the ependymoma genome at a higher resolution was required to identify age-specific anomalies.

Table 1.4: Abnormal karyotypes reported in the literature of 57 paediatric primary intracranial ependymomas.

Sex	Age (yrs)	Site	Hist	Karyotype[a]	Study
F	2	BS		47,XX,+17	Agamanolis et al., 1995
M	0-18	PF		50,XY,+7,+8,+9,+9	Bhattacharjee et al., 1997
F	8			47,XX,+i(1)(q10)	Bigner et al., 1997
F	8			45,XX,-9/46,idem,+r	
F	3			45,X,-X,t(10;11;15)(p12;q13;p12)	Dal Cin et al., 1998
F	6			46,X,-Y,+7,-10,-14,+18,+20,-22,+mar	Debiec-Rychter et al., 2003
M	4			45,XY,der(1)t(1;3)(q24;p13),der(6)t(2;6)(p11;q22)	
M	1			45,XY,-22	
M	8			44,X,-Y,-22	
M	2	PF		46,XY,+del(1)(p22),-16	Griffin et al., 1988
F	16	PF		55,XX,+2,+3,+5,+7,+8,+9,+11,+19,+21	Griffin et al., 1992

Sex	Age (yrs)	Site	Hist	Karyotype[a]	Study
M	9			46,XY,+2mar	Jenkins et al.,1989
M	4			46,XY,inv(11)(?p11-13?q13-14)	
M	5	PF	CL	46,fra(X)(q27.3)c LT	Kramer et al., 1998
M	1	ST	CL	46,XY,?der(6)	
F	8	PF	CL	46,XX,der(3)t(1;3)(q23;q27-29)	
F	6	PF	AN	46,XX,trp(1)(q22q31),der(6)t(1;6)trp(1)	
M	3	PF	CL	46,XY,der(1;6)(q10;p10)	
F	2	ST	CL	46,XX,?der(6)	
M	2			46,X,-Y,der(1)t(Y;1)(q12;q11)	Mazewski et al., 1999
F	3			49,XX,add(1)(p36),+i(1)(q10)x2,add(2)(p25), inv(2)(p25q21),?t(3;3)(q29;q25), -7,del(7)(q32), -9,- 11,+add(14)(p11),+3mar/54,idem,+inv(2), +del(7),+add(14), add(16)(q23),+17,+21,+3mar	
F	4			46,XX,der(6)t(1;6)(q11;q11)/46,XX,der(14)t(1;14) (q11;p11)	
F	11			45,X,-X,add(19)(p13),add(22)(q?13)	
F	0			46,XX,r(14)(p11q?23),add(16)(q11)/46,XX,der(14) t(?1;14)(?q11;p11),add(16)	
F	7			47,XX,+i(1)(q10)	
F	6	PF		45,XX,-22	Neumann et al., 1993
M	15	PF		46,XY,add(20)(p?)	
M	3	PF		45,XY,der(6)t(6;16)(q11;p11),-16	
F	12	PF		45,X,t(X;18)(p11;q11),t(1;20)(q21;q13),t(2;17) (p11;p11),add(5)(p?),t(12;18) (p11;q11),-13,13, t(13;14)(q11;p11),del(14)(q?),add(17),der(21) t(17;21) (p11;q1?)	
M	10	PF		50,XY,+8,+9,+15,+19	
M	11	PF		93,XXYY,-6,+13,+20	
M	8	PF		48,XY,+1,t(1;2)(p33;q21),t(11;18)(q13;q21), +mar/48,X,-Y,+der(1)t(1;8) (p31;q22),t(1;8), del(6)(q15),add(16)(p?),+2mar	
M	0-18	PF		46,Y,t(X;22)(p22;q11)	Roberts et al., 2001
M	0-18	PF		46,XY,add(1)(p12)	
M	3	PF		39-51,XY,del(2)(q?34),t(2;4)(q34;q35), +del(6)(q25),+12,-17	Rogatto et al., 1993
F	3	ST		46,XX[5]/40-51, XX,del(X)(p21), der(X)del(X)(p21)del(X)(q26),-X,del(2)(q34), t(2;10)(p25;p12),del(4)(q21q25),del(4)(p14),+6, +10,+16,-17,+22,-22[cp25]	
M	1			46,XY,t(11;17)(q13;q21)	Sainati et al., 1996
M	1			50-77,XXY,del(1)(p22),dup(1)(p13p32)x2, +i(7)(q10),add(9)(p?),t(12;21),inc	
M	3	BS		47,XY,+11	Sawyer et al., 1994
F	4			61-62,XXX,-1,-2,-5,-8,+9,-10,-12,-14,-21,-22 /61-62,idem,tas(3;11) (q29;q25) /61-62,idem,tas(4;22)(p16;p13) /61-62,idem,tas(6;11)(q27;q25) /61-62, idem,tas(9;11)(q34;q25) /61-62,idem,tas(9;17)(q34;q25)	
M	6			46,XY,add(17)(p13)	Stratton et al.,1989
M	3	PF		45,XY,add(9)(q34),-17,add(17)(p13),add(22)(q11)	

Sex	Age (yrs)	Site	Hist	Karyotype[a]	Study
F	5	PF		47,X,-X,+del(1)(q21),t(1;3)(p34;q21), +add(7)(p11),+15,del(16)(q13),inv(17) (p11q21), -19/46,X,-X,t(1;3),del(4)(q25q31),add(9)(q34), del(10)(p13),der (13)t(13;17)(p12;q21),-15,-17, del(18)(p11),+mar/??,X?,t(1;2)(q21;q35),inc	Stratton et al.,1989
F	1			46,XX,add(11)(q13)	Thangavelu et al.,1998
M	13			46,XY,del(5)(q34),+7,-8/45,XY,del(5),-17/44,XY, del(5),-17,-22	Thiel et al.,1992
F	9			48-49,XX,+16,+16,+20,-22,+mar	
F	3			40-44,X,-X,-8,t(11;12)(q13;q24),-15,-17, -22/40-44,idem,der(6;17) (p10;q10)/40-44, idem,dic(6;17)(q13;p13)	Urioste et al., 2002
F	2			46,XX,t(2;22)(p12;q13)/45,idem,-10	Vagner-Capadano et al., 1992
M	6			43-46,X,-Y,+7,+19,-22/44-46,X,-Y,+mar	
F	8			46,X,-X,+7,+16,-22	
F	8			46,X,-X,+mar	
M	9			42-47,XY,+mar	
M	6			46-48,XY,t(1;7)(q25;q35),+2,del(3)(q13),der(5), del(6)(q12),+7,+11,-12 /45,XY,-22,46-48,XY, t(1;7)(q25;q35),+2,del(3)(q13),der(5),del(6)(q12), +7,+11,-12	
M	2			45,XY,-22/54,XY,+5,+7,+11,+13,+14,+15,+19,+21 54,XY,+5,+7,+11,+13,+14,+15,+19,+21	
M	2	PF		46,XY[16]	Wernicke et al., 1995
M	13	PF		46,XY[12]/46,XY,del(5)(q34),+7, -8[5]/45, XY, del(5)(q34),-17[7]/44,XY,del(5)(q34),-17,-22[8]	
F	18	ST		46,XY,der(5)t(5;?)(p;?),der(8)t(8;17)(q24;q23), del(17)(q23)[8]/46,XY,der(5)t(5;?)(p;?),-8, der(8)t(8;17)(q24;q23),-11,del(17)(q23), del(22)(q13)[4]	
M	6			47,XY,+8,+13,-22	Weremowicz et al., 1992

Adapted from (Dyer 2007). [a] = Normal cell lines not shown; M = male; F = female; Hist = histology; PF = posterior fossa; ST = supratentorial; CL = classic; AN = anaplastic.

1.5.2 Comparative genomic hybridisation meta-analysis

Over the past decade conventional CGH has been collectively performed on 303 primary and 71 recurrent ependymomas from children and adults in 13 studies (Table 1.5). A meta-analysis of resulting data was undertaken in order to obtain a more global perspective of the ependymoma genome (Kilday, Rahman et al. 2009). The analysis revealed distinct genomic profiles for paediatric and adult ependymomas and location specific genomic anomalies in paediatric ependymomas.

Table 1.5: Summary of studies analysing ependymomas by conventional CGH

Study	Primary paediatric	Recurrent paediatric	Primary adult	Recurrent adult
Pezzolo et al., 2008	20	-	-	-
Rickert et al., 2006	10	-	3	-
Carter et al., 2002	17	11	44	14
Dyer et al., 2002	42	11	-	-
Grill et al., 2002	13	3	-	-
Jeuken et al., 2002	5	-	15	-
Scheil et al., 2001	7	5	10	4
Ward et al., 2001	28	12	-	-
Hirose et al., 2001	11	7	23	3
Granzow et al., 2001	1	1	-	-
Zheng et al., 2000	7	-	21	-
Shlomit et al., 2000	3	-	-	-
Reardon et al., 1999	23	-	-	-
Total	**187**	**50**	**116**	**21**

For each study, the number of samples analysed are shown and categorised according to whether the patient was paediatric (aged equal to or below 16 years) or adult (above 16 years) and whether the tumour was primary or recurrent.

1.5.2.1 Paediatric versus adult ependymoma

Even accounting for differences in clinical factors and therapy, paediatric ependymomas have a worse prognosis than those in adults (Hamilton and Pollack 1997; Ritter, Hess et al. 1998; Korshunov, Golanov et al. 2002; Louis, Ohgaki et al. 2007). One interpretation of this is that paediatric and adult ependymomas are biologically distinct and evidence from the CGH meta-analysis supported this (Kilday, Rahman et al. 2009). Distinct patterns of imbalance across the genome were identified between primary paediatric and adult ependymomas (Figures 1.6 and 1.7). Paediatric ependymomas most frequently demonstrated gain of chromosomes 1q, 7 and 9 and loss of chromosomes 22, 3, 9p,13q, 6q, 1p, 17 and 6, whereas the commonest genomic aberrations in adult ependymomas were gain of chromosomes 7, 9, 12, 5, 18, X and 2 and loss of 22/22q, 10, 13q, 6 and 14q.

One striking difference between the two age groups revealed by the meta-analysis was genomic gain of 1q which was present in over 20 % of paediatric ependymomas but in only 8 % of their adult counterparts (p < 0.004, Fisher's exact test). In contrast, gains of chromosomes 7, 9 and 12 appeared more prevalent in adults (p < 0.001, Fisher's exact test). The genomic imbalances seen in ependymomas from different regions of the CNS also reflected this distinct pattern (Figures 1.8A – F), suggesting an inter-relationship between patient age and tumour location (Kilday, Rahman et al. 2009). Paediatric

ependymomas tend to arise in an intracranial location where 1q gain predominates, while adult ependymomas have a predilection for a spinal location, characterised by whole chromosome aberrations. Gain of 1q, together with loss of chromosomes 6q and 22, were also a common finding in recurrent paediatric intracranial ependymal tumours (Figure 1.9A – C) (Kilday, Rahman et al. 2009) which implies that genes on these chromosomes may be involved in a more aggressive phenotype.
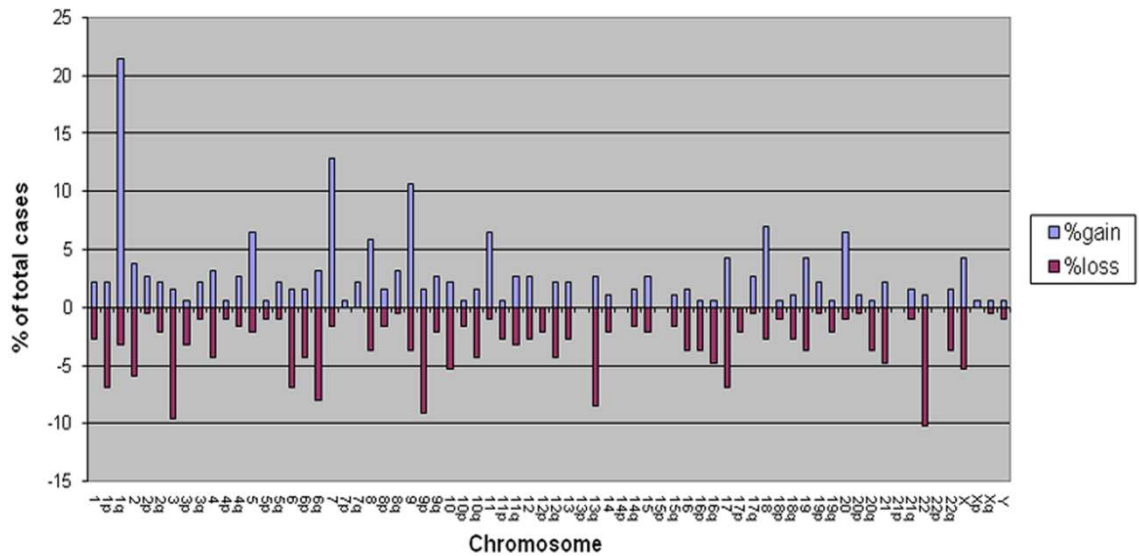
Figure 1.6: Genomic anomalies in 187 primary paediatric ependymomas detected by CGH analysis. The most frequent genomic gains involve chromosome 1q (40/187), 7 (24/187), and 9 (20/187). The most frequent genomic losses involve chromosome 22 (19/187), 3 (18/187), 9p (17/187), 13q (16/187), 6q (15/187), 1p (13/187), 17 (13/187), and 6 (13/187).
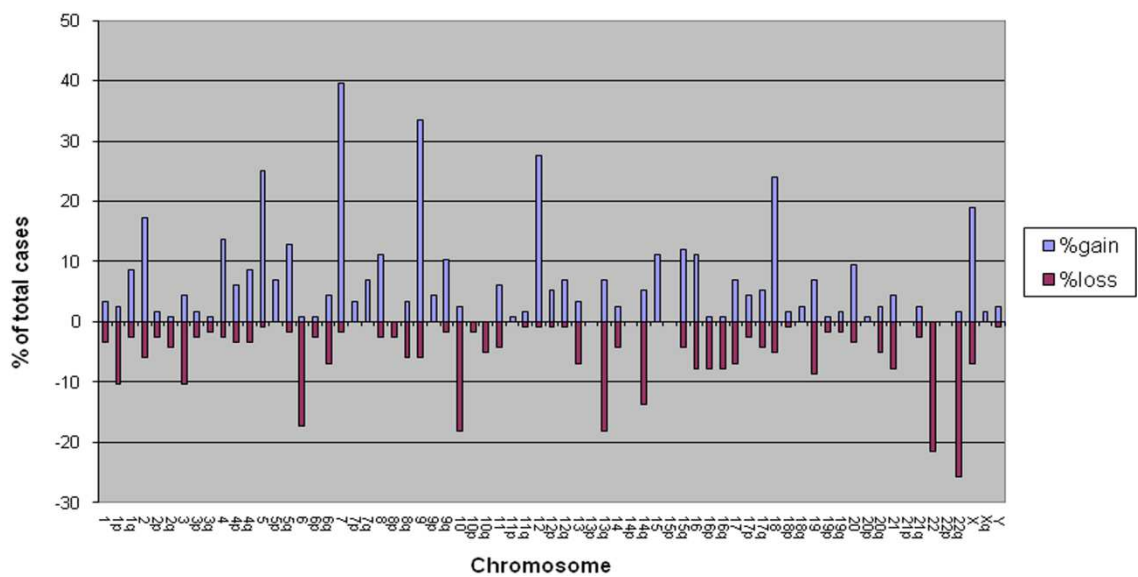
Figure 1.7: Genomic anomalies in 116 primary adult ependymomas detected by CGH analysis. The most frequent genomic gains involve chromosome 7 (46/116), 9 (39/116), 12 (32/116), 5 (29/116), 18 (28/116), X (22/116), and 2 (20/116). The most frequent genomic losses involve chromosome 22q (29/116), 22 (25/116), 10 (21/116), 13q (21/116) 6 (20/116), and 14q (16/116).
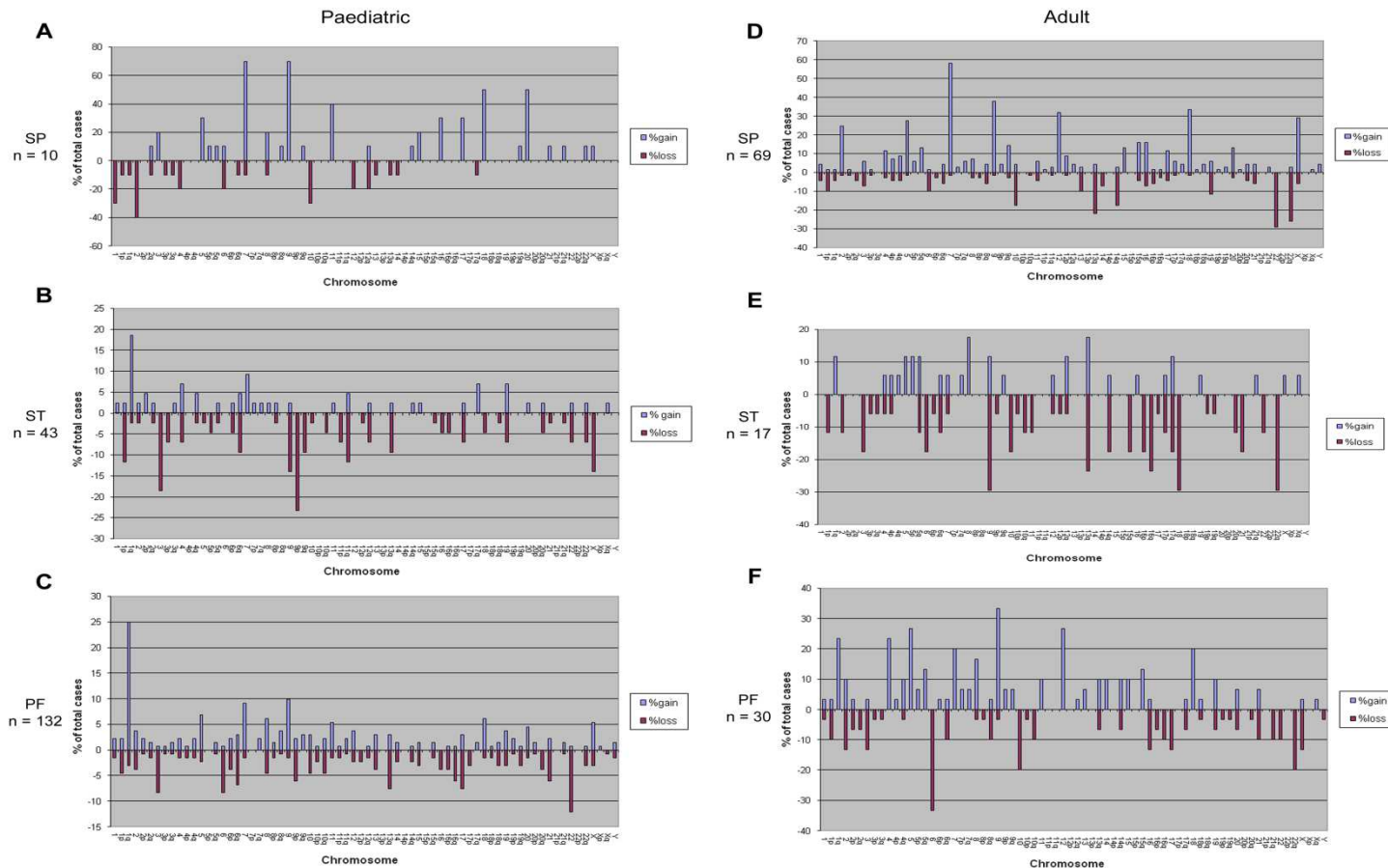
Figure 1.8: Genomic anomalies in 301 primary ependymomas, grouped by tumour location within the CNS and age, as detected by CGH. The paediatric group (A–C) consisted of 10 spinal (SP), 43 supratentorial (ST) and 132 posterior fossa (PF) tumours. The adult group (D–F) consisted of 69 spinal, 17 supratentorial and 30 posterior fossa tumours. The most frequent anomalies in paediatric spinal ependymomas (A) included gains of chromosomes 7 (7/10), 9 (7/10), 18 (5/10), and 20 (5/10) and losses of chromosomes 2 (4/10), 1 (3/10), and 10 (3/10). Frequent imbalances in paediatric supratentorial tumours (B) were gain of chromosome 1q (8/43) and loss of chromosome 9p (10/43), 3 (8/43), and X (6/43). The most common anomalies in paediatric posterior fossa tumours (C) were gain of chromosome 1q (33/132), yet loss of chromosomes 22 (16/132), 3 (11/132), 6 (11/132), 13q (10/132), and 17(10/132). The most frequent anomalies in adult spinal ependymomas (D) included gains of chromosomes 7 (40/69), 9 (20/69), 18 (23/69), 12 (22/69), X (20/69), 5 (19/69), and 2 (17/69) and loss of chromosome 22 (20/69), 22q (18/69), 13q (15/69), 14q (12/69), 10 (12/69), and 19 (8/69). Frequent imbalances in adult supratentorial tumours (E) were gain of chromosome 13q (3/17) and 8 (3/17) and loss of chromosome 9 (5/17), 18 (5/17), 22q (5/17), 13q (4/17), and 16q (4/17). The most common anomalies in adult posterior fossa tumours (F) were gain of chromosomes 9 (10/30), 5 (8/30), 12 (8/30), 1q (7/30), 4 (7/30), 7 (6/30), and 18 (6/30) yet losses of chromosome 6 (10/30), 22q (6/30), and 10 (6/30).

32
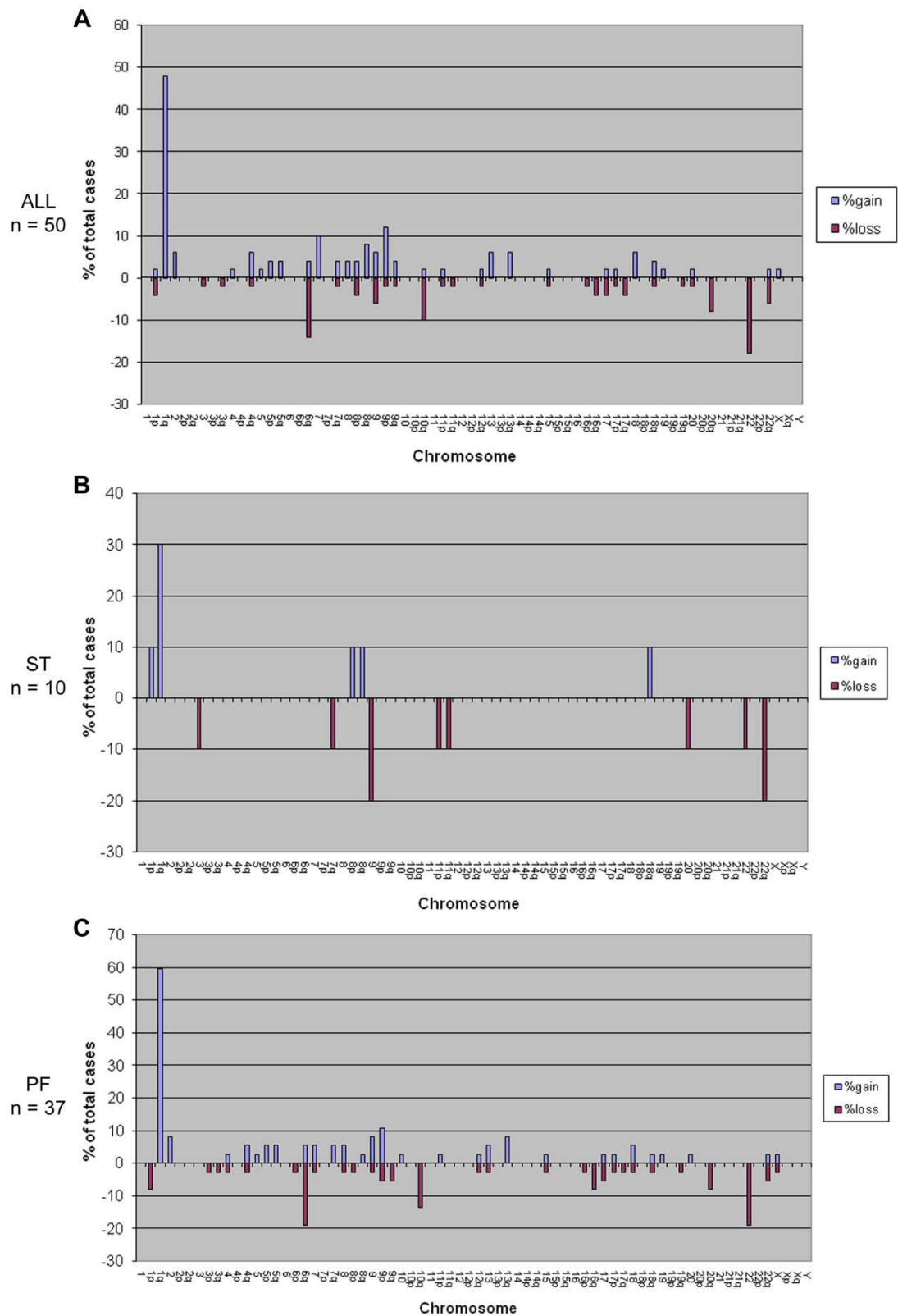
Figure 1.9: Genomic anomalies in 50 intracranial recurrent paediatric ependymomas detected by CGH analysis. Analysis was conducted collectively (A) then subgrouped according to supratentorial location (B: 10 tumours), or posterior fossa location (C: 37 tumours). The location of three tumours was not specified in the literature. Overall (A), the most frequent genomic gain involved 1q (24/50), and the most frequent genomic losses involved chromosomes 22 (9/50) and 6q (7/50). In addition, location-specific anomalies included the loss of chromosome 9 in 2/10 supratentorial tumours (B), and the loss of chromosomes 6q (7/37) and 10q (5/37), which were exclusive to recurrent posterior fossa tumours (C).

In addition to the chromosomal location of genomic imbalances seen, the other important distinction between paediatric and adult ependymoma related to the number and complexity of genomic anomalies observed. By comparing the meta-analysis data from all primary ependymoma CGH studies it was observed that adult tumours displayed nearly twice as many chromosomal aberrations as paediatric tumours, corresponding with a higher mean number of genomic anomalies per tumour (Table 1.6) (Kilday, Rahman et al. 2009). This is supported by the discovery from CGH analyses that a 'balanced' genomic profile, without chromosomal gain or loss, can be seen in 36 – 58 % of paediatric ependymomas and is associated with children under three years of age. By contrast, a balanced genome is found in less than 10 % of adult cases (Reardon, Entrekin et al. 1999; Hirose, Aldape et al. 2001; Ward, Harding et al. 2001; Carter, Nicholson et al. 2002; Dyer, Prebble et al. 2002; Modena, Lualdi et al. 2006).

Table 1.6: Comparison of the number of genomic imbalances in 116 adult and 187 paediatric primary ependymomas by CGH.

Age Group	Number of primary ependymomas	Total number of gains	Total number of losses	Total number of genomic anomalies	Mean number of anomalies per tumour
Adult	116	507	362	869	7.5
Paediatric	187	331	376	707	3.8

The meta-analysis also revealed that the genomic imbalances characteristic of adult and spinal ependymomas regularly involve whole chromosomal rearrangements, unlike the partial and complex imbalances frequently seen in paediatric cases and several aggressive adult cancers (Figures 1.6, 1.7, 1.8A, 1.8D) (Isola, Kallioniemi et al. 1995; Rooney, Boonsong et al. 2001; Kilday, Rahman et al. 2009). This may reflect intermediate ploidy, a phenomenon associated with a favourable outcome in acute lymphoblastic leukaemia, where imbalances involves whole chromosomes (Chessels, Swansbury et al. 1997). This explanation is supported, yet refined further in the paediatric population by a CGH analysis of ependymomas from 53 children (Dyer, Prebble et al. 2002). This work identified three genetically distinct subgroups within the examined cohort: the 'balanced' genomic group already discussed which was significantly associated with an age below three years, a 'structural' group demonstrating infrequent and often partial genomic imbalances and a third 'numerical' group showing frequent (13 or greater) whole chromosome gains and losses similar to

those often seen in adult ependymomas. The subdivisions were significantly associated with prognosis, with the numerical and structural groups demonstrating the best and worst patient outcome respectively. Consistent with this observation, almost all recurrent ependymomas analyzed exhibited a structural profile.

1.5.2.2 Paediatric ependymomas from different CNS locations

Despite histological similarities, ependymomas arising from the spinal, infratentorial and supratentorial compartments of the central nervous system demonstrate diverse clinical behaviour (Miller, Young et al. 1995). Indeed, the conventional CGH meta-analysis (Figures 1.8A – C) demonstrated location-specific differences between paediatric ependymomas, suggesting biological tumour heterogeneity (Kilday, Rahman et al. 2009). As stated, paediatric spinal ependymomas (Figure 1.8A) frequently demonstrated whole chromosomal imbalances such as gain of chromosomes 7, 9, 11, 18 and 20 or loss of chromosomes 1, 2 and 10, whereas intracranial ependymomas were characterized by gain of chromosome 1q and often showed loss of chromosomes 22, 3, 9p and 13q (Figure 1.8B – C).

While gain of chromosome 1q and loss of chromosome 3 were shared features (Figure 1.8B – C), genetic diversity also appeared to exist between paediatric ependymomas from different intracranial locations. The genomic loss of 9p occurred preferentially in supratentorial ependymomas (Figure 1.8B), while posterior fossa tumours often demonstrated loss of chromosomes 22, 6 and 17 (Figure 1.8C). In addition, the loss of chromosome 6q at recurrence appeared exclusive to posterior fossa ependymomas (Figure 1.9C) (Kilday, Rahman et al. 2009).

1.5.3 Higher resolution genomic analyses of paediatric ependymoma

A limited number of array CGH and SNP array studies have been performed on ependymomas, with even fewer exclusively examining paediatric tumours. Nevertheless, the conclusions drawn from these studies have supplemented those from conventional CGH work.

Identified common regions of chromosomal gain in ependymoma using aCGH have included 1p34, 1q, 2q23, 3p14, 3q29, 5p15.33, 6p21, 7p21, 7q11.23 – 22.1, 7q34, 8q11.2, 9p24.3 – qter, 11q13 – q23, 12p, 12q13.13 – 13.3, 13q21.1, 14q11.2, 14q32.2, 15q21.3, 16p11.2, 16p13.3, 16pter, 17q21, 18, 19p13.3, 20p12, Xp21.2 and Xq26.3. Frequent regions of loss were discovered at 5q31, 6q25.3, 6q26, 7q36, 9p21, 9p23, 9q24.31, 10q, 14q, 15q21.1, 16q24, 17p13.3, 19p13.2, 22q12 and 22q13.3. In addition, focal regions of amplification, encompassing particular genes of interest, were identified such as 3q25.2 (DNASE1L3), 7p11.2 (EGFR), 6p21.32 (NOTCH4), 11p13 (TYR), 9p24 (MYCN), 11q22 (YAP1, BIRC2,BIRC3), 16q22.2 (PRM1), 17p13.3 (CDC6) and 19p13.3 (VAV1). (Taylor, Poppleton et al. 2005; Mendrzyk, Korshunov et al. 2006; Modena, Lualdi et al. 2006; Mack and Taylor 2009; Puget, Grill et al. 2009).

Using array CGH, ependymomas with a 'balanced' genomic profile were again identified and associated with a young paediatric age group (Taylor, Poppleton et al. 2005; Mendrzyk, Korshunov et al. 2006; Modena, Lualdi et al. 2006; Puget, Grill et al. 2009), while the ability to cluster tumours into three genomic profiles according to the quantity and structural complexity of identified aberrations was also replicated (Mendrzyk, Korshunov et al. 2006; Puget, Grill et al. 2009). One paediatric analysis of 59 ependymomas accomplished this using the same criteria as the conventional CGH study of Dyer et al. (Dyer, Prebble et al. 2002; Puget, Grill et al. 2009), whereas another study of 68 ependymomas from a mixed age cohort identified tumours with up to two chromosomal aberrations, greater than two imbalances and an almost diploid state or greater than two imbalances and aneuploidy (Mendrzyk, Korshunov et al. 2006).

Two large array CGH studies confirmed that chromosome 1q gain was characteristic of childhood anaplastic ependymoma, by examining 171 tumours from mixed age cohorts (Taylor, Poppleton et al. 2005; Mendrzyk, Korshunov et al. 2006). Two analyses of 83 paediatric ependymomas refined this further, reporting 9q34 gain and 22q13 loss as features of tumours from older children, while gains involving 9q and 11q13 and loss of chromosome 16 were associated with infant ependymomas (Modena, Lualdi et al. 2006; Puget, Grill et al. 2009). Gains of 1q, 9q34 and loss of loci within 6q and chromosome 19 have also been correlated with disease progression in the paediatric age group (Modena, Lualdi et al. 2006; Puget, Grill et al. 2009; Peyre, Commo et al. 2010).

Tumours from different CNS locations were again noted to harbour distinct genomic imbalances. Spinal ependymomas often demonstrated gain of chromosome 16 and loss of chromosome 14q, whereas intracranial tumours were associated with 1q gain and 6q loss (Taylor, Poppleton et al. 2005; Mendrzyk, Korshunov et al. 2006). Within the paediatric population, gain of 9q33 and 9q34, together with loss of 17p13.3, were a feature of posterior fossa tumours, while loss of regions within chromosome 9p was characteristic of supratentorial ependymomas (Modena, Lualdi et al. 2006; Puget, Grill et al. 2009).

The only SNP array study of ependymomas to date is an analysis of 204 tumours from a mixed age cohort, integrating 500K SNP array copy number data with gene expression profiling performed on a sample subset (Johnson, Wright et al. 2010). This corroborated the location specific genomic aberrations detected by the discussed previous work. Indeed, while supratentorial tumours demonstrated the highest number of focal anomalies, broad genomic imbalances were most common amongst spinal tumours which were characterised by the gain and overexpression of genes on chromosomes 4, 7, 9, 12, 15 and 18q, together with the loss and underexpression of genes on 22q (Johnson, Wright et al. 2010). In addition, the deletion and underexpression of genes on chromosomes 3, 9 and 22q, or genes on chromosome 6q, were features of specific supratentorial and posterior fossa subgroups respectively (Johnson, Wright et al. 2010).

Gene expression profiling has attempted to establish candidate ependymoma oncogenes and tumour suppressor genes within regions of imbalance demonstrated from genomic analyses. For example, by assimilating gene expression and array CGH data, two paediatric ependymoma studies identified genes with potentially dysregulated copy number driven expression (Modena, Lualdi et al. 2006; Puget, Grill et al. 2009). Similarly, the integration of expression array data for a subset of 83 tumours from the SNP array cohort of 204 ependymomas enabled Johnson et al. to propose 130 candidate tumour suppressor genes and 107 oncogenes (Johnson, Wright et al. 2010). Various putative candidate genes identified from these and other gene expression analyses are detailed throughout the remainder of this work, while possible explanations for the aforementioned age and location-specific genomic aberrations are also discussed in subsequent chapters.

1.5.4 Methylation analyses of ependymoma

Various studies have analysed the methylation status of several selected genes with purported tumour suppressive functions in ependymoma (Table 1.7)

Table 1.7: Methylated genes identified in ependymoma

Study	Number of tumours studied	Methylated gene(s) identified (percentage of tumours methylated)	
Alonso et al., 2003	7	DAPK (9q21.33)	(57 %)
		MGMT (10q26.3)	(28 %)
		GSTP1 (11q13.2)	(28 %)
		CDKN2A/P14^{ARF} (9p21.3)	(28 %)
		TIMP3 (22q12.3)	(28 %)
		P73 (1p36.2)	(14 %)
		RB1 (13q14.2)	(14 %)
Gonzalez-Gomez et al., 2003	1	RB1 (13q14.2)	
Rousseau et al., 2003	108	CDKN2A/P16^{INK4A} (9p21.3)	(21 %)
	71	CDKN2B (9p21.3)	(32 %)
Alonso et al., 2004	27	TIMP3 (22q12.3)	(40 %)
		P73 (1p36.2)	(33 %)
		THBS1 (15q14)	(30 %)
		CDKN2A/ P16^{INK4A} (9p21.3)	(18 %)
		MGMT (10q26.3)	(20 %)
		NF2 (22q12.2)	(<10 %)
		CASP8 (2q33.1)	(<10 %)
		RB1 (13q14.2)	(<10 %)
Waha et al., 2004	52	HIC1 (17p13.3)	(83 %)
Hamilton et al., 2005	35	RASSF1A (3p21.3)	(86 %)
	20	CASP8 (2q33.1)	(20 %)
		MGMT (10q26.3)	(5 %)
		P73 (1p36.2)	(5 %)
Michalowski et al., 2006	27	RASSF1A (3p21.3)	(56 %)
		TNFRSF10D (8p21.3)	(36.4 %)
		CASP8 (2q33.1)	(30 %)
		FHIT (3p14.2)	(22 %)
		RARB (3p24.2)	(14.8 %)
		BLU (3p21.3)	(13.6 %)
		CDKN2A/P16^{INK4A} (9p21.3)	(11.1 %)
		TNFRSF10C (8p21.3)	(9.5 %)
		DAPK (9q21.33)	(7.4 %)
Modena et al., 2006	18	CDKN2A/P16^{INK4A} (9p21.3)	(11 %)
		HIC1 (17p13.3)	(83 %)

Frequently methylated genes identified to date include the pro-apoptotic genes RASSF1A, P73 and CASP8, the metalloproteinase inhibitor TIMP3 and genes implicated in transcriptional repression (HIC1), cell cycle regulation (CDKN2A) and DNA repair (MGMT). However, only the hypermethylation of gene regulatory regions (relative to normal DNA controls) have been established from these studies, whereas recent data suggests that hypomethylation of intergenic regions may also be associated

with tumourigenesis and progression in several tumours, including ependymoma (Xie, Wang et al. 2010). Moreover, all of the previous studies used a methylation specific polymerase chain reaction (PCR) technique and, due to technological limitations, restricted their analysis to pre-selected candidate genes which were chosen predominantly because of their altered methylation states in other malignancies. The advent of DNA microarrays has now enabled the methylation status of CpG dinucleotide sites within multiple genes across the genome to be analysed simultaneously.

1.5.5 Ependymoma-initiating cells

As stated earlier in this chapter (section 1.1.3), there is increasing awareness that the genetic and molecular diversity demonstrated by tumours, such as ependymomas, are the consequence of being organised as a developmental hierarchy, originating from tumour initiating cells with stem-like characteristics.

Embryonic and postnatal ependymal cells have previously been shown to originate from a subset of radial glia, indicating their potential as such neural stem or precursor cells (Spassky, Merkle et al. 2005). Gene expression analysis has shown that signature genes of supratentorial and spinal ependymomas are similarly expressed by murine embryonic radial glial cells in the corresponding region of the CNS (Taylor, Poppleton et al. 2005; Poppleton and Gilbertson 2007). Using protein markers representative of early differentiation stages in the subventricluar zone, Taylor et al. also found that ependymoma derived neurospheres displayed both a stem (CD133+/Nestin+) and glial (Rc2+/Blbp+) immunophenotype, indicating a radial glial rather than neuroepithelial derivation for this tumour. The capability of these radial glial-like cells to self renew and exhibit multilineage differentiation in vitro, while form orthotopic tumours in vivo (Taylor, Poppleton et al. 2005; Johnson, Wright et al. 2010) renders them a convincing candidate for the ependymoma-initiating cell within which an initial malignant transformation event occurs.

Further evolution of this concept comes from recent cross-species transcriptomic analysis. By correlating the mRNA expression profile of human ependymomas to those

of a distinct range of murine radial glia and adult neural stem cells, Johnson et al. identified neural stem cells with specific regional, genetic and developmental characteristics as potential cells of origin for particular ependymoma variants (Johnson, Wright et al. 2010). Human spinal ependymomas had a gene expression profile that closely corresponded to that of adult spinal murine neural stem cells. In contrast, supratentorial tumours had a transcriptome correlating with embryonic radial glia derived from the cerebrum of INK4A/ARF$^{-/-}$ mice. Activation of EphB2 signalling in this latter radial glia murine subtype was capable of ependymoma initiation upon xenograft transplantation in NOD-SCID mice. The histological appearance and transcriptomic profile of the resulting tumours modelled that of a previously identified human supratentorial ependymoma subgroup (Johnson, Wright et al. 2010), providing evidence that regionally defined radial glia are susceptible to malignant transformation from driver mutations.

It could be argued that the mouse may not represent a physiologically relevant microenvironment for engraftment and growth of human tumours, thus potentially resulting in an underestimation of cell types capable of tumour initiation. An additional unresolved question is whether the cell type displaying the CD133+/Nestin+/RC2/BLBP+ immunophenotype with a capacity for self-renewal in vitro and tumour development in murine xenotransplantation is the same cell type that is ependymoma initiating in humans. Although likely, definitive experiments have not yet been established. It also remains unknown whether normal ependymal cells can transform directly into a neoplastic cell, although in vitro analysis of other CNS tumours suggest the malignant transformation of mature cells is improbable (Alcantara Llaguno, Chen et al. 2009; Jacques, Swales et al. 2010).

If the ependymoma-initiating cell hypothesis proves true, future therapeutic strategies may deviate away from being centred around the primary tumour mass and focus on targeting the putative ependymoma-initiating cells, a topic discussed further in the final chapter.

1.5.6 Cell signalling pathways

Various cell signalling pathways have been reported to be involved in ependymoma pathogenesis. These include Sonic Hedgehog (Modena, Lualdi et al. 2006; Palm, Figarella-Branger et al. 2009), Wnt (Suarez-Merino, Hubank et al. 2005; Palm, Figarella-Branger et al. 2009; Peyre, Commo et al. 2010) and the EPHB-EPHRIN tyrosine kinase pathway (Taylor, Poppleton et al. 2005; Johnson, Wright et al. 2010). However, the most consistently implicated pathway is the Notch signalling pathway (Taylor, Poppleton et al. 2005; Modena, Lualdi et al. 2006; Poppleton and Gilbertson 2007; Palm, Figarella-Branger et al. 2009; Puget, Grill et al. 2009; Johnson, Wright et al. 2010; Peyre, Commo et al. 2010).

The Notch pathway is thought to contribute to the regulation of cell proliferation and determination of cell fate in the CNS by mediating cell to cell interaction (Figure 1.10). In summary, a membranous ligand for Notch (Delta/Jagged/Serrate) on the surface of a cell binds to the corresponding Notch receptor of a neighbouring cell. This prompts enzyme induced Notch cleavage (via tumour necrosis factor alpha converting enzyme and gamma-secretase), resulting in the cytoplasmic release of its intracellular domain (Notch-ICD) which ultimately penetrates the nucleus. The transcription factor recombination signal-binding protein suppressor of hairless (RBPSUH/CSL) then binds with the Notch-ICD fragment to activate the transcription of downstream targets, such as the HES gene family. The encoded HES proteins are transcriptional repressors from the basic helix-loop-helix (bHLH) family which negatively regulate genes known to induce neuronal differentiation (Radtke and Raj 2003). Conversely, in the absence of Notch/HES activation, other transcription factors from the bHLH family that promote differentiation can be expressed. These include the ASH proteins which facilitate neuronal maturation through the downstream activation of genes such as those of the Neurogenin family and NeuroD (Scotting 2004).
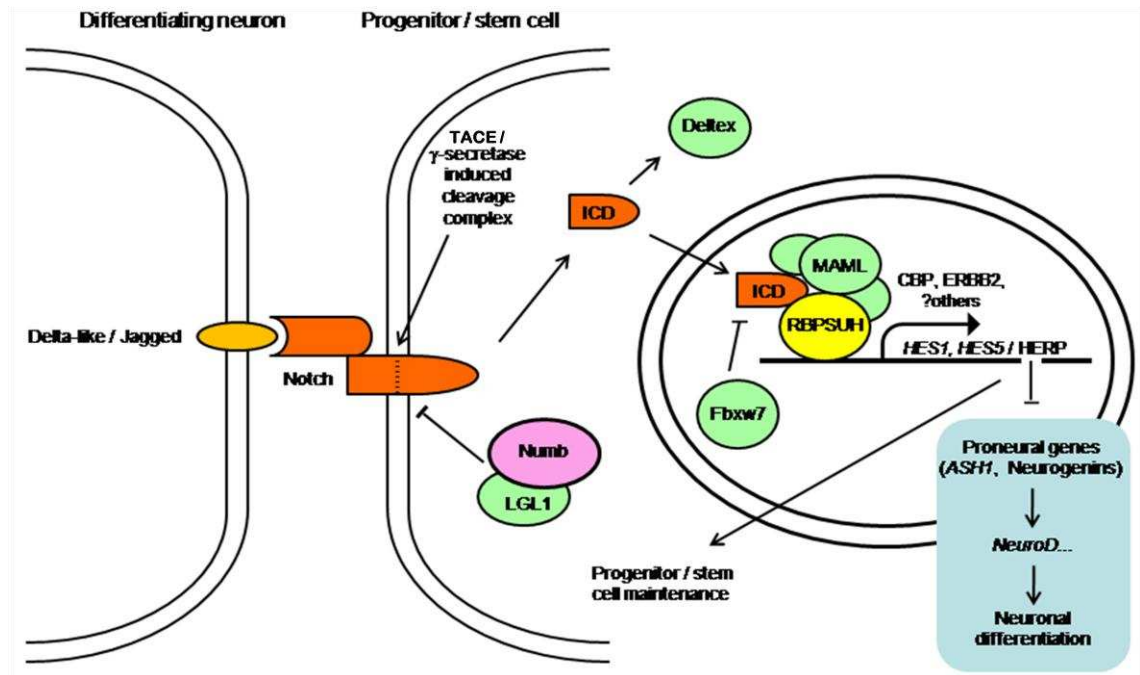
Figure 1.10: Overview of the Notch signalling pathway. Adapted from (Bray 2006). See text for an overview of the pathway and definition of particular components. In addition to activation, the pathway can be inhibited by mechanisms that include the activation of the cytoplasmic Notch inhibitor Numb, or degradation of the Notch-ICD by activated nuclear Fbxw7. ICD = intracellular domain, LGL1 = lethal giant larvae 1, Fbxw7 = F box-w7, MAML = Mastermind like, RBPSUH = recombination signal-binding protein suppressor of hairless, CBP = CREB binding protein.

In this context, activation of the Notch pathway can thereby promote progenitor cell proliferation and inhibit differentiation. Indeed, maintenance of the self-renewal and multipotential characteristics of adult neural stem cells can be induced by the upregulation of the Notch ligand Jagged-1, whereas NOTCH1 (9q34) deletion reduces the stem cell proportion within a cell population (Hitoshi, Alexson et al. 2002; Nyfeler, Kirch et al. 2005). Inactivation of the Notch inhibitor NUMB1 in the neural stem cells of lethal giant larvae-1 (LGL1) null mice has also resulted in asymmetric cell division and CNS neoplastic lesions (Klezovitch, Fernandez et al. 2004). Moreover, the sustained proliferation of radial glia, the purported ependymoma-initiating progenitors, has been achieved through retroviral vector induced activation of NOTCH1 in the embryonic murine subventricular zone (Gaiano, Nye et al. 2000) or upregulation of the Notch target ERBB2 (17q12), a transcriptional repressor belonging to the HERP gene family which also inhibits neuronal differentiation (Ever and Gaiano 2005).

Gene expression and IHC analyses of ependymomas also support involvement of Notch signalling in ependymoma pathogenesis and progression by demonstrating the overexpression of several pathway members such as Delta-like and Jagged ligands

(DLL1 (6q27), DLL3 (19q13.2), JAG1 (20p12.2) and JAG2 (14q32.33)), Notch receptors (NOTCH1, NOTCH2 (1p11.2)) and downstream effectors and targets (RBPSUH (5q11), MAML2 (11q21), HES1, HES2 (1p36.31), HES5, HEY2 (6q22.1), DTX1 (12q24.13), LFNG (7p22.3), FBXW7 (4q31.3), C-MYC (8q24.21), ERBB2) (Gilbertson, Bentley et al. 2002; Taylor, Poppleton et al. 2005; Modena, Lualdi et al. 2006; Palm, Figarella-Branger et al. 2009; Puget, Grill et al. 2009; Johnson, Wright et al. 2010; Peyre, Commo et al. 2010).

1.5.7 Biological prognostic markers for paediatric ependymoma

Several putative biological prognostic markers for ependymoma have been suggested using immunohistochemistry (IHC) and genomic analysis on retrospective cohorts (Table 1.8). However, only a few of these candidates have been analysed in sufficient numbers of childhood ependymoma to allow consideration as prognostic markers in this age group. Protein markers include human telomerase reverse transcriptase (hTERT) and its nuclear chaperone Nucleolin (Tabori, Ma et al. 2006; Ridley, Rahman et al. 2008), members of the receptor tyrosine kinase 1 (RTK1) family (Gilbertson, Bentley et al. 2002) and the Ki-67 labelling index (Bennetto, Foreman et al. 1998; Gilbertson, Bentley et al. 2002). While the number and pattern of genomic imbalances in paediatric ependymoma have a proposed prognostic significance (Dyer, Prebble et al. 2002; Puget, Grill et al. 2009), specific aberrations reported to be implicated in disease progression and patient outcome include gains of chromosome 1q and 9q, and loss of 6q and 22q.

Neoplastic cells have a capacity for unlimited proliferation and the ability to avoid replicative senescence, both of which require maintenance of telomere length (Cong and Bacchetti 2000). As previously stated, telomeres are nucleoproteins which cap human chromosomal termini, thereby contributing to genomic stability (Counter, Avilion et al. 1992; Hackett, Feldser et al. 2001). In normal somatic cells successive mitotic divisions result in telomere erosion, culminating in either senescence or apoptosis (Harley, Futcher et al. 1990; Feldser and Greider 2007). In contrast telomere maintenance is present in almost all types of malignant cells, mediated by the enzyme telomerase (Shay and Bacchetti 1997). Telomerase functions as a homodimer and contains a catalytic component with reverse transcriptase activity (hTERT) which may be implicated in

paediatric tumourigenesis development through avoidance of transcriptional repression by histone deacetylation in utero (Cong and Bacchetti 2000).

Amplification of the hTERT locus (5p13.3) has been reported in a proportion of intracranial ependymoma cases analyzed by array CGH (Mendrzyk, Korshunov et al. 2006). Moreover, an IHC analysis of 65 paediatric ependymomas appeared to find increased hTERT expression as an independent predictor of adverse patient survival (Tabori, Ma et al. 2006). However, re-evaluation of the antibody (NCL-hTERT) used to detect hTERT in this study verified its actual target to be Nucleolin, a phosphoprotein that acts as a chaperone for hTERT during its transport from nucleolus to nucleoplasm (Wu, Dudognon et al. 2006). Indeed, a subsequent retrospective IHC assay of hTERT, Nucleolin, RTK1 members, Ki-67 and Survivin expression in 80 paediatric intracranial ependymomas only established prognostic significance for altered Nucleolin expression, as detected by NCL-hTERT and a second independent monoclonal antibody (Ab13541 Abcam). Ependymomas exhibiting high Nucleolin expression were associated with a poor outcome when compared to tumours with low Nucleolin expression, with a five year event-free survival for patients of 31 % versus 74 % respectively (Ridley, Rahman et al. 2008). The study also demonstrated telomerase-induced telomere maintenance in over 70 % of cases, with telomere lengthening apparent in over half of the relapsed tumours. However, neither telomerase activity, nor telomere length correlated with patient outcome, possibly due to the small cohort size (Ridley, Rahman et al. 2008).

The RTK1 family of proteins, which includes ERBB2, ERBB3, ERBB4 and epidermal growth factor receptor (EGFR), are responsible for a variety of cellular processes such as cell division, motility and survival. Indeed, more recent data suggests ERBB2 overexpression may potentiate radial glia proliferation (Poppleton and Gilbertson 2007). Members of the RTK1 family have been assessed as potential prognostic markers in paediatric ependymoma. An IHC assay of 59 tumours revealed ERBB2/ERBB4 co-expression, together with a raised Ki-67 tumour proliferation index, was associated with reduced paediatric patient survival although ERBB2/ERBB4 co-expression alone did not correlate with outcome (Gilbertson, Bentley et al. 2002). Support for EGFR as a prognostic marker comes from two studies analysing mixed age cohorts. An immunohistochemical analysis of 46 ependymomas proposed EGFR as a relapse marker for low grade tumours, whilst an array CGH study of 68 ependymomas correlated poor

prognosis with frequent genomic gains and high-level amplifications covering the EGFR locus (Korshunov, Golanov et al. 2002; Mendrzyk, Korshunov et al. 2006). However, these findings contrast with results from the aforementioned IHC assay performed exclusively on paediatric ependymomas which did not confirm a prognostic value for RTK1 markers (Ridley, Rahman et al. 2008). These findings suggest a possible difference in the mechanism of EGFR, ERBB2 and ERBB4 expression between paediatric and adult ependymomas, reinforcing that these two subgroups are distinct genetic entities and raising the possibility that anti-RTK1 therapy may prove disappointing for those of a younger age at diagnosis. Current trials of Laptinib (GW572016), a dual inhibitor of EGFR and ERBB2 receptor signalling, within the Paediatric Brain Tumour Consortium and Collaborative Ependymoma Research Networks, may help to resolve this concern.

Expression of the proliferative marker Ki-67 has been identified as a putative prognostic marker in studies of ependymoma in children (Bennetto, Foreman et al. 1998; Figarella-Branger, Civatte et al. 2000; Gilbertson, Bentley et al. 2002; Zamecnik, Snuderl et al. 2003). In contrast, other analyses have found no association between this marker and patient survival (Shuangshoti, Rushing et al. 2005; Ridley, Rahman et al. 2008), thereby making definitive conclusions on its use as a paediatric prognostic marker difficult. Moreover, Ki-67 has been shown to correlate with tumour grade, supporting the view that increased proliferation appears a feature of histological anaplasia (Rushing, Brown et al. 1998; Suzuki, Oka et al. 2001; Suri, Tatke et al. 2004; Ridley, Rahman et al. 2008).

The CGH meta-analysis revealed chromosome 1q gain as the most frequent genomic aberration in primary and recurrent paediatric intracranial ependymoma (Figures 1.6 and 1.9) (Kilday, Rahman et al. 2009), justifying its further assessment as a prognostic marker in children. Genomic gain of chromosome 1q is not an observation restricted to paediatric ependymomas. Indeed, this aberration is one of the most common in human cancer (Schutte, Carpten et al. 2001). Chromosome 1q gain has been identified in other paediatric brain tumours such as CNS PNETs (Pfister, Remke et al. 2007), and medulloblastomas (Lo, Rossi et al. 2007), while it has been associated with an unfavourable outcome in Wilms' tumours, neuroblastomas and Ewing's sarcomas,

suggesting a generic role in paediatric tumour progression and recurrence (Hirai, Yoshida et al. 1999; Hing, Lu et al. 2001; Ozaki, Paulussen et al. 2001).

It has been suggested that smaller regions of chromosome 1q may be of particular functional and prognostic importance, thereby warranting further assessment. Indeed, gain of 1q21 − 32 in ependymoma has been associated with an anaplastic histology and adverse prognosis (Carter, Nicholson et al. 2002; Dyer, Prebble et al. 2002; Mendrzyk, Korshunov et al. 2006; Korshunov, Witt et al. 2010). Specifically, two array CGH studies of ependymomas from sizeable mixed age cohorts demonstrated that gain of 1q25, as validated by fluorescent in situ hybridisation (FISH), represented an independent prognostic marker for both recurrence free and overall survival (Mendrzyk, Korshunov et al. 2006; Korshunov, Witt et al. 2010). One such analysis also identified a particular area of recurrent gain at 1q23.3 (Mendrzyk, Korshunov et al. 2006). Gene expression studies have demonstrated the upregulation of candidate oncogenes genes located within this 1q21 − 32 region, including POGZ (1q21.3), S100 family members (1q21.3), HSPA6 (1q23) TPR (1q25), LAMB1 (1q31), PRELP (1q32), CHI3L1 (1q32), and GAC1 (1q32) (Korshunov, Neben et al. 2003; Suarez-Merino, Hubank et al. 2005; Karakoula, Suarez-Merino et al. 2008; Rand, Prebble et al. 2008; Johnson, Wright et al. 2010). While GAC1 amplification has already been implicated in the pathogenesis of malignant gliomas (Almeida, Zhu et al. 1998), overexpression of the nucleoprotein encoding gene TPR has been associated with adverse outcome in a small study of paediatric patients with intracranial ependymoma (Karakoula, Suarez-Merino et al. 2008). In addition to the 1q21 − 32 locus, the recent SNP array analysis of 204 paediatric and adult ependymomas also identified a novel region of amplification at 1q42.13 in a subpopulation of posterior fossa ependymomas harbouring potential oncogenes such as the RAS superfamily member ARF1 (Johnson, Wright et al. 2010).

Gain of 9q33 − 34 and the incorporated putative oncogenes TNC and NOTCH1 have been associated with paediatric ependymoma recurrence (Puget, Grill et al. 2009), which contrasts with reports of 9q gain as a favourable prognostic marker for ependymoma patients from a mixed age cohort (Korshunov, Witt et al. 2010). Similar dichotomous results are evident in ependymoma for genomic deletions involving chromosome 6. While loss of the entire chromosome, or specifically 6q25, have been identified as independent markers of improved survival for both children and adults

(Monoranu, Huang et al. 2008; Korshunov, Witt et al. 2010), 6q deletion was a feature of relapsed ependymomas from paediatric CGH analyses (Figure 1.9) and deletion of the 6q23 and 6q25.2 loci have been associated with disease progression (Rajaram, Gutmann et al. 2005; Peyre, Commo et al. 2010).

Within chromosome 22, the 22q13.1 and 22q13.3 loci have been highlighted as frequent regions of genomic loss in ependymoma (Rousseau-Merck, Versteege et al. 2000; Huang, Starostik et al. 2002; Modena, Lualdi et al. 2006; Johnson, Wright et al. 2010). Gene expression analyses within these loci have revealed the underexpression of several candidate tumour suppressor genes with roles in gene silencing (TNRC6B, CBX7 (22q13.1)), cell cycle regulation (MCM5 (22q13.1)), DNA repair (G22P1 (22q13.2)) and neurotransmitter metabolism (SULT4A1 (22q13.3)) (Korshunov, Neben et al. 2003; Suarez-Merino, Hubank et al. 2005; Modena, Lualdi et al. 2006; Johnson, Wright et al. 2010). In addition, a real-time quantitative PCR analysis of 47 paediatric intracranial ependymomas revealed frequent loss of C22orf2 (22q13.1) and RAC2, (22q13.1), the latter being associated with a reduced overall survival in the cohort (Karakoula, Suarez-Merino et al. 2008).

The identification of biological correlates of clinical outcome from a retrospective population is an important finding. However, in order to confirm validity, these need to be tested in an independent, cohort treated in a homogenous manner, ideally within the setting of a clinical trial. Moreover, the detection method for any identified marker must be rapid, robust, standardized and easy to interpret to allow diagnostic laboratories to identify patients suitable for a particular therapeutic proposal. At present, published prospective validation is lacking for putative biological markers in paediatric ependymoma.

Table 1.8: Statistically significant putative biological prognostic markers (immunohistochemical and genomic) for paediatric ependymoma.

Immunohistochemical marker	Source	Prognostic Feature	Patient outcome (p-value)	Use as a prognostic marker in paediatric ependymoma	
				Strength	Weakness
Apoptotic index (AI - using ISEL)	Korshunov et al., 2002 (cohort expansion of 1999 study)	Apoptotic index < 1 %	Worse PFS for entire cohort (p = 0.002) and high grade tumours (p = 0.0001)	Large cohort size (n = 112). Statistical significance in univariate analysis.	No statistically significant association in multivariate analysis Mixed age cohort. No prospective validation.
bcl-2	Zamecnik et al., 2003	Increased expression	Worse PFS and OS (p < 0.001)	Paediatric cohort analyzed exclusively. Statistical significance in univariate analysis.	No statistically significant association in multivariate analysis. Small cohort size (n = 31). No prospective validation. Other studies refute prognostic association (Verstegen et al., 2002)
Bromodeoxyuridine (BrdU) Labelling Index	Asai et al., 1992	Labelling index > 1 % - inverse correlation between BrdU index and time to recurrence	Early recurrence rate (p < 0.05)	Statistical significance in univariate analysis	Mixed age cohort Small cohort size (n = 32) No prospective validation
Cyclin D1 Labelling Index	Zamecnik et al., 2003	Labelling Index > 5 %	Worse PFS (p = 0.049)	Paediatric cohort analyzed exclusively. Statistical significance in univariate analysis.	No statistically significant association for OS in univariate analysis or OS & PFS in multivariate analysis. Small cohort size (n = 31). No prospective validation. Other studies refute association with prognosis (Prayson et al., 1999)
EVI1	Koos et al., 2011	Increased expression	Worse PFS in posterior fossa ependymomas (p = 0.02)	Statistical significance in multivariate analysis.	Mixed age cohort. Small cohort size (n = 28). No prospective validation.
Glial Fibrillary Acidic Protein (GFAP)	Figarella-Branger et al., 1991	Increased expression in > 30 % tumour cells GFAP/ vimentin ratio < 1	Improved OS (p < 0.05) Worse OS (p < 0.05)	Paediatric cohort analyzed exclusively. Statistical significance in univariate analysis.	Small cohort size (n = 16). Other paediatric study refutes prognostic association (Zamecnik et al., 2003). No prospective validation.
Human Telomere Reverse Transcriptase (hTERT)	Tabori et al., 2006	Increased expression	Reduced PFS and OS (p = 0.002)	Relatively large cohort size (n =87). Paediatric cohort exclusively analyzed. Statistical significance in multivariate analysis.	All studies No prospective validation Re-evaluation of antibody used to detect hTERT verified the actual target to be Nucleolin (Wu et al., 2006)
	Mendrzyk et al., 2006	Increased expression	Worse OS (p = 0.01)	Large cohort size (n = 170). IHC validated genomic gains at hTERT locus detected by aCGH.	Mendrzyk study Mixed age cohort. No statistically significant association in multivariate analysis.

Immunohistochemical marker	Source	Prognostic Feature	Patient outcome (p-value)	Use as a prognostic marker in paediatric ependymoma	
				Strength	**Weakness**
Hypoxia related tissue factors	Preusser et al., 2005	High hypoxia score (increased expression of 2 or 3 of the markers VEGF, carbonic anhydrase 9 (CA9) and hypoxia – inducible factor 1 alpha (HIF-1α)	Worse OS (p = 0.0402)	Large cohort size (n = 100). Statistical significance in univariate analysis.	No statistically significant association for individual markers in univariate analysis. No statistically significant association for hypoxia score in multivariate analysis. Mixed age cohort. No prospective validation.
Ki-67/MIB-1 Labelling Index	Rezai et al., 1996	High MIB-1 labelling index (13 % versus 2 %)	Tumour tendency for CSF dissemination (p = 0.02)	Statistical significance in univariate analysis; large cohort size (n = 140).	All studies No prospective validation. Other paediatric studies refute association with prognosis (Shuangshoti et al., 2005; Ridley et al., 2008). Variable LI cut-offs to define prognostic groups.
	Ritter et al., 1998	MIB-1 LI > 20 %	Worse OS (p = 0.0013)	Statistical significance in CART analysis.	Rezai, Ritter, Ho, Preusser and Verstegen studies Mixed age cohorts.
	Ho et al., 2001	MIB-1 LI > 9 %	Worse OS and PFS (p < 0.001)	Relatively large cohort (n = 81). Significance in univariate analysis.	Zamecnik and Ritter studies Small cohort sizes (n = 36 & 34).
	Zamecnik et al., 2003	MIB-1 LI > 7 %	Worse OS and PFS (p = 0.002)	Paediatric cohort analyzed exclusively. Statistical significance in multivariate analysis.	Verstegen study No statistically significant association in multivariate analysis.
	Preusser et al., 2005(cohort expansion of other 2005 study)	MIB-1 LI > 20.4 %	Worse OS in univariate analysis (p = 0.00001) and multivariate analysis (p = 0.010)	Statistical significance multivariate analysis. Large cohort (n = 100).	
	Verstegen et al., 2002	MiB-1 LI > 1 %	Worse OS in univariate analysis (p = 0.02)	Statistical significance univariate analysis. Relatively large cohort (n = 51).	
	Benetto et al., 1998	Elevated Ki-67 LI in infratentorial cases (> 25 %)	Worse OS (p < 0.002)	Paediatric cohorts analyzed exclusively. Relatively large cohort (n = 74). Statistical significance in multivariate analysis.	All studies No prospective validation. Variable LI cut-offs to define prognostic groups. Cohort sizes remain inadequate for reliable multivariate analysis. Other studies refute association with prognosis (Ridley et al., 2008; Prayson et al., 1999).
	Gilbertson et al., 2002	Ki-67 LI > 25 %	Worse OS (p < 0.009)	Paediatric cohorts analyzed exclusively. Relatively large cohort (n = 89).	Gilbertson and Fig-Branger studies No statistically significant association in multivariate analysis.
	Figarella-Branger et al., 2000	Ki-67 LI > 1 %	Worse PFS (p = 0.006)	Paediatric cohorts analyzed exclusively.	Figarella-Branger study Small cohort size (n = 37).

Immunohistochemical marker	Source	Prognostic Feature	Patient outcome (p-value)	Use as a prognostic marker in paediatric ependymoma	
				Strength	Weakness
Metallothionine	Korshunov et al., 1999	Increased expression	Lower recurrence risk (p = 0.005)	Statistical significance in multivariate analysis. Relatively large cohort size (n = 76).	Mixed age cohort. No prospective validation. Cohort size remains inadequate for reliable multivariate analysis.
Metalloproteinase (MMP) – 2	Snuderl et al., 2008	Increased expression	Reduced EFS in completely resected tumours (p = 0.02)	Statistical significance in univariate analysis. Paediatric cohorts analyzed exclusively.	No statistically significant association in multivariate analysis. No statistically significant association for OS. Small cohort size (n = 28). No prospective validation.
Metalloproteinase (MMP) – 14	Snuderl et al., 2008	Increased expression	Reduced OS in completely resected (p = 0.046)	Statistical significance in multivariate analysis. Paediatric cohorts analyzed exclusively.	Small cohort size (n = 28). No prospective validation.
Mos protein	Athanasiou et al., 2003	Labelling index > 10 %	Worse PFS (p = 0.05)	Statistical significance in univariate analysis.	No statistically significant association for OS. No multivariate analysis. Mixed age cohort. Small cohort size (n = 34). No prospective validation.
NEFL (Neurofilament light polypeptide 70)	Andreiuolo et al., 2010	Strong expression (> 5 % positive cells)	Better PFS (p = 0.007)	Statistical significance in univariate analysis. Paediatric cohorts analyzed exclusively.	No statistically significant association for OS. Small cohort size (n = 34). No multivariate analysis. No prospective validation.
Nucleolin	Ridley et al., 2008	Increased expression	Reduced EFS and OS (p = 0.007)	Relatively large cohort size (n = 80). Paediatric cohort exclusively analyzed. Statistical significance in multivariate analysis.	No prospective validation.
CDKN2A / p14 ARF Labelling Index	Korshunov et al., 2002 (cohort expansion of 1999 study)	Labelling index < 10 %	Worse PFS for entire cohort (p = 0.0001) and high grade tumours (p = 0.0001). Higher recurrence risk (p = 0.001).	Large cohort size (n = 112). Statistical significance in multivariate analysis. CDKN2A reported tumour suppressor gene in various cancers.	Mixed age cohort. No prospective validation. Other work refutes prognostic association of CDKN2A (Rajaram et al., 2004).
p27/Kip 1 Labelling Index	Korshunov et al., 2002 (cohort expansion of 1999 study)	Labelling Index < 20 %	Reduced PFS in entire cohort (p = 0.01) and high grade tumours (p = 0.00003)	Large cohort size (n = 112). Statistical significance in univariate analysis.	No statistically significant association in multivariate analysis. Mixed age cohort. No prospective validation.

Immunohistochemical marker	Source	Prognostic Feature	Patient outcome (p-value)	Use as a prognostic marker in paediatric ependymoma	
				Strength	Weakness
p53	Korshunov et al., 2002 (cohort expansion of 1999 study)	Immunohistochemical positivity (LI = 14 to 32 %)	Worse PFS for entire cohort (p = 0.0001) and high grade tumours (p = 0.0001). Higher recurrence risk (p = 0.01).	Large cohort size (n = 112). Statistical significance in multivariate analysis.	All studies No prospective validation. p53 antibody detects mutant and wild type p53. Mutations of p53 appear rare in ependymoma. Other work refutes prognostic association (Shuangshoti et al., 2005)
	Verstegen et al., 2002	Increased expression > 1 %	Worse OS in univariate analysis (p = 0.02) and multivariate analysis(p = 0.013)	Relatively large cohort size (n = 51). Statistical significance in multivariate analysis.	Verstegen and Korshunov studies Mixed age cohorts. Zamecnik study Small cohort (n = 31). No statistical significance in multivariate analysis.
	Zamecnik et al., 2003	Increased expression	Worse PFS (p<0.001)	Paediatric cohort analyzed. Statistical significance in univariate analysis.	
P-glycoprotein	Korshunov et al., 1999	Increased expression	Lower recurrence risk (p = 0.02)	Statistical significance in multivariate analysis. Relatively large cohort size (n = 76).	Mixed age cohort No prospective validation Cohort size remains inadequate for reliable multivariate analysis.
Receptor Tyrosine Kinase 1 family (EGFR, ERBB2 – 4)	Korshunov et al., 2002 (cohort expansion of 1999 study)	Immunohistochemical positivity	Worse PFS for entire cohort (p = 0.005) and low grade tumours (p = 0.002)	Large cohort size (n = 112).	All studies No statistically significant association in multivariate analysis. No prospective validation. Other work refutes prognostic association of all RTK1 family in paediatric cases (Ridley et al., 2008)
	Gilbertson et al., 2002	Combination of incomplete tumour resection and ERBB2/ERBB4 co-expression or Ki-67 LI > 25 %	Worse OS (p < 0.0001)	Relatively large cohort size (n = 59). Paediatric cohort exclusively analyzed.	Gilbertson study ERBB2/ERBB4 co-expression alone not associated with poor outcome.
	Mendrzyk et al., 2006	Immunohistochemical positivity	Worse OS (p=0.002)	Large cohort size (n = 170). IHC validated genomic gains and amplifications at EGFR locus detected by aCGH.	Korshunov and Mendrzyk studies Mixed age cohorts.
Survivin	Preusser et al., 2005	Increased expression	Worse OS (p = 0.0032)	Statistical significance on univariate analysis. Relatively large cohort size (n = 63).	No statistically significant association in multivariate analysis. Cohort size remains inadequate for reliable multivariate analysis. Mixed age cohort. Other paediatric studies refute prognostic association (Ridley et al 2008) or show association between low expression and aggressive disease (Alturas et al., 2003)

Immunohistochemical marker	Source	Prognostic Feature	Patient outcome (p-value)	Use as a prognostic marker in paediatric ependymoma	
				Strength	Weakness
Tenascin	Korshunov et al., 2002 (cohort expansion of 1999 study)	Immunohistochemical positivity	Worse PFS for entire cohort (p = 0.0001) and low grade tumours (p = 0.0001). Higher recurrence risk (p = 0.001).	Large cohort size (n = 112). Statistical significance in multivariate analysis.	Korshunov study Mixed age cohort. No prospective validation.
	Zamecnik et al., 2004 (cohort expansion of 2003 study)	Immunohistochemical detection in intercellular spaces and blood vessel walls	Worse PFS (p = 0.012)	Paediatric cohort analyzed. Statistical significance in multivariate analysis.	Zamecnik study Small cohort size (n = 36). No prospective validation.
Topoisomerase-II alpha (Ki-S1) Labelling Index	Korshunov et al., 2002 (cohort expansion of 1999 study)	Labelling Index > 5 %	Worse PFS for entire cohort (p = 0.001) and low grade tumours (p = 0.0001)	Large cohort size (n = 112). Statistical significance in univariate analysis.	All studies No statistically significant association in multivariate analysis. No prospective validation. Variable LI cut offs between studies.
	Zamecnik et al., 2003	Labelling Index > 12 %	Worse PFS and OS (p < 0.001)	Paediatric cohort analyzed exclusively. Statistical significance in univariate analysis.	Korshunov study Mixed age cohort. Zamecnik study Small cohort (n = 31).
Vitronectin	Zamecnik et al., 2004 (cohort expansion of 2003 study)	Immunohistochemical detection at tumour invasion front	Worse PFS (p = 0.005)	Paediatric cohort analyzed exclusively. Statistical significance in univariate analysis.	Small cohort size (n = 36) No statistically significant association in multivariate analysis. No prospective validation.
Vascular Endothelial Growth Factor (VEGF)	Korshunov et al., 2002 (cohort expansion of 1999 study)	Immunohistochemical positivity	Worse PFS for entire cohort (p = 0.003) and low grade tumours (p = 0.001)	Large cohort size (n = 112). Statistical significance in univariate analysis.	No statistically significant association in multivariate analysis. Mixed age cohort. No prospective validation. Other work refutes prognostic association (Pietsch et al., 1997).
Genomic/Genetic marker	Source	Prognostic Feature	Patient outcome (P value)	Use as a prognostic marker in paediatric ependymoma	
				Strength	Weakness
Genomic gain of 1q	Carter et al., 2002	Gain of 1q (CGH)	Worse OS for intracranial tumours (p < 0.05)	Statistical significance in univariate analysis.	All studies Mixed age cohort. No prospective validation.
	Mendrzyk et al., 2006	Gain of 1q25 (aCGH/FISH)	Worse PFS (p < 0.001) and OS (p = 0.003) in intracranial tumours	Statistical significance in multivariate analysis. Supported by other work (Karakoula et al., 2008, Korshunov et al., 2010)	Carter study Small cohort size (n = 31) No statistically significant association in multivariate analysis.
		Gain of 1q21.1-32.1 (aCGH/FISH)	Higher recurrence rate in intracranial tumours (p < 0.001)	Relatively large cohort (n = 68). FISH validated aCGH findings.	
	Korshunov et al., 2010	Gain of 1q (aCGH/FISH)	Worse PFS (p < 0.001) and OS (p < 0.001) in intracranial tumours	Large cohort (n = 122). Statistical significance in multivariate analysis.	

Genomic/Genetic marker	Source	Prognostic Feature	Patient outcome (p-value)	Use as a prognostic marker in paediatric ependymoma	
				Strength	Weakness
Genomic gain of 9q33 and 9q34	Puget et al., 2009	9q33 / 9q34 gain (aCGH/FISH)	Associated with tumour recurrence (9q3, p = 0.003; 9q34, p = 0.009)	Relatively large cohort size (n = 59). Statistically significant on supervised classification. Paediatric cohort exclusively analyzed.	No patient survival data. No prospective validation. Conflicting results from other aCGH studies (Korshunov et al., 2010).
Genomic gain of 9q	Korshunov et al., 2010	Gain of 9q (aCGH/FISH)	Improved PFS (p = 0.01) and OS (p = 0.002)	Large cohort (n = 122). Statistical significance in univariate analysis.	Mixed age cohort. No statistically significant association in multivariate analysis. No prospective validation. Conflicting results from other aCGH work (Puget et al., 2009).
Genomic gain of 15q	Korshunov et al., 2010	Gain of 15q (aCGH/FISH)	Improved PFS (p = 0.04 – multivariate analysis) and OS (p = 0.002 – univariate analysis)	Large cohort size (n = 112). Statistical significance for EFS in multivariate analysis.	Mixed age cohort. No statistical significance for OS in multivariate analysis. No prospective validation.
Genomic gain of 18q	Korshunov et al., 2010	Gain of 18q (aCGH/FISH)	Improved PFS (p = 0.002) and OS (p = 0.001)	Large cohort (n = 122). Statistical significance in univariate analysis.	Mixed age cohort. No statistically significant association in multivariate analysis. No prospective validation.
Genomic loss of 6q25	Monoranu et al., 2008	Presence of 6q25 deletion in anaplastic intracranial tumours (microsatellite markers)	Improved OS (p = 0.013)	Statistically significant in multivariate analysis. Paediatric cohort exclusively analyzed. Support from other work (Korshunov et al., 2010)	Very small cohort size (n = 15). Mixed age cohort. No prospective validation. 6q loss feature of recurrent, not primary paediatric disease on CGH meta-analysis. Conflicting results from other work (Rajaram et al., 2005).
Genomic loss of chromosome 6	Korshunov et al., 2010	Loss of chromosome 6 (aCGH/FISH)	Improved PFS (p = 0.01) and OS (p = 0.003)	Large cohort (n = 122). Statistical significance in univariate analysis. Support from other work (Monoranu et al., 2008)	Mixed age cohort. No statistically significant association in multivariate analysis. No prospective validation. Conflicting results from other work (Rajaram et al., 2005).
Genomic loss of aCGH BAC clones on chromosome 19	Modena et al., 2006	Loss of BAC clones on chromosome 19 (aCGH)	Increased incidence of recurrence (p < 0.0001)	Statistically significant on Fisher's exact test. Paediatric cohort exclusively analyzed.	Small cohort size (n = 24). No survival analysis. No prospective validation.
Homozygous CDKN2A deletion	Korshunov et al., 2010	CDKN2A homozygous deletion (aCGH/FISH)	Worse PFS (p = 0.03 – univariate analysis) and OS (p = 0.02 – multivariate analysis)	Large cohort size (n = 122). Statistical significance for OS in multivariate analysis.	No statistical significance for EFS in multivariate analysis. Other work refutes prognostic association of CDKN2A (Rajaram et al., 2004).

Genomic/Genetic marker	Source	Prognostic Feature	Patient outcome (p-value)	Use as a prognostic marker in paediatric ependymoma	
				Strength	Weakness
LOC374491 gene	Sowar et al., 2006	Over-expression (expression microarray)	Prediction of recurrence ($p = 8.08 \times 10^{-5}$)	Statistically significant on PAM analysis. Paediatric cohort exclusively analyzed.	Very small cohort size (n = 13). No prospective validation. No survival data. Not identified in unsupervised or supervised hierarchical clustering of cohort gene expression.
NFκB2 gene	Sowar et al., 2006	Under-expression (expression microarray)	Prediction of recurrence ($p = 2.05 \times 10^{-5}$)	Statistically significant on PAM analysis. Paediatric cohort exclusively analyzed.	Very small cohort size (n = 13). No prospective validation. No survival data. Not identified in unsupervised or supervised hierarchical clustering of cohort gene expression.
Number and complexity of genomic aberrations	Dyer et al., 2002	'Structural' group of tumours with few and partial chromosomal imbalances (CGH)	Worse OS than 'numerical' tumours ($p = 0.05$) or 'balanced' tumours ($p = 0.02$)	Statistically significant in multivariate analysis. Paediatric cohort exclusively analyzed.	Small cohort size (n = 42). No prospective validation.
PLEK gene	Sowar et al., 2006	Under-expression (expression microarray)	Prediction of recurrence ($p = 6.29 \times 10^{-5}$)	Statistically significant on PAM analysis. Paediatric cohort exclusively analyzed.	Very small cohort size (n = 13). No prospective validation. No survival data. Not identified in unsupervised or supervised hierarchical clustering of cohort gene expression.
Protein 4.1 family	Rajaram et al., 2005	Deletion of 4.1G locus – 6q23 (FISH)	Worse EFS ($p = 0.009$)	Statistically significant in univariate analysis Large cohort (n = 84)	Mixed age cohort. No multivariate analysis. No prospective validation. Contradicting survival findings from other studies (Monoranu et al., 2008, Korshunov et al., 2010).
RAC2 (22q13) TPR (1q25)	Karakoula et al., 2008	Loss of RAC2 in paediatric intracranial tumours (qPCR) Amplification of TPR in paediatric intracranial tumours (qPCR)	Shorter OS ($p = 0.0492$ Shorter OS ($p < 0.0001$)	Statistically significant in multivariate analysis. Paediatric cohort exclusively analyzed.	Small cohort size (n = 47). No prospective validation.

Studies shown are those performed exclusively on paediatric or patients or have included children in the study cohort which have achieved statistical significance (defined as $p \leq 0.05$). Included are each candidate's strength and weakness as a prognostic marker for paediatric ependymoma. For studies analyzing the same markers in sequentially larger cohorts, the study analyzing the largest patient number is shown with reference to the smaller related series. EFS = event-free survival, PFS = progression free survival, OS = overall survival.

1.6 Summary and aims

Paediatric ependymomas remain a clinical management challenge. Improvements in the understanding of ependymoma biology are important to refine the current histological classification systems, define novel prognostic groups and encourage targeted therapeutic development. However, previous studies have often analysed mixed age cohorts, which may contribute to ambiguous results as evidence exists for biological disparity between ependymomas from children and adults. A review of current knowledge regarding the biology of paediatric ependymoma, encompassing several aspects of this chapter, has been published (Kilday, Rahman et al. 2009).

By using the Affymetrix$^®$ 500K SNP array and Illumina$^®$ GoldenGate$^®$ Cancer Panel I methylation array platforms, the initial aim of this study was to determine the scale and nature of genomic and epigenetic aberrations present in paediatric ependymomas collectively and then correlate specific abnormalities with clinical, histopathological and prognostic subgroups. The results of this work are detailed in chapters three, four and five. The SNP array analysis enabled a higher resolution analysis of the ependymoma genome than conventional or array based CGH studies. Whilst other researchers have adopted this technique to examine ependymoma (Johnson, Wright et al. 2010), this was the first to predominantly use neoplastic and constitutional DNA from the same patient to more accurately define tumour-specific genomic imbalances and correlate such aberrations with patient outcome. Moreover, epigenetic array analysis of a paediatric ependymoma cohort has yet to be published.

A final objective was to prospectively validate the prognostic significance of a panel of putative biological markers in paediatric ependymoma, including candidates established from the array analyses. This was performed using immunohistochemistry and FISH on two age-defined tissue microarray (TMA) ependymoma cohorts obtained from children treated uniformly according to the independent clinical trials UKCCLG CNS 1992 04 and SIOP CNS 1999 04. The results obtained are shown in chapter six.

Chapter seven summarises the findings of the entire study and considers potential future work evolving from this research.

CHAPTER 2

MATERIALS AND METHODS

2.1 DNA extraction

2.1.1 Frozen ependymoma cohort

DNA was isolated from tumours which had been snap frozen at surgery and blood samples collected from paediatric ependymoma patients. The DNA was stored at -80ºC for use on either the 500K SNP arrays and/or the methylation arrays. All tumour and blood specimens were collected with patient and parental consent, together with regional and national ethics approval.

In total, 122 snap frozen tumours were collected from 10 United Kingdom Children's Cancer and Leukaemia Group (UK CCLG) registered medical institutions and the Cooperative Human Tissue Network (CHTN) within the United States of America. The UK CCLG centres contributing samples to this study included Nottingham, Birmingham, Cambridge, Bristol, Newcastle, Liverpool, London (Great Ormond Street Hospital), Cardiff, Leeds and Southampton. Forty-five constitutional blood samples also had DNA extracted for SNP analysis from patients contributing tumour specimens (i.e. paired blood samples). The specific number of samples used on each array and the clinical information for each respective patient cohort is presented in the relevant result chapters.

All tumour tissues were reviewed locally at each institution by a neuropathologist in accordance with current WHO criteria (Louis, Ohgaki et al. 2007) to confirm the histological appearance of ependymoma and establish tumour grade. As further verification, all tumours were also centrally reviewed by two neuropathologists at Queen's Medical Centre Histopathology Department, Nottingham (Professor James Lowe and Dr. Keith Robson).

2.1.2 Tumour tissue assessment

To ensure the solid tissue samples to be used for DNA isolation were representative of viable tumour, a small section was cut from the tissue segment to be used for DNA extraction and smeared between two glass microscope slides.

Both tumour tissue cutting and smearing were performed in a recirculating laminar flow preparation station (Labcaire PCR hood 8), which had been pre-cleaned with 100 % (v/v) ethanol and exposed to ultraviolet light for 30 minutes before use to minimise contamination. Each tissue sample for cutting was transferred into the preparation station from -80 ℃ storage in a 1.5 ml vial (Eppendorf, UK) and placed onto a sterile petri-dish (Corning, USA) which was rested on a small box of dry ice to ensure the sample remained frozen, thereby preventing degradation of genetic material. A sterile scalpel blade (Swann-Morton, Sheffield, UK) was used to cut the tissue segments. This had also been cleaned in 100 % (v/v) ethanol before use. The tissue segment cut to assess tumour viability was subsequently smeared between two glass microscope slides which were immediately removed from the station and placed into a coplin jar of Carnoy solution for fixation. The slides were immersed in Harris haematoxylin (SurgiPath, UK) for 30 seconds, washed in water for 10 seconds and then placed in lithium carbonate (Sigma, UK) for 10 seconds. The slides were washed again in water for 10 seconds and subsequently stained in 10 % (v/v) eosin (ProSciTech, AU) for 20 seconds. Following a third wash in water for 10 seconds, the slides were dehydrated in 95 % (v/v) ethanol, then 100 % (v/v) ethanol and finally xylene (Fisher Scientific, UK), each for 10 seconds. Mounting of the slides was performed using coverslips (SLS, UK) and DPX mountant (Surgipath, UK). Stained slides were left to dry on a hotplate. Under neuropathological review, only tissue samples with smears demonstrating viable tumour underwent DNA extraction (Figure 2.1).
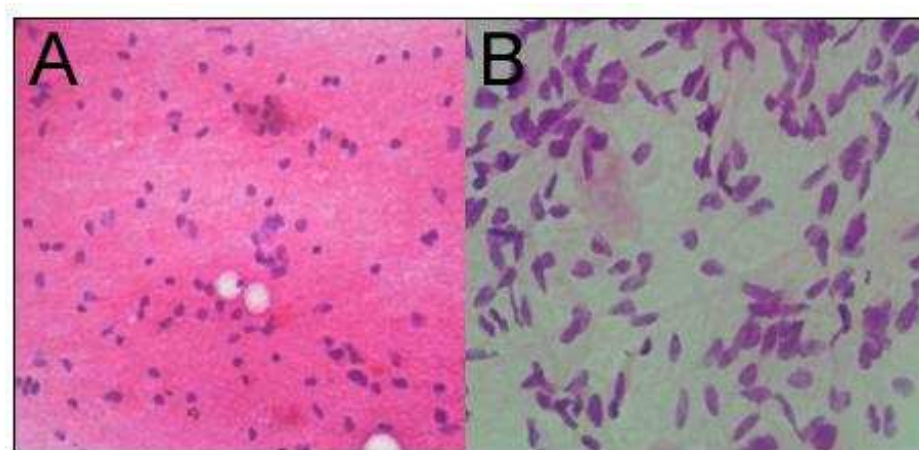
Figure 2.1: Haematoxylin and eosin stained smear of a normal brain specimen (A) and an ependymoma specimen (B). Objective x40.

2.1.3 DNA extraction from frozen tumour tissue

Approximately 10 mg of each tumour sample was homogenized in 1.5 ml vials (Eppendorf, UK) containing 0.5 ml of DNA lysis buffer (see section 2.1) and 0.1 ml of 20 mg/ml proteinase K (Sigma, UK). During homogenisation (performed using the Powergen 125 hand-held homogeniser, Fisher Scientific, UK), vials were surrounded by dry ice to keep samples frozen and prevent degradation of nucleic acid material. After homogenisation, samples were incubated overnight at 37 °C and at 10 g in a Thermomixer (Eppendorf, UK). Following incubation, phase lock gel (PLG) tubes (Eppendorf, UK) were spun at 12,000 g in a microcentrifuge (Sigma, UK) for 30 seconds. A 0.7 ml aliquot of each sample was added directly to the pre-spun PLG tubes in addition to an equal volume (0.7 ml) of phenol/chloroform isoamyl alcohol (Sigma, UK). The contents of each tube were mixed by inversion until a transiently homogenous suspension was formed. Microcentrifugation at 16,000 g for five minutes then separated the aqueous and organic phases. The nucleic-acid-containing aqueous upper phase was transferred by pipette (Gilson, USA) to a fresh 1.5 ml vial. DNA was precipitated on the addition of equal volumes of ice-cold isopropanol followed by vial inversion and DNA pellets were formed by mirocentrifugation at 12,000 g for five minutes. After supernatant removal by pipette, each resulting pellet was twice washed in 0.5 ml of 70 % (v/v) ethanol, mixed by vial inversion and microcentrifuged at 12,000 g for five minutes. Ethanol was then removed from each vial and the pellets were air dried for 20 minutes. Addition of 50 µl ddH$_2$O re-suspended the dried DNA pellets in a 1.5 ml vial.

2.1.4 DNA extraction from blood

The blood specimens used had been stored frozen in anticoagulated vials (Vacutainer, BD, USA). After thawing at room temperature, DNA was extracted from lymphocytes by the transfer of 1 ml of blood (Greiner, UK) to 14 ml of distilled water. The resulting mixture was incubated on ice for five minutes, using osmosis to lyse red cells whilst lymphocytes remained intact. The tubes were then spun in a centrifuge (Sigma 3 – 18K, Sigma, UK) for five minutes at 660 g. Supernatant was discarded by pipette, leaving a pellet which was subsequently transferred to a 1.5 ml vial. To each pellet was added 0.5

ml of DNA lysis buffer and 0.1 ml of 20 mg/ml proteinase K and the extraction protocol was continued from the homogenisation process stated above in section 2.1.3.

2.1.5 DNA quantification and quality control

Each DNA sample was quantified using a spectrophotometer (ND-1000, Nanodrop, UK) which also provided qualitative information for the sample by means of 260 nm/280 nm and 260 nm/230 nm absorbance ratios. A 260/280 ratio of $1.8 - 2.2$ and a 260/230 ratio of $2.0 - 2.2$ indicate intact DNA without contamination. Low 260 / 280 and 260/230 ratios suggest RNA and phenol contamination respectively.

DNA quality was secondly analysed by electrophoresis on a 1 % agarose gel (Figure 2.2). This was made by initially adding 0.5 g agar (Sigma, UK) to 50 ml of 1x TAE buffer in a 250 ml conical flask (Pyrex, SciLabware, Germany). The agar was dissolved by heating in a microwave (Proline, SM18, 750W) for two minutes on full power. After cooling, 0.5 µl of ethidium bromide (Sigma, UK) was added and the mixture was briefly swirled before being poured into a gel mould incorporating sample wells and allowed to set. The gel was added to a gel tank (Flowgen, UK) filled with 1 x TAE buffer. A 1 µl aliquot of sample DNA, combined with 4 µl of gel loading solution (Sigma, UK), was added to each of the sample wells adjacent to one well containing 5 µl of 10 kb hyperladder (Bioline, UK). The gel was electrophoresed at 120 V for 20 minutes. A single band above 10 kb for each sample indicated DNA which was not degraded.
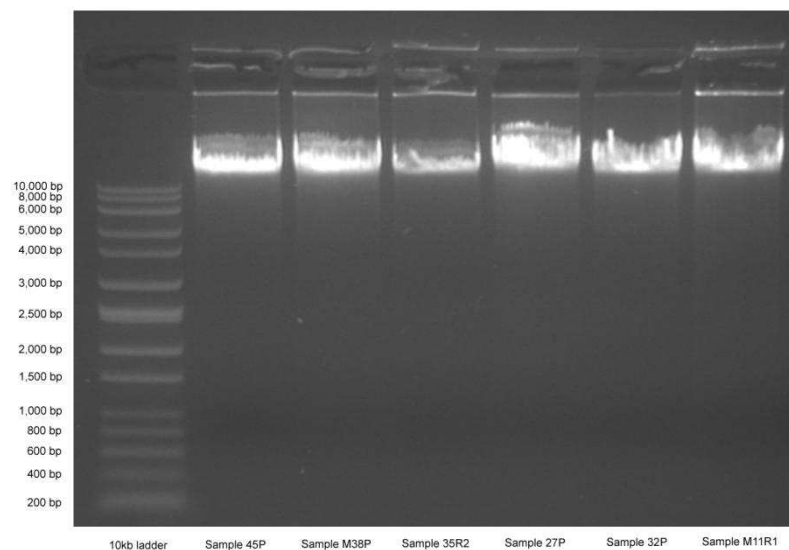
Figure 2.2: 1 % agarose gel of DNA samples extracted from six paediatric ependymomas. The presence of a single band above 10 kb indicated the presence of DNA that was free from degradation.

2.2 The 500K Affymetrix® SNP array

2.2.1 500K SNP array protocol overview

The Genechip® Human Mapping 500K Array Set reliably and accurately detects greater than 500,000 single nucleotide polymorphisms (SNPs) in samples of genomic DNA. Assays were performed at the Hartwell Centre, St. Jude's Research Hospital, Memphis, USA and the Almac Biotechnology Centre, Belfast, Northern Ireland, UK under the under the auspices of Dr. Richard Gilbertson and Dr. Andrea McCulla. Both centres adopted the same protocol, an overview of which is given.

A complete mapping set was comprised of two 250K arrays, each with a corresponding 250K assay kit, incorporating either the restriction enzyme NspI or StyI. This enabled the complete set to achieve maximum genomic coverage and reduce the complexity of the genome by amplifying NspI and StyI fragments using a single PCR primer from only 250 ng DNA per enzyme. The two assays in the mapping set were processed independently to minimise error, thereby generating genotype calls for more than 250,000 SNPs for each array of the set.

A human genomic DNA control of known genotype (Reference Genomic DNA, no. 103) was used as an experimental positive control on each 250K array for every 10 complete chips processed and aided troubleshooting by providing comparative results to those generated by DNA extracted from samples. Fifty SNP probes within each 250K array acted as internal controls by producing genotype calls that intentionally remain consistent irrespective of the DNA hybridising to the array. If the genotype calls generated by these SNP probes post hybridisation proved different to that expected, the result for the entire mapping set was to be discarded. Each NspI and StyI array housed a further 50 SNP probes that acted as controls for cross-checking genotype calls produced from the same sample, thereby confirming identification in cases of confusion.

The five day protocol consisted of 10 main steps (Figure 2.3). DNA was initially digested by the restriction enzyme for each 250K array. End linkers were then ligated to the digested DNA fragments which were amplified by PCR. Following purification,

fragmentation reduced the size of the PCR products (from 250 – 1100 bp to 180 bp), which were subsequently end-labeled prior to hybridisation onto each array. After overnight hybridisation, washing of the arrays removed any unbound DNA fragments. This was followed by Streptavidin Phycoerythrin (SAPE) staining. Each array was then scanned to produce raw signal intensity data.
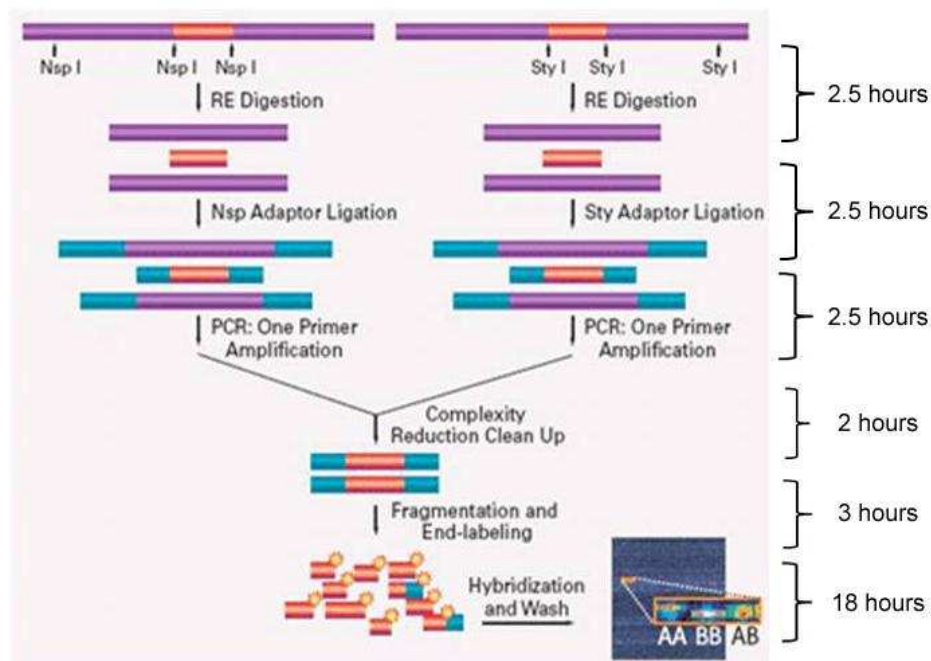
Figure 2.3: Schematic overview of the Affymetrix® 500K assay. Adapted from the Genechip® Human Mapping 500K Set User Guide (Affymetrix 2006).

2.2.2 SNP array scanning and initial data interpretation

The arrays were scanned using the Affymetrix® Genechip® Scanner 3000, controlled by GCOS (Affymetrix® Genechip® Operating Software). Arrays were automatically scanned upon loading. On completion, a .dat file was produced containing fluorescent intensity data for the probes on the arrays. Examination of the .dat file image at this stage was an important quality control checkpoint. The aim was to verify hybridisation of B2 oligo, the main component of an Oligonucleotide Control Reagent (OCR) which was incorporated into each assay's reagent cocktail prior to overnight hybridisation with sample or control DNA. Successful hybridisation of B2 oligo to the array produced characteristic fluorescent designs on the resulting .dat file image including an alternating intensity pattern around the border, a checkerboard pattern at every corner (Figure 2.4A) and the specific array's name at the lower left corner (Figure 2.4B). B2

oligo hybridisation also generated a fluorescent grid over the file image for each array which aided probe alignment. Arrays lacking these features or demonstrating other artefacts were discarded at this stage.
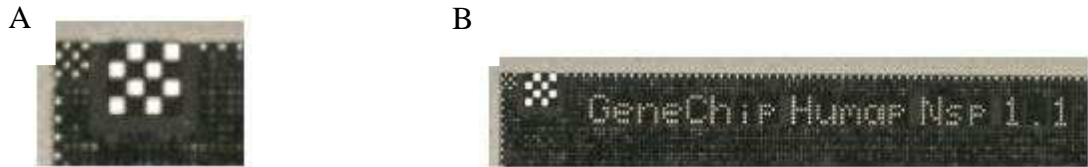
Figure 2.4: Fluorescent designs produced by B2 oligo hybridisation on each 250K array. A checkerboard pattern is produced on the corner of each array (A) while the array name is found in left lower corner of each array (B). Reproduced from the Genechip® Human Mapping 500K Set User Guide (Affymetrix 2006).

2.2.3 SNP array analysis – GTYPE

A fluorescent signal value was generated for every probe on the 250K arrays. Thus, in addition to generating .dat files, GCOS also automatically produced .cel files. Each .cel file contained the results of probe fluorescent intensity from the .dat file in a binary format only compatible for analysis with related Affymetrix® computer programs such as GTYPE (Affymetrix® Genechip® Genotyping Analysis Software). Using the 'batch analysis' function which incorporated an algorithm called BRLMM (Bayesian Robust Linear Model with Malanobis distance classifier), GTYPE incorporated the array data from the .cel files into .chp files which contained information such as probe set perfect match (PM) and mismatch (MM) intensities and genotype calls (AA, AB, BB) (Figure 2.5). This allowed the SNP call rate to be evaluated for each array. The SNP call rate for an array was defined as the percentage of SNPs on a 250K array with a reliable genotype when compared to the total number of SNPs on the array. SNP call rates of 93 % or above were deemed successful, however lower values were possible because of the genomic abnormalities present in tumour DNA. If the calculated SNP call rate was below 85 %, the array was viewed as unreliable and discarded from the analysis.

Mapping Array Report
Report File Name - C:\Program Files\Affymetrix\GeneChip\Affy_Data
Date: 04/23/09 15:38:03

Total number of SNPs:	262264
Total number of QC Probes:	4
Probe array type:	Mapping250K_Nsp

SNP Performance

CEL Data	Called Gender	SNP Call	AA Call	AB Call	BB Call
Sample A_Nsp.brlmm	male	97.66%	39.69%	22.40%	37.91%
Sample B_Nsp.brlmm	female	97.39%	36.59%	28.48%	34.92%

QC Performance

CEL Data	AFFX-5Q-123	AFFX-5Q-456	AFFX-5Q-789	AFFX-5Q-ABC
Sample A_Nsp.brlmm	1092.5	458	2492	3309.5
Sample B_Nsp.brlmm	594	482.5	1678	2093

Shared SNP Patterns

CEL Data	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	SNP10
Sample A_Nsp.brlmm	AB	AA	AA	AB	AA	BB	BB	AB	AA	BB
Sample B_Nsp.brlmm	AA	AB	AB	AB	AA	BB	AA	AA	AB	BB

Figure 2.5: GTYPE batch analysis. Results for the NspI array of two samples (A and B) displaying the SNP call rates (red circle) and the genotype calls for 10/50 control SNP probes used to cross-check genotype calls between the two 250K arrays (red brackets). The genotype calls of these 50 SNP probes should be identical for both arrays performed on each sample.

2.2.4 SNP array analysis – CNAG

The .cel and .chp files for each 250K array were subsequently imported and combined using the computer program CNAG (Copy Number Analyser for Genechip®) (Nannya, Sanada et al. 2005). After initially normalising probe signal intensity to account for factors such as background noise and varying intra and inter-array performance, CNAG incorporates an algorithm that calculates the relative copy number (N) at the ith SNP between two samples from the $\log_2$ ratio of the normalised signal intensities (S):

$$N = \log_2\left[\frac{S(\text{test sample})}{S(\text{reference sample})}\right] \quad \text{for ith SNP}$$

For cases where both tumour and normal DNA were available from the same patient, the constitutional DNA was used as the reference sample. However, for cases where only unpaired analysis was possible, the reference was generated from a pool of the available normal tissue DNA. Further normalisation of this data was then peformed to account for potential artefacts incurred during the PCR process, thereby strengthening the signal to noise ratio obtained. The hidden Markov model was implemented on the combined data sets from both 250K arrays so that each tumour copy number $\log_2$ ratio, relative to either patient-matched constitutional DNA or the pool of unmatched references, was designated to one of seven copy number states:

a) Copy number state 0 = homozygous deletion

b) Copy number state 1 = hemizygous deletion

c) Copy number state 2 = copy neutral (diploid)

d) Copy number state 3 = single copy gain

e) Copy number state 4 = multiple copy gain

f) Copy number state 5 = amplification

g) Copy number state 6 = amplification

Specifically, the '*extract data*' function in CNAG transformed the .chp data for each array sample into .cfh files which were deposited into an 'array data' folder. From the .cfh files, a '*sample manager*' function in CNAG generated .cfs ('self' referenced) and .cfn ('non-self' referenced) data files, using either paired, allele specific or unpaired, non-allele specific reference files respectively. These were stored in a 'results files' folder containing the relative copy number values for the SNPs from each 250K array. Finally, a '*display samples*' function enabled the combination of data from both arrays of a mapping set, while allowing visualisation of the data on a chromosome ideogram (Figure 2.6). In addition, a .txt file containing the copy number results for each sample was generated and stored in an 'array output' folder.
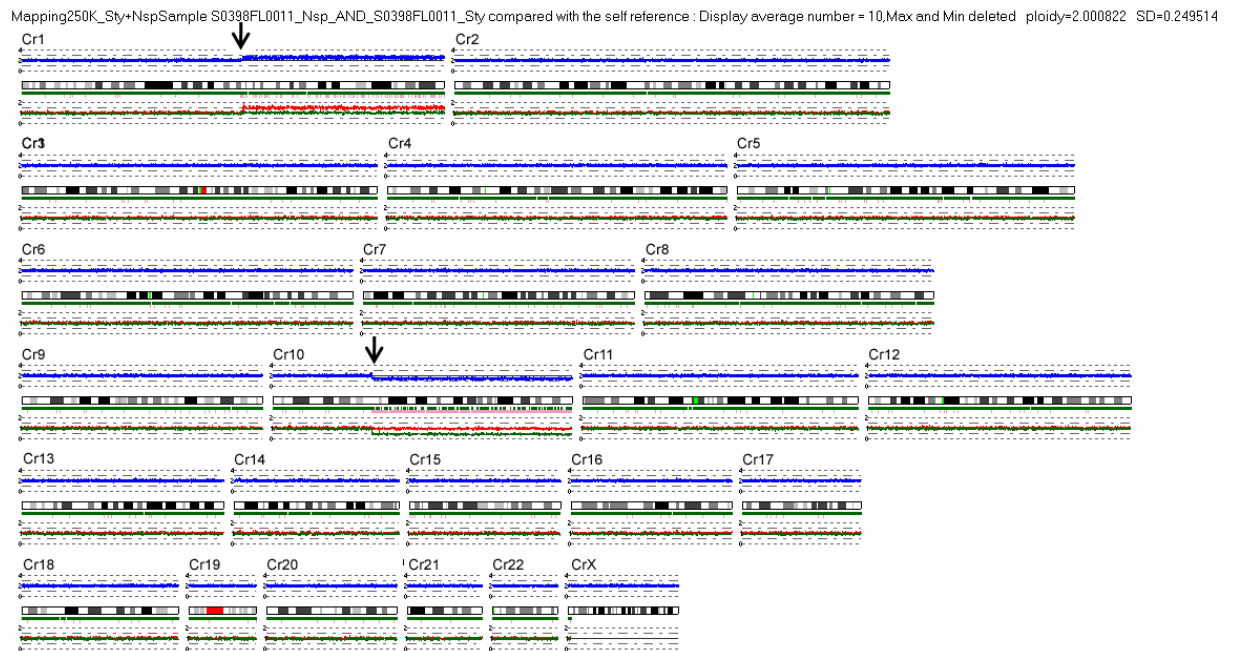
Figure 2.6: CNAG generated chromosome ideogram for the combined 250K NspI and StyI arrays for sample 30P after comparison with patient matched constitutional DNA. Copy number and $\log_2$ ratio line graphs can be seen above and below each chromosomal caricature respectively. In this case, single copy number gain of chromosome 1q and hemizygous deletion of chromosome 10q (arrows) are clearly visible.

65

2.2.5 SNP array analysis – data visualisation software

The ependymoma .txt files generated in CNAG were imported into the data handling and visualisation program Spotfire Decision Site®, in conjunction with a 500K annotation .txt file generated by Affymetrix® (Netaffx file build 07.12.07). This annotation file enabled classification of each probe of the 500K array according to features including probe identification code and number in genomic order, actual physical position within the genome, locality in relation to the nearest gene, the name of such genes and whether the probe location was within a region of known normal copy number variability. The Spotfire Decision Site® program allowed this large dataset to be summarised and viewed, through tables and heatmaps, to identify the copy number value of each 500K probe for each tumour sample. This facilitated the identification of chromosome arm, cytoband and gene copy number alterations either universally or in clinically relevant subgroups, in addition to LOH and aUPD detection. Any heatmap visualisations produced were transformed to .tif (Tag Image File Format) files using Adobe® Photoshop 6.0.

The in-house SNPview program was used as another means of visualising the data generated by CNAG and was devised by Dr Alain Pitiot and Francois Morvillier at the Brain and Body Centre, University of Nottingham, UK in collaboration with the Children's Brain Tumour Research Centre. SNPview was used to visualise the data generated by CNAG on chromosome ideograms encompassing the genome. After importation, the CNAG .txt files were converted to .snp files and incorporated into the chromosome ideograms. As with the Spotfire Decision Site®, visualisations were transformed to .tif files using Adobe® Photoshop 6.0. An overview of the data analysis is shown (Figure 2.7). Details of genomic imbalance analysis methods for chromosome arms, cytobands and genes are given in the relevant results chapters (Chapter 3, section 3.2.1.2 and Chapter 4, section 4.2.1.2). CNAG generated copy number data for all of the probes analysed across the SNP array cohort of 63 paediatric ependymomas, together with a filtered version of the 500K annotation file incorporating selected probe classification features, are included in Appendix 10A.
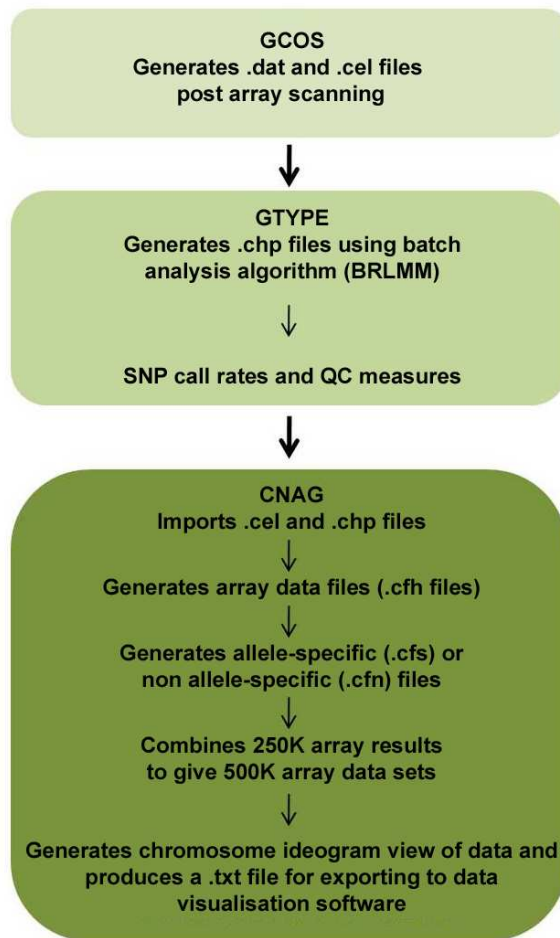
Figure 2.7: SNP array analysis overview using GCOS, GTYPE and CNAG.

2.3 Illumina® GoldenGate® Cancer Panel I assay for methylation

2.3.1 GoldenGate® Cancer Panel I assay protocol overview

Assays were performed according to manufacturer's instructions at the Wellcome Trust Centre for Genetics, Oxford, UK under the guidance of project manager Dr. Joseph Trakolo. An overview of the protocol adopted is given below.

The Illumina® GoldenGate® Cancer Panel I methylation assay is a modification of the SNP genotyping assay (Fan, Oliphant et al. 2003). It enables simultaneous assessment of the methylation status of 1,536 independently selected CpG dinucleotide sites across the genome, underpinned by utilising a bead-based array called the Illumina® Sentrix-Array Matrix® (SAM) (Bibikova, Lin et al. 2006).The SAM configures 96 individual

fibre-optic bundle arrays to match the well spacing of a standard 96 well plate format, allowing this number of samples to be run per assay. Each bundle array is comprised of 50,000 fibre strands, each accommodating a single synthetic bead within a 3 μm chemically etched well at the end of each strand. 1,536 unique bead types are represented with an average 30-fold redundancy. Each bead is covered in hundreds of thousands of copies of covalently attached oligonucleotide probes which are concatamers of a 29 base pair 'address' sequence and a CpG site-specific sequence. After bead insertion into the array, a hybridisation process is used to physically map or 'decode' each bead, using the address sequence. This process validates the performance of the array and is used as a quality control measure prior to use.

Of the 1,536 CpG sites chosen for analysis on the GoldenGate® Cancer Panel I methylation assay, 98 % (1,505/1,536) of the sites are targeted from 807 genes mapping across the genome which have been implicated in the pathogenesis of cancer. These include tumour suppressor genes, oncogenes and genes associated with DNA repair, differentiation, cell cycle control, imprinting and apoptosis. The majority of these genes (576, 71.4 %) contain two or more CpG sites interrogated by the methylation assay. The remaining 31 CpG sites analysed by the assay are designed as internal quality bead controls for each experiment performed. Although several CpG sites are present within a gene, particularly within CpG islands (Takai and Jones 2002) , the specific CpG sites chosen for the GoldenGate® Cancer Panel I methylation assay by Illumina® are those allowing robust assay design (i.e. a design score above a pre-determined threshold). Other CpG sites in close proximity to the CpG site of interest are assumed to have the same methylation status as that of the target site, based on previous evidence (Rakyan, Hildmann et al. 2004; Eckhardt, Lewin et al. 2006).

The four day protocol consisted of seven main steps (Figure 2.8).
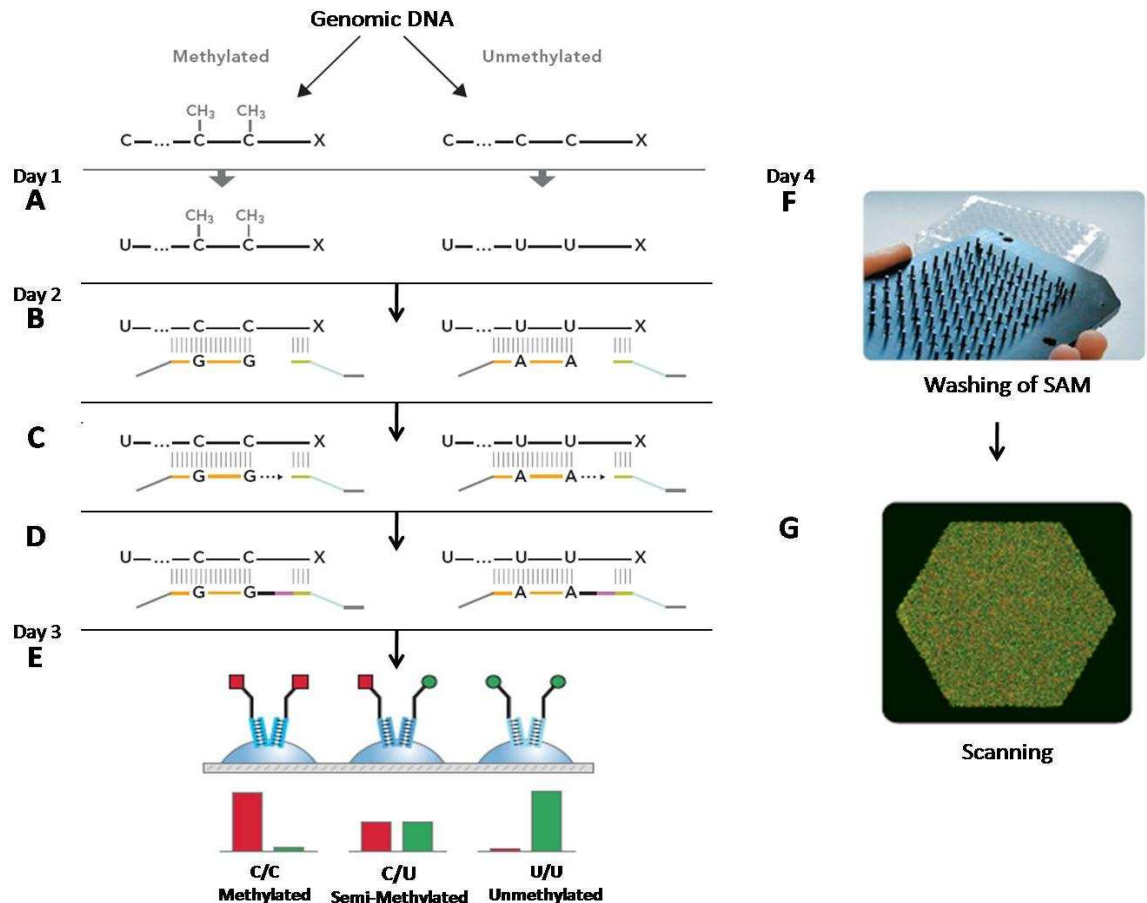
Figure 2.8: Schematic overview of the Illumina® GoldenGate® Cancer Panel I assay for methylation. This chronologically illustrates DNA bisulphite conversion (A), oligonucleotide annealing, extension and ligation (B – D), product labeling and bead array hybridization (E), SAM washing (F) and finally array scanning (G). See text for a more detailed explanation of the process. Figures B – D demonstrate allele-specific oligonucleotides (ASOs) in gold, while locus specific oligonucleotides (LSOs) are shown in green. Ligation of extended ASOs to their corresponding LSOs is designated purple in Figure D. In Figure E, the Illumina® 'address code' is coloured blue. Modified from Illumina® SNP Genotyping –the Illumina® Goldengate® Assay Workflow (Illumina 2006).

Five hundred nanograms of genomic DNA underwent bisulphite conversion. An aliquot of the converted DNA, (corresponding to 250 ng of original genomic DNA) was then used on the GoldenGate® assay. Following bisulphite treatment, unmethylated cytosines (C) were converted to uracils (U), while methylated cytosines remain unchanged (Figure 2.8A). Due to the similarity in hybridisation properties between uracil and thymine (T), the methylation status of a particular cytosine was evaluated through the use of a C/T polymorphism genotyping assay. For each CpG target site, two allele-specific oligonucleotide (ASO) probes and a locus specific oligonucleotide (LSO) probe were designed such that an ASO – LSO pair corresponding to either the methylated or unmethylated state of each CpG site existed (Figure 2.8B). All of the oligonucleotides were pooled and allowed to anneal to the target DNA sequence, with all loci being assayed simultaneously. Extension occurred from the matched ASO

69

towards the corresponding LSO followed by ASO – LSO ligation (Figure 2.8C and D). The ligated products were then amplified by PCR using Cy3- (green) and Cy5- (red) fluorescently labelled universal primers. The resulting fluorescent products were subsequently hybridised to the SAM, through the decoding 'address' sequence on the bead array, which hybridised to a complementary sequence residing within the LSO sequence of each product (Figure 2.8E). The signal intensity ratio between the Cy5- and Cy3- fluors distinguished methylated (red) and unmethylated (green) loci.

2.3.2 Methylation array scanning

Following hybridisation and washing, the Illumina[®] Beadarray[TM] Reader (Illumina, San Diego, USA) scanned each array and produced an intensity data file (.idat) (Figure 2.9). This file contained the fluorescent intensity, location and identification data for every bead of the 96 arrays, together with a .tif image of the signal intensities for each array. The information contained within each ependymoma sample's .idat file was subsequently exported for data interpretation and visualisation to both the BeadStudio v3.2 methylation module[TM] (Illumina, San Diego, USA) and the open source software package Bioconductor[®] (Gentleman, Carey et al. 2004), run within the R software environment for statistical analysis. The beadarray program within the Bioconductor[®] package was used via the readillumina application of R, as it provided a framework for analysing data from bead based arrays (Dunning, Smith et al. 2007).
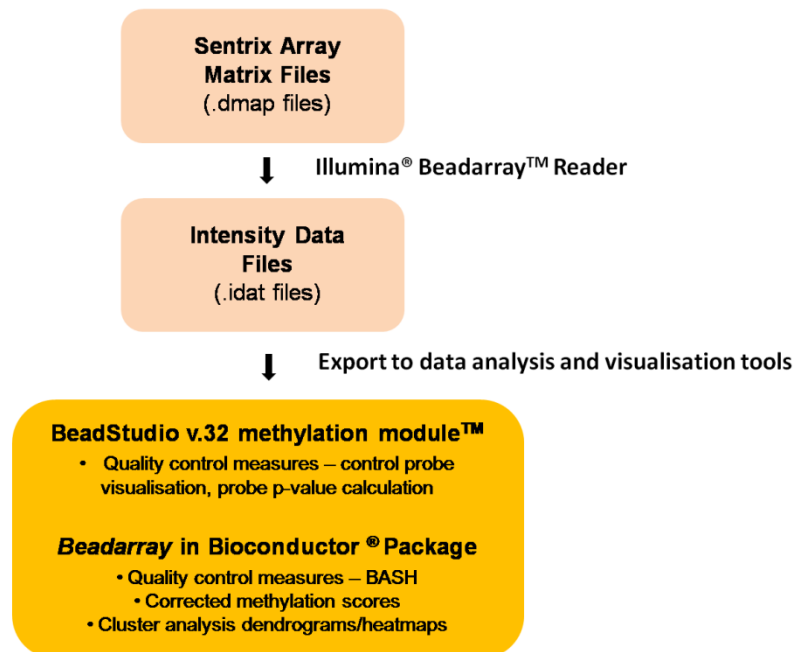
Figure 2.9: Basic workflow for the Illumina[®] GoldenGate[®] Cancer Panel I assay for methylation.

2.3.3 Methylation array quality control

Initially, the .idat array file data exported to the Illumina® Beadstudio methylation module[TM] and the beadarray Bioconductor® program underwent quality assessment.

As stated previously, 31 beads on each array act as internal controls, assessing every step of the assay process. These beads can be categorised into nine groups (Table 2.1). The Illumina® Beadstudio methylation module[TM] produced a control summary graph for these nine sets of control beads across the entire array cohort included in this project (Figure 2.10). This allowed a visual overview of the efficacy of the experiment. In addition, each of the nine sets of control beads were viewed and assessed individually across the cohort to identify samples with aberrant signal intensities that warranted exclusion from further analysis (Figure 2.11).

Table 2.1: The nine groups of internal control probes present on each array of a SAM®.

Bead Control Type	Description
Bisulphite conversion controls	Tests for the presence of unconverted genomic DNA in assay samples.
First hybridisation controls	Test the specificity of annealing ASOs with different Tm to the same DNA locus.
Allele – specific extension controls	Test the extension efficiency of properly matched versus mismatched ASOs.
Extension gap control	Tests the efficiency of extending 15 bases from the 3' end of the ASO to the 5' end of the LSO.
Second hybridisation controls	Tests the hybridisation of single-stranded assay products to address code sequences on the bead array.
Assay intensity controls	Shows average signal intensity for all probes on an array.
Gender – specific methylation controls	Designed against X-linked genes. Cy3- and Cy5- signal should be detected in females; Cy3- only in males.
PCR contamination detection controls	Divided into 4 types. Only one type is added to each oligo pool. When an experiment is run, only one contamination control should have a high signal. If two or more controls have high signal, significant contamination has occurred.
Negative controls – targets sequences lacking CpG dinucleotides	Define the methylation assay background. LSO probes can hybridise to bisulphite converted DNA but ASOs are randomly permutated and should not hybridise to the DNA template. As a result, an amplifiable target should not be formed and the signal should be low.

Reproduced from the Illumina® Genome Studio Methylation Module[TM] v1.0 User Guide (Illumina 2009).
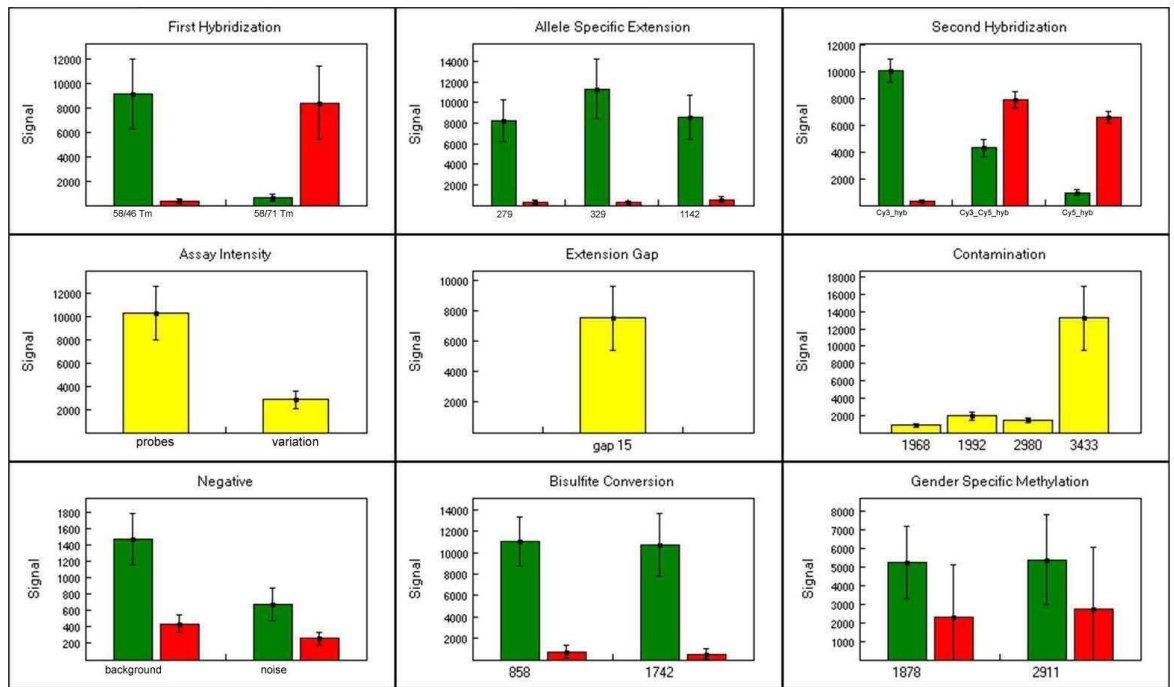
Figure 2.10: Control summary graph of the ependymoma cohort run on the Illumina® Golden Gate® Cancer Panel I. This allows visualisation of the signal intensities (measured in relative fluorescence units or RFU) for the nine groups of internal controls present on each array, summarised across the sample cohort.
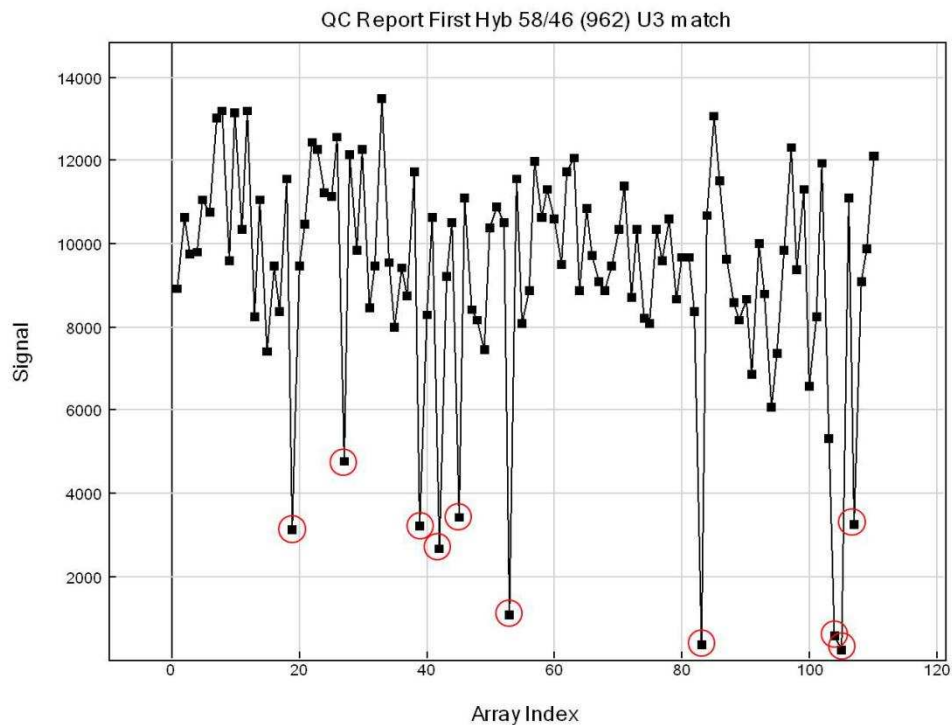
Figure 2.11: Graph demonstrating one of the first hybridisation probe signal intensities for each sample across the ependymoma methylation cohort. The x-axis denotes sample number (array index). The y-axis denotes signal intensity in RFU. Exact intensities for each sample are marked by the small black squares. Outlier samples with aberrant signal intensities are identified by a red ring.

72

The Illumina® Beadstudio methylation module™ also allowed a probe detection p-value to be calculated which characterised the chance of a target bead summary signal value being distinguishable from that of background controls. Significance was defined as a p-value < 0.05. Summary signal values were calculated by averaging the fluorescent intensity of all replicate beads for a bead type on an array. This was performed by the Illumina® default method, where outlying beads with a summary signal value that was more than three absolute deviations from the median (MADs) for that bead type were excluded. The background intensity value for each array was calculated by averaging the summary signals of the in-built negative control beads, which were designed to be thermodynamically equivalent to the target probes but lack a specific target in the transcriptome. Tumour samples were excluded from further analysis if more than 1 % of the 1,505 target bead probes had a detection probe p-value ≥ 0.05.

Quality assessment of the arrays was also assessed using the beadarray program within the Bioconductor® software package. In addition to using boxplots and image plots to identify and remove outlier arrays (Figure 2.12A – D), spatial artefacts on each array were identified using the Beadarray Subversion of Harshlight (BASH) (Cairns et al 2007). This latter feature is not available through the Illumina® Beadstudio methylation module™. Using the median intensity signals calculated from the replicates of each bead type and incorporating a formulaic knowledge of the direct neighbours of each bead, BASH generated an 'error image' for each of the hexagonal arrays such that significantly large spatial artefacts, defined visually as encompassing greater than 25 % of the array area could be identified. Samples contributing such artefacts were thereby excluded from further analysis (Figure 2.12E).
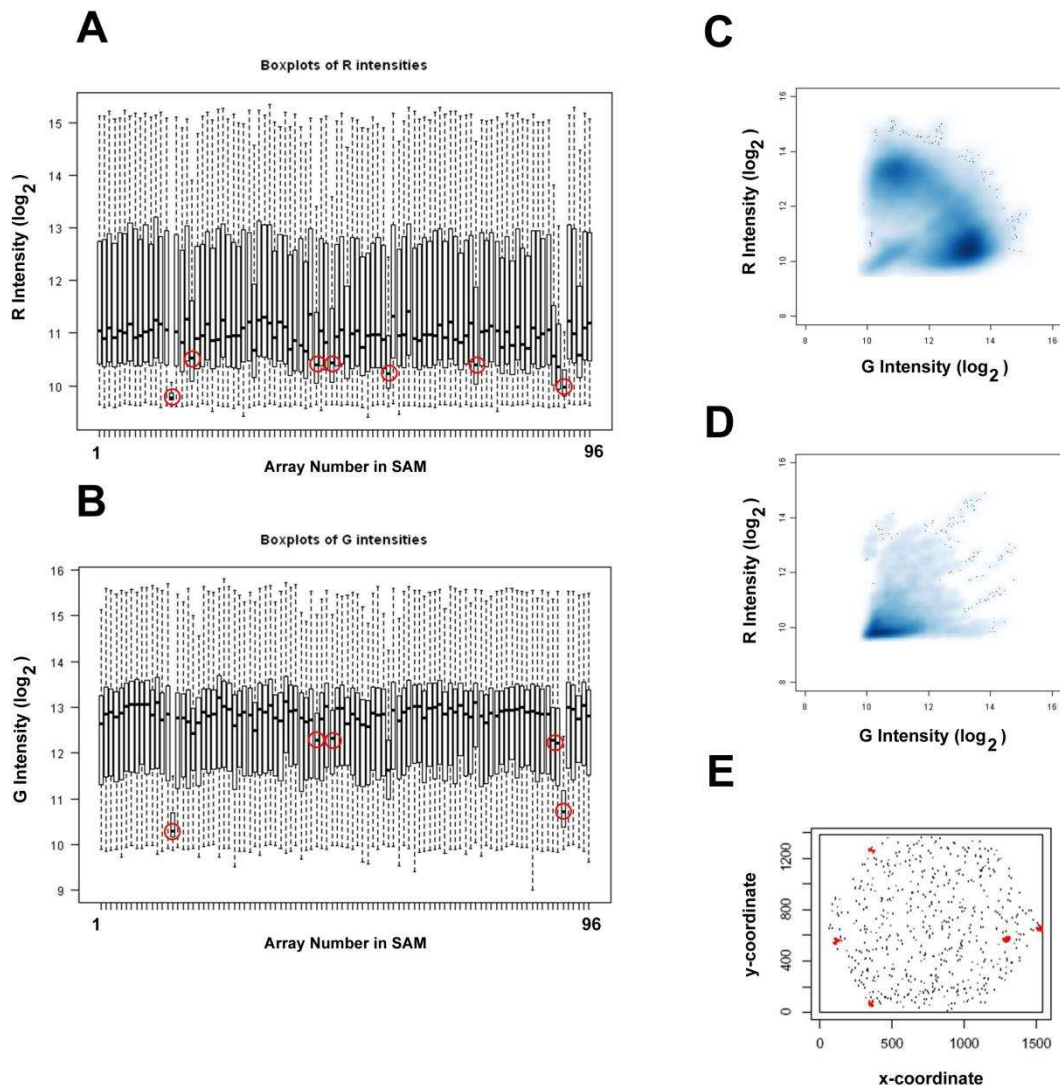
Figure 2.12: Quality assessment of the methylation array using the Bioconductor® beadarray program. Figures A and B are box plots demonstrating the median Cy3- and Cy5- intensity (measured as $\log_2$ (intensity)) of all CpG probes for 96 ependymoma samples. Outliers with aberrant signal intensities are highlighted (red rings). Figures C and D are image plots extrapolated from the box plots demonstrating a suitable (C) and unsuitable (D) array. Median green and red fluorescence (measured as $\log_2$ (intensity)) are denoted on the x and y axes respectively. Ideally, two focal regions should be seen, representing the fluorophore intensity around which most medians fall for the 96 samples on the SAM. Figure E represents a BASH induced image of 1 of the 96 hexagonal arrays in a SAM, highlighting areas of spatial artefact in red. If > 25 % of the array was attributable to artefact, the sample was excluded from further analysis.

2.3.4 Methylation array initial data interpretation

After quality control measures were completed, data analysis was performed, again using the Bioconductor® beadarray program. This incorporated an annotation file (Illumina® Goldengate® Cancer Panel I CpG List 2007) which enabled each of the 1,505 CpG probes to be classified according to features such as probe identification code and location within the genome, locality in relation to CpG islands and the name and function of the encompassing gene.

Initially, bead summary signal values were calculated for both red and green fluorophores across all of the samples analysed. This was carried out using the MADs based Illumina® default method described previously. Background normalisation was performed by subtracting the average signal of the in-built negative control beads from the target probe Cy3- and Cy5- signals. This enabled a corrected methylation score or Beta value (β) to be attributed to each CpG bead probe, based on the following equation:

$$\beta = \frac{\text{Cy5 (corrected red intensity value)}}{\text{Cy3 (corrected green intensity value)} + \text{Cy5 (corrected red intensity value)} + 100}$$

The Beta score could range from zero in the case of completely unmethylated CpG target sites, to one in completely methylated sites. The denominator of the equation is an absolute value to compensate for any negative values resulting from any global background subtraction. A bias of 100 is added to the denominator to standardise the Beta score when both Cy3- and Cy5- values are small.

The computer script enabling array quality assessment and probe background correction using the Bioconductor® beadarray program within the R software environment is included in Appendix 10B. Further downstream analysis including cluster analysis, heatmap generation and differential methylation analysis between clinical subgroups was also performed using the beadarray program within the Bioconductor® package. These methods are described either in the statistics section of this chapter (section 2.7.7) or in the relevant chapters (Chapter 3, section 3.2.2.2 and Chapter 4, section 4.2.2.1).

2.4 Real time quantitative Polymerase Chain Reaction

Real time quantitative PCR (qPCR) was used to validate the genomic aberrations of 15 selected genes from the 500K SNP microarray analysis (Chapter 4, section 4.3.5). Initially, the genomic sequence of the target genes and two control genes which had a normal SNP copy number in the tumours chosen for a particular target gene's validation (ULK4 and AJAP1), were identified using the web-based genomic database Ensembl (http://www.ensembl.org). This program also allowed exons encompassed within regions interrogated by sequential SNPs of interest to be located and selected for primer

design. For each gene, the sequence of the exon of choice was imported into Primer3 (http://frodo.wi.mit.edu/) which generated forward and reverse primer pair sequences with a GC content of 40 – 60 %, such that the optimal product length was between 100 – 150 bp. The sequences of the primer pairs for each candidate exon were subsequently analysed for matches elsewhere in the genome using a BLAST search (http://www.ensembl.org/Multi/blastview), in order to prevent incorrect product amplification (Figure 2.13). All primers were provided by Operon (MWG Eurofins, Germany).
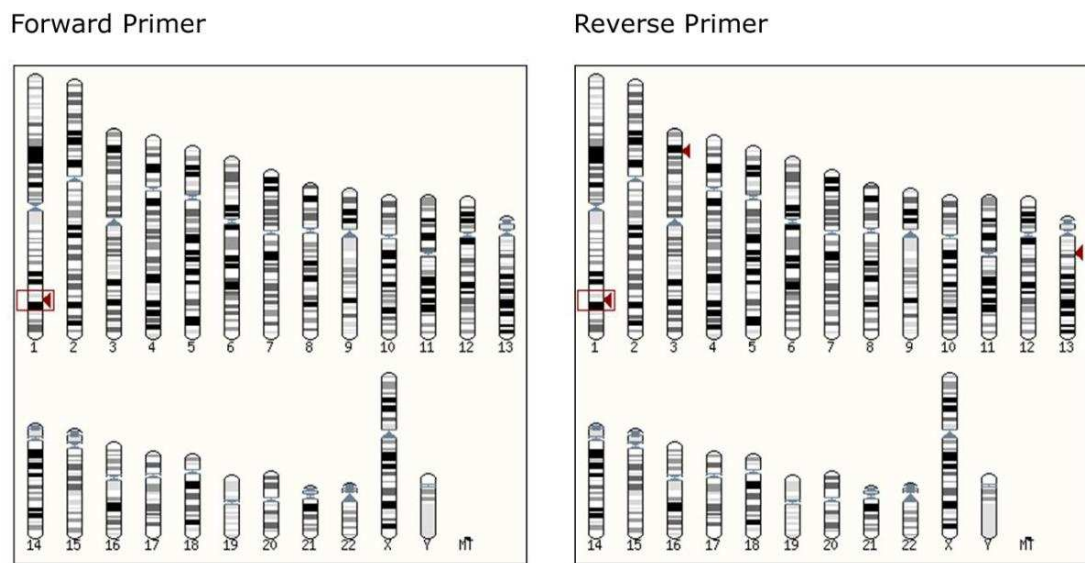
Figure 2.13: An example of a Blast search using the genomic database Ensembl (http://www.ensembl.org/Multi/blastview). This highlights regions across the genome with identical or similar sequences to that of the selected primer pair, to avoid inappropriate product amplification. Results for the primer pair interrogating exon 1 of the NSL1 gene on chromosome 1q32.3 are shown. Red arrows represent regions across the genome with similar sequences to that of the chosen primer. A boxed red arrow signifies an identical sequence match to that of the primer. The forward primer has only one boxed hit across the genome, located at chromosome 1q32.3. The reverse primer has three hits, yet the only boxed hit is also on chromosome 1q32.3. Therefore only one product will be produced by this primer pair, incorporating the appropriate region of interest.

The primer sequences for the target and control genes selected for this study are shown in Table 2.2.

Table 2.2: Primer sequences for 15 genes of interest identified by the 500K SNP array analysis, in addition to those of two control genes, AJAP1 and ULK4.

Gene name	Locus	Exon	Product size (bp)	Forward (5' – 3')	Reverse
AJAP1	1p36.32	3	123	CCCACCACAGAGACTGAGTT	TGTCTTCCGTGTTTGTAGGA
ULK4	3p22.1	37	116	TGTTTCATCCAAGTGCCTGT	GGTCCTCCTTGGATGTCAGT
PRUNE	1q21.2	3	100	CTGCGAGGTGACATTGTCTT	CAGCCTGGTATAATGCATGG
BNIPL	1q21.2	3	107	TGTCTGCCGCCTTTTAATCT	CTTGACCTACCTGCCTGTGA
CHI3L1	1q32.1	4	102	AGGCTCTGCATACAAACTGG	GTACAGAGGAAGCGGTCAAG
NAV1	1q32.1	3	109	CTTTATGAGCCCGAATGGAG	ACTTGGACAGCGTCTTGGAC
NSL1	1q32.3	1	124	ACAGTTCCGACGAAAAATGG	AAAGTCTTCTCGGGGAGTGG
FILIP1	6q14.1	1	104	CCTAATCTCCGAACCCAACA	ATCCCGATCAGCAGAAATTG
FRK	6q22.1	4	104	TTTAAGCGATTGGGATCTGG	ATTGTTCTCGCACCTGGTTT
HOXA5	7p15.2	3	127	TGAAGAAGCCCTGTTCTCGT	TCAAGACAAAGCCATTCAGG
CDKN2A	9p21.3	5	133	CAACGCACCGAATAGTTACG	GAGAATCGAAGCGCTACCTG
TXN	9q31.3	1	105	AGCTCTGTTTGGTGCTTTGG	GTAGCGCGTACCTTGCTCTC
DNAJC25	9q31.3	2	101	CTGCATATGCCCAAGAGGAT	ACAAGCAGGGCAATAAGCTC
COL4A1	13q34	2	113	CCCTCGCTTTGAAAGTGTTC	CCTTTTGTCCCTTCACTCCA
TELO2	16p13.3	18	117	CTTTGTAGCGAGGCCAGGT	CCCAGGCCCAGTGTATTTTT
SEH1L	18p11.21	6	102	GTGGGCCAACAAAGTTTGAA	ATGCTAGCACCGTTCCTGTT
PPARA	22q13.31	2	116	ACTCTGGGTCTTCGGGTTGT	GGAAAGCACCTTCTGAGTCG

2.4.1 Primer Optimisation

The optimal annealing temperatures for all qPCR primers were established using a temperature gradient on a thermal cycler (C1000 thermal cycler incorporated in the CFX96 Real-Time System, Bio-Rad, UK). For each primer set optimisation, a 144 µl PCR megamix was made up in the pre-cleaned recirculating laminar flow preparation station exposed to ultraviolet light for 30 minutes before use. The megamix contained iQ^{TM} Custom SYBR Green Supermix (Bio-Rad, UK), forward primer, reverse primer and double distilled water. The megamix was vortexed (Whirlmixer, Fisher, UK) for one minute, then spun at 16,000 g for 20 seconds in a microcentrifuge. Finally, 24 µl of the megamix was aliquoted into five wells of an unskirted white 96-well PCR plate (Bio-Rad, UK). A surplus 24 µl remained available in case of error. Four temperatures per primer set were selected (ranging between 56 – 62 ℃). A 1 µl aliquot of control DNA (Promega, UK) was added to four of the five wells containing the megamix. A 1 µl aliquot of ddH$_2$O was added to the fifth well as a negative control. The contents of each PCR reaction are shown in Table 2.3.

Table 2.3: Primer PCR reaction set up

Reagent	Volume (µl)
iQ^{TM} Custom SYBR Green Supermix (Bio-Rad, UK) – 2 x Reaction buffer with dNTPs, iTaq DNA polymerase, 6 mM $MgCl_2$, SYBR Green I, fluoroscein and stabilisers	12.5
Forward primer	2.5 (0.5 µM)
Reverse primer	2.5 (0.5 µM)
ddH$_2$O	6.5
DNA (10 ng/µl) or ddH$_2$0	1

The total reagent volume in each well was 25 µl. The initial megamix volume was 144 µl and comprised of 75 µl of iQ^{TM} Custom SYBR Green Supermix, 15 µl of forward primer, 15 µl of reverse primer and 39 µl of double distilled water. dNTPs = deoxyribonucleotide triphosphates, $MgCl_2$ = magnesium chloride, ddH$_2$0 = double distilled water. Note: the primer pairs for DNAJC25 and NSL1 worked effectively at a lower concentration of 0.1 µM.

A temperature gradient program was run under the conditions shown in Table 2.4.

Table 2.4: Temperature gradient PCR program

No. of cycles	Duration	Temperature (^{o}C)
1	10 minutes	95
40	30 seconds	95
	1 minute	56 – 62
	1 minute	72

PCR products were electrophoresed on a 2 % (w/v) agarose gel (1 g agar, 50 ml TAE buffer and 0.5 µl ethidium bromide). The brightest band seen at the expected product size led to the identification that 57 or 58 °C was the optimal annealing temperature for the primers used (Chapter 4, Table 4.21). Every target gene primer set used had an annealing temperature that was suitable for either control gene primer pair.

2.4.2 Primer efficiencies

The efficiency of a primer incorporates both the amount of product which is made following a given number of cycles and the rate at which the PCR reaction occurs. An efficiency of 100 % means that doubling of the target occurs with each PCR cycle. Assessment of primer efficiency is important since normalisation of the target gene against a control gene with a known normal copy number in both tumour and blood is recommended for quantitative assessment (Pfaffl 2001). Two primer sets are required regardless of tissue type, which is a potential source of variability in quantification. While the commonly used ΔΔCt equation for quantification always assumes 100 %

primer efficiency, the Pfaffl equation incorporates individual calculated primer efficiencies in order to provide a more accurate yet robust quantitative analysis. Utilising the Pfaffl equation, primer efficiencies were assessed in this study. Serial dilution stocks of control DNA containing 10 ng/µl, 5 ng/µl, 2.5 ng/µl, 1.25 ng/µl and 0.625 ng/µl were prepared for use. For each primer efficiency assessment, a PCR megamix was prepared as described in section 2.5.1. This time, 24 µl of the megamix was aliquoted into 18 wells of an unskirted white 96-well PCR plate. For each DNA concentration, PCR reactions were performed in triplicate in order to reduce pipetting error. Therefore 1 µl from each of the five concentrations was placed into three separate wells on the plate, with 1 µl of double distilled water being added to the remaining three wells as a negative control. The selected PCR program incorporated each primer's optimal annealing temperature and a final melting curve reaction for confirmation of product specificity by the dissociation pattern produced (Table 2.5).

Table 2.5: Primer efficiency PCR program

No. of cycles	Duration	Temperature (°C)
1	10 min	95
40	30 sec	95
	1 min	Annealing temp
	1 min	72
Melt Curve	Increments of 5°C every 5 seconds	Annealing temp to 95°C

Each PCR reaction generated a C(t) value which represents the number of cycles at which reaction fluorescence exceeds the background fluorescence threshold (Figure 2.14).

Since each reaction was performed in triplicate, a mean C(t) for the primer set was calculated at each DNA concentration. Using Microsoft Excel 2007 (Microsoft, USA), a scatter plot of log DNA concentration versus mean C(t) was produced for each primer set, such that a linear regression line could be generated (Table 2.6 and Figure 2.15).
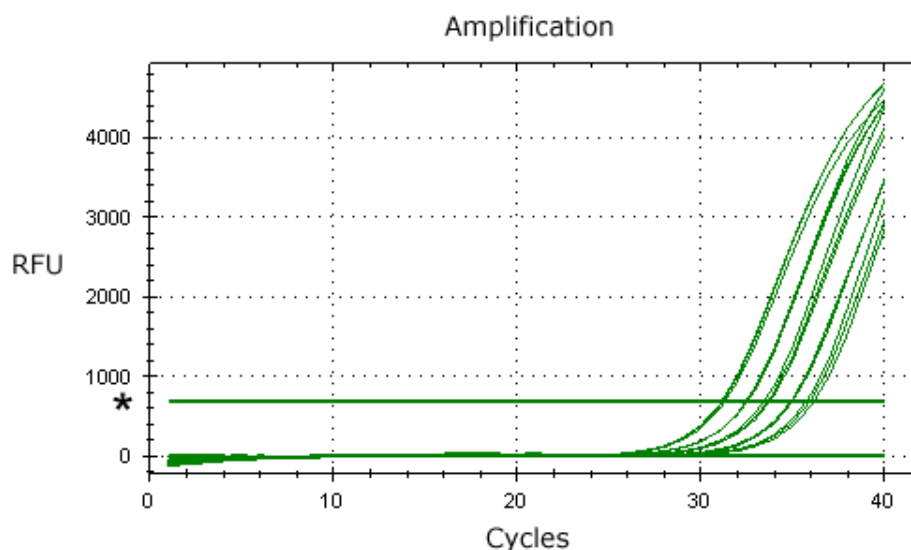
Figure 2.14: Amplification plot generated by the CFX-96 real time system to allow calculation of primer efficiencies. The results generated by using the DNAJC25 gene primer set on serial dilutions of control DNA is shown as an example. A fluorescence threshold for SYBR green was set by the CFX-96 computer program to a baseline reading of 750 relative fluorescence units (RFU) (asterisk). Each curve represents the amount of SBYR green fluorescence produced per PCR cycle and consequently the amount of PCR product made from each DNA concentration. The C(t) value for each reaction was the cycle value at which the reaction fluorescence exceeded the fluorescence threshold value.

Table 2.6: An example of mean C(t) calculation for a primer pair from serial dilutions of DNA.

DNA concentration (ng/µl)	Log DNA concentration	DNAJC25 C(t)	Mean DNAJC25 C(t)
10	1	24.91	24.93
		24.9	
		24.99	
5	0.7	25.93	25.93
		25.98	
		25.87	
2.5	0.4	26.88	26.78
		26.84	
		26.62	
1.25	0.1	28.05	27.97
		27.97	
		27.89	
0.625	-0.2	28.7	28.94
		29.09	
		29.05	

The DNAJC25 primer set is shown as an example.

DNAJC25 primer pair

$y = -3.353x + 28.25$
$R^2 = 0.997$
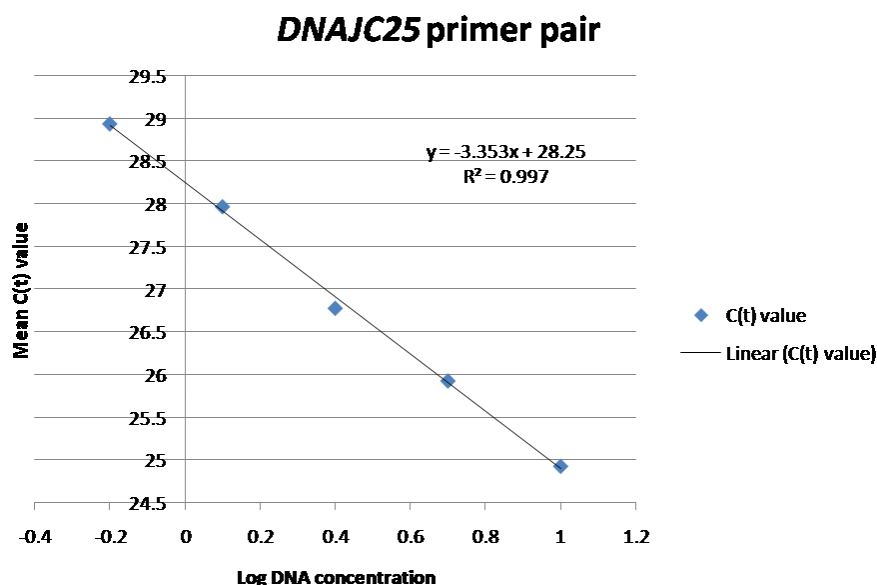
Figure 2.15: Scatter plot of mean C(t) values for the DNAJC25 primer pair at serially diluted concentrations of DNA template. The log of DNA concentration is plotted against the mean C(t) values obtained for each reaction. This generates a linear regression line, the slope of which can be calculated from the equation y = mx+c. In this case the gradient = -3.353. The goodness of fit measure (R^2) in this case was 0.997. In all primer sets used, this value was 0.96 or higher.

The gradient of the resulting line was incorporated into the following equation to allow calculation of primer efficiency as a percentage (Pfaffl 2001):

Primer efficiency $\qquad$ **Ef = $(10^{(-1/Slope)})$ -1**

Example:

$\qquad$ DNAJC25 $\qquad\qquad$ Ef = $(10^{(-1/-3.353)})$ -1

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ = 0.99

$\qquad$ To obtain percentage, Ef x 100 $\quad$ = 99 %

Before inclusion in the Pfaffl equation, each Ef value was converted into an E number signifying the number of copies that would be produced by the primer set per PCR cycle. This was achieved by adding one to the Ef value. For instance, a primer with an efficiency of 100 % (Ef = 1) would have a final E value of two, generating two copies per PCR cycle. For the DNAJC25 example:

$\qquad$ DNAJC25 $\qquad\qquad\qquad$ E = 1.99

Primer efficiencies between 90 and 110 % were deemed satisfactory. The efficiencies of the primer sets used for SNP microarray validation are shown in Table 4.21 (Chapter 4, section 4.3.5), all achieved at an annealing temperature of 57 or 58 °C as stated above.

2.4.3 PCR reactions on tumour and blood samples

PCR was subsequently performed for the target and matched control genes on DNA from paediatric ependymomas already run on the 500K SNP microarray and on patient matched constitutional DNA where appropriate and possible. The PCR megamixes and optimal reaction conditions were as described in section 2.4.2. The precise number of samples used varied according to the target gene being analysed (detailed in Chapter 4, section 4.3.5).

For intra-assay accuracy, individual tumour and blood PCR was again performed in triplicate. For those genes demonstrating a significant correlation between SNP array copy number and qPCR derived copy number (as explained in sections 2.4.4 and 2.7.3), inter-assay reproducibility was assessed by replicating the entire PCR twice, or thrice if a difference of greater than one was found between the two previous qPCR copy number values. To avoid specimen depletion, genes not demonstrating copy number correlation were not subjected to repeated PCR experiments.

2.4.4 Quantification of target gene copy number in tumour samples

The established E values for both the target and reference gene primers were used in the Pfaffl equation (Pfaffl 2001), alongside the calculated mean C(t) values for both target and control genes in ependymoma and blood in order to identify the copy number ratio (R) for the particular target gene as follows:

The Pfaffl equation:

$$R = \frac{E_{target}^{\Delta Ct \, (Ct \, target \, gene \, blood \, - \, Ct \, target \, gene \, tumour)}}{E_{control}^{\Delta Ct \, (Ct \, control \, gene \, blood \, - \, Ct \, control \, gene \, tumour)}}$$

In the equation, R is the relative ratio of the difference in fluorescent signal between the target gene in the blood and tumour samples when compared to the control gene. Therefore if R = 1, this suggested no quantifiable difference in fluorescence intensity between the target gene in tumour and blood DNA relative to the difference identified in

the control gene, the latter being selected as it had demonstrated a diploid copy number in both tissue samples on the 500K SNP microarray. Since normal copy number is two, the qPCR derived copy number for each tumour sample was calculated by doubling the resulting R value and rounding to the nearest whole number. Thus a qPCR copy number of $0 < 0.5$ = homozygous loss, ≥ 0.5 to < 1.5 = hemizygous loss, ≥ 1.5 to < 2.5 = normal (diploid), ≥ 2.5 to < 6 = genomic gain and ≥ 6 = amplification.

2.5 Fluorescent in situ hybridisation

Fluorescent in situ hybridisation (FISH) was performed on TMAs of formalin fixed paraffin embedded (FFPE) tumour to validate the findings of chromosome 1q gain in selected ependymomas from the SNP array analysis (Chapter 3, section 3.3.6). The technique was also used to establish a prognostic role for gain of chromosome 1q25 in paediatric ependymoma (Chapter 6, section 6.3.4.1).

The TMAs were constructed (using the manual tissue arrayer I, Beecher, USA) by coring viable and representative tumour regions previously identified from FFPE ependymoma tissue blocks by a neuropathologist (Prof. James Lowe). Sections of 4 µm thickness were obtained for the microarray and cores were 0.6 mm in diameter. Where possible, triplicate cores per tumour sample were taken. In some cases tumour heterogeneity had resulted in two different sets of triplicate cores being constructed for a particular sample. These were labelled triplicate 'a' and triplicate 'b'.

Detection of 1q25 was performed using a commercial LSI 1p36/LSI 1q25 dual colour probe (Vysis, USA). The Spectrum Orange 1p36 probe hybridises to a region extending from the SHGC57243 locus to a point telomeric to the EGFL3 locus. The Spectrum Green 1q25 probe hybridises to a region extending from a point telomeric to the ABL2 locus to a point upstream of the SHGC-1322 locus. FFPE normal tonsil tissue was assessed in conjunction with the TMA slides as a control to demonstrate that, in normal cells, diploid signals for both LSI probes should be predominantly observed.

Prior to use, each TMA slide, together with a positive control slide, were baked in an oven (Windsor incubator, Lamb, UK) at 60 ºC overnight. The slides were then

deparaffinised in xylene for 15 minutes and dehydrated in 100 % (v/v) ethanol for a further 15 minutes. After this, the slides were placed in a coplin jar of 10 % (v/v) neutral buffered formalin (Sigma, UK) for one hour. While the slides remained in fixative, a plastic container filled with sodium citrate was heated to boiling point in a steamer (Tefal, UK). On removal from the formalin, the slides were washed in warm tap water for five minutes before being placed into the boiling sodium citrate for one hour. After this, the slides were transferred back into warm running tap water for five minutes before being immersed in a coplin jar containing pepsin solution at room temperature for 30 minutes. The slides were washed in cold running tap water for five minutes before being rinsed in distilled water for one minute at room temperature and then air dried. Slides were placed in Carnoy fixative for 30 minutes and again air dried. Under reduced lighting, the FISH probe (Vysis, UK) was applied to the tissues on each slide. Between 10 µl and 20 µl was applied, depending on the size of the tonsil tissue or TMA, to ensure probe coverage of the slide. A coverslip was placed over each slide. The slides were then placed in a hybridiser (Dako, UK) which was set to conditions of 90 ℃ for 12 minutes followed by 37 ℃ overnight. The following day under reduced light, the slides were placed in a coplin jar of 2 x SSC for 30 minutes. The coverslips were then removed and the slides were placed in 4 M urea in 2 x SSC for two minutes. After this, the slides were washed for one minute in 2 x SSC, one minute in distilled water and finally air dried. 20 µl of DAPI (750 µg/ml) was applied to each slide followed by coverslip placement.

Visualisation of results was performed using a Nikon Eclipse 90i microscope (Nikon, Japan) and a Hamamatsu ORCA-ER camera (Hamamatsu, Japan). Resulting images were acquired using Improvision Volocity 5.0 software (Perken Elmer, USA). Images were taken at x 40 objective. Three filters were used: DAPI (Ex 340 – 380 nm), Spectrum Green (Ex 497 – 509 nm), and Texas Red (Ex 540 – 580 nm). As the camera was black and white, a colour was applied to each image based on the wavelength of the filter: DAPI at 435 nm (blue), Spectrum Green at 535 nm (green), and Texas Red at 588 nm (orange/red). Maximum light exposure levels were set using the tonsil control tissue. Where possible, 3 images were taken per tumour core. The FISH scoring regimen used was based on previous studies using FISH on paraffin embedded tumour tissue microarrays (Bubendorf, Kononen et al. 1999; Neben, Korshunov et al. 2004; Korshunov, Witt et al. 2010). Only samples where signals were obtained for greater

than 80 % of available nuclei were analysed. A minimum of 100 nuclei were counted per tumour sample, except for one case (9904 – 39P) where only 62 nuclei were assessed due to the cellularity and quantity of tissue available. The number of green and red probe signals visualised for each nucleus was recorded. Chromosome 1q25 gain was defined to include widespread or focal gain. Widespread gain was observed when at least 15 % of the counted nuclei per sample contained three or more copies of the 1q25 probe, regardless of the 1p36 probe. Focal 1q25 gain was demonstrated when at least 15 % of the nuclei counted within one tumour core contained 3 or more copies of the probe. During prognostic marker assessment, a second scorer (B.M.) was used to verify initial scores blindly in a proportion of cases.

2.6 Immunohistochemistry

Immunohistochemistry (IHC) was performed to establish the expression of particular proteins encoded by putative prognostic gene candidates resulting from the SNP array analysis. In addition, IHC was used to examine a panel of potential ependymoma prognostic markers suggested from the literature (Bennetto, Foreman et al. 1998; Figarella-Branger, Civatte et al. 2000; Gilbertson, Bentley et al. 2002; Zamecnik, Snuderl et al. 2003; Ridley, Rahman et al. 2008; Puget, Grill et al. 2009) (Chapter 6).

Immunohistochemistry was carried out using the Dako Chemate Envision Antigen Detection kit (Dako, UK), except for Tenascin-C when the Vector ABC reagent kit was used (Vector, UK). The optimal conditions for each antibody used are shown in Table 2.7.

Table 2.7: The five antibodies used in the immunohistochemical analysis.

Name	Supplier	Host	Clonality	Dilution	Incubation conditions	Staining pattern	Positive control
Nucleolin	Abcam® (ab13541)	Mouse	Monoclonal	1:400	Overnight at 4 °C	Nuclear	Tonsil
Tenascin-C (E-9)	Santa-Cruz Biotec® (sc-25328)	Mouse	Monoclonal	1:50	1 hour at room temperature	Extracellular matrix / Cytoplasmic	Epidermoid carcinoma
Ki-67	DAKO® (M7240)	Mouse	Monoclonal	1:50	1 hour at room temperature	Nuclear	Tonsil
PRUNE	Sigma-Aldrich® (HPA028411)	Rabbit	Monoclonal	1:800	1 hour at room temperature	Cytoplasmic/ Membranous	Breast carcinoma stromal tissue
NAV1	Sigma-Aldrich® (HPA018127)	Rabbit	Monoclonal	1:350	1 hour at room temperature	Cytoplasmic/ Membranous	Small intestine

Each TMA slide was deparaffinised in xylene for 15 minutes, then hydrated by immersion in 100 % (v/v) ethanol for five minutes, 95 % (v/v) ethanol for four minutes and finally running warm tap water for 10 minutes. Antigen retrieval for Nucleolin and Ki-67 IHC was performed using a pressure cooker. Approximately 1.5 litres of sodium citrate (pH 6.0) was placed in a pressure cooker (Kern, China) and heated until boiling on a digital hot-plate (Cimarec, Cole-Parmer, UK). At this point the slides were added and the cooker lid was sealed until full pressure was attained. The slides were treated at full pressure for 1 minute then allowed to cool in the sodium citrate for 20 minutes. Antigen retrieval for Tenascin-C, PRUNE and NAV1 IHC was performed using pre-boiled sodium citrate (pH 6.0) in a steamer (Tefal, UK). These slides were treated for 30 minutes and then cooled in the sodium citrate for a further 45 minutes.

Following antigen retrieval techniques, a protein blocking agent was applied to each slide for 15 minutes. Normal goat serum was used for all cases except Tenascin-C IHC where normal horse serum diluted to 1:100 (v/v) in PBS solution was used (Vector, UK). After a five minute wash with PBS, endogenous peroxidase blocking solution (Dako, UK) was then applied to the slides for five minutes, followed by another five minute PBS wash. Since the Tenascin-C IHC made use of an avidin-biotin complex (ABC) reaction to detect the primary antibody, a further avidin/biotin block was applied at this stage (Vector, UK). Slides were immersed in avidin solution for 15 minutes, dipped in PBS solution then incubated in biotin solution for a further 15 minutes.

Once blocking was completed, 100 µl of primary antibody (or enough to cover the TMA) was applied to each slide, followed by incubation (see Table 2.7). Negative controls for each antibody tested had antibody diluent (Dako, UK) applied instead of primary antibody. After incubation, the slides were immersed in PBS for five minutes then secondary antibody was applied to each slide for 30 minutes, followed by a further wash in PBS for five minutes. The secondary antibody used was HRP anti-rabbit/mouse (Dako, UK) in all cases except for Tenascin-C where 100 µl of biotinylated universal antibody anti-rabbit/mouse (Vector, UK), diluted in 5 ml of normal horse serum, was used. At this stage during the Tenascin-C IHC experiments, Vectastain avidin-biotin complex reagent (Vectastain-Elite ABC kit, Vector, UK) was subsequently applied for 30 minutes, followed by a five minute PBS wash.

Following secondary antibody application, 100 μl of DAB (20 μl 3-3'-diaminobenzidine and chromogen in 1 ml substrate buffer, (Dako, UK)) was applied for five minutes to each slide, followed by immersion in PBS for five minutes. The slides were then stained in Harris haematoxylin for 10 seconds, washed in water for 10 seconds, then placed in lithium carbonate for 10 seconds. After this, the slides were washed again in water for 10 seconds and subsequently dehydrated in 95 % (v/v) ethanol, then 100 % (v/v) ethanol and finally xylene, each for 10 seconds. Mounting of the slides was performed using coverslips and DPX mountant. The slides were then visualised using an Olympus BX41 light microscope (Olympus, UK) at x 10 and x 40 objectives.

Ki-67 and Nucleolin labelling indices (LIs) were defined as the percentage of tumour cells with immunopositive nuclei divided by the total number assessed. A minimum of 100 nuclei were counted per core. For samples with two sets of triplicate cores, the higher mean LI value for each triplicate was used as the designated LI for that case. For Nucleolin analysis, a 50 % LI threshold was used to designate low and high expression cohorts, while Ki-67 expression was defined as low (≤ 1 %), moderate ($2 - 4$ %) or high (≥ 5 %). This scoring method and the LI thresholds used replicated that of a previous retrospective immunohistochemical analysis of Nucleolin and Ki-67 in paediatric ependymoma (Ridley, Rahman et al. 2008). Tenascin-C staining was extracellular and defined as negative, weak or moderate to strong, based on the area of highest staining, albeit focal or widespread. The expression of NAV1, and PRUNE were cytoplasmic or membranous and graded as negative, weak, moderate or strong respectively. For each putative prognostic marker assessed, repeat analysis was performed blindly by a second scorer on a proportion of the total cases (L.R. for Nucleolin and Ki-67, J.B for NAV1 and PRUNE, F.A. for Tenascin-C).

2.7 Statistics

Statistical analysis was predominantly performed using SPSSv16 (Statistical Package for the Social Sciences, Version 16, Chicago, USA), with the exception of differential methylation analysis between clinical subgroups, which was performed through the R statistical software environment.

2.7.1 Associations of variables in two-way frequency tables

Fisher's exact test was used to determine associations between combinations of clinical and biological variables for a particular patient cohort, in two-way frequency tables. Clinical variables included intracranial ependymoma location within the central nervous system (infratentorial or supratentorial); patient age at diagnosis (above or below three years of age); patient sex (male or female), tumour grade according to centralised pathological WHO classification (Grade II or III); tumour resection status (complete or incomplete) and the primary or recurrent status of each tumour (using first recurrent cohorts unless stated). Biological variables included chromosomal arm imbalances, copy number imbalances at a cytoband and gene level, gain of chromosome 1q25 as detected by FISH and immunohistochemical staining patterns. In each case, a two-tailed significance p-value was generated. Significance was defined as a p-value below 0.05, while a p-value of 0.05 – 0.1 indicated a trend towards significance.

2.7.2 Associations with patient age

As patient age is a continuous variable, a comparison of mean age values between certain clinical and biological categories described above was assessed using either an independent sample t-test with 95 % confidence intervals or a one-way analysis of variance (ANOVA) if more than two subgroups were assessed simultaneously, for instance when comparing spinal, supratentorial and posterior fossa ependymomas. ANOVA post-hoc comparisons were performed using the Tukey HSD test and the effect size of the ANOVA result was calculated as an eta squared value. A value greater than 0.14 represented a large effect (Pallant 2007). Significance was defined as a p-value below 0.05.

The Mann-Whitney test was performed to compare the number of broad genomic imbalances between children with posterior fossa ependymomas who were aged either above or below three years. Again, significance was defined as a p-value below 0.05.

2.7.3 Correlation of SNP array and qPCR results

Spearman's rank correlation coefficient (SRCC) was used to determine whether the CNAG derived SNP array copy number results for 15 selected genes in a set of tumour samples were confirmed by qPCR, as described in section 2.4. A SRCC value (R) of zero indicated no correlation whilst a value of one signified an exact correlation. A value of R greater than 0.65 accompanied by a p-value less than 0.05, defined a positive correlation.

2.7.4 Correlation of methylation array technical replicates

The Kendall-tau coefficient was used to assess the correlation of corrected Beta methylation scores between 2 sets technical replicate samples of DNA analysed on the methylation array, thereby assessing intra-assay variability. The DNA replicates were from high grade glioma samples occupying surplus well space in one of the SAM's used for the ependymoma analysis. A SRCC value (R) of zero indicated no correlation whilst a value of one signified an exact correlation. Significance was defined as a p-value below 0.05.

2.7.5 Kappa measure of agreement

When examining the putative prognostic markers by FISH or immunohistochemistry, the degree of agreement between two independent scorers was assessed using the Kappa measure of agreement. A Kappa value above 0.5 was defined as a moderate agreement, while a value above 0.8 was defined as a very good agreement (Pallant 2007). Significance was defined as a p-value below 0.05.

2.7.6 Survival analysis

To assess the univariate prognostic impact of the clinical and biological variables described above in a particular patient cohort, Kaplan-Meier survival curves were created with significance values established by the log rank test. Event-free survival (EFS) time was calculated from the date of original diagnosis to the time taken for the first event

(recurrence or death) to occur, or the date of last follow-up (censorship) if event free. Overall survival was calculated from the date of diagnosis to the date of death, or to the time of censorship if the patient remained alive. Significance was defined by a p-value below 0.05, while a p-value of 0.05 − 0.1 indicated a trend towards significance. Multivariate prognostic analysis was performed using the Cox proportional hazard regression model to determine patient survival comparisons and factor hazard ratios with 95% confidence intervals. P-values less than 0.05 defined indicated a significant result.

2.7.7 Differential methylation analysis between clinical subgroups

Corrected Beta values were calculated following quality control measures and array background normalisation, as discussed in sections 2.3.3 − 2.3.4. Differentially methylated loci between selected clinical subgroups or certain sample subgroups established from unsupervised methylation cluster analysis (Chapter 3, section 3.3.7) were identified via the R statistical software environment. The clinical subgroups analysed and the results obtained are shown in Chapter 4, section 4.3.4.

Differential analysis was performed using Mann-Whitney U tests with a p-value < 0.05, after Benjamini-Hochberg False Discovery Rate correction for multiple testing (Benjamini, Drai et al. 2001), with an additional filter introduced so that the average change in Beta values between analysed subgroups was greater than 0.34. A computer script was devised for this aspect of the data analysis by Dr. Edward Schwalbe and Dr. Steve Clifford at the Northern Institute for Cancer Research, Newcastle (Appendix 10C).

CHAPTER 3

GENOME-WIDE CHARACTERISATION OF GENOMIC AND EPIGENETIC ABERRATIONS IN PAEDIATRIC EPENDYMOMA

3.1 Introduction

Notwithstanding the evolution of improved neuro-surgical techniques and adjuvant therapeutics, the prognosis for children diagnosed with ependymoma remains inferior in relation to other paediatric brain tumours (Zacharoulis, Levy et al. 2007; Messahel, Ashley et al. 2009; Wright and Gajjar 2009). In order to achieve significant progress in the prediction of both paediatric ependymoma behaviour and subsequent patient outcome, a better understanding of underlying tumour biology through high resolution molecular characterisation is required, in conjunction with improved patient recruitment into large prospective clinical trials.

Although evidence exists for genomic distinctions between certain clinical subgroups of ependymoma (Chapter 1, sections 1.5.2 – 1.5.3), the data is derived from studies which were either performed on relatively small cohorts, included paediatric and adult patients or were of a lower genomic resolution compared to SNP array analysis. High resolution analysis has been performed on this tumour group (Johnson, Wright et al. 2010), yet no SNP microarray study has been performed exclusively on paediatric ependymomas, examining both neoplastic and constitutional DNA from the same patient to identify tumour-specific genomic aberrations and loss of heterozygosity. In addition, no SNP array analysis has ever correlated detected anomalies with patient outcome. This study aimed to address these compounding issues by examining a relatively large series of paediatric ependymomas and relating the genomic imbalances detected with accompanying clinical, pathological and patient survival data.

Similarly, to date there has been no reports of epigenetic microarray work being exclusively performed on paediatric ependymomas, hence this genome-wide methylation analysis of over 1,500 selected CpG dinucleotides on a sizeable tumour cohort was also the first and largest study of its kind.

The dual array based analysis was undertaken to initially establish the broad genomic and methylation characteristics of paediatric ependymomas, prior to a higher resolution assessment detailed in Chapter 4, in order to explore the following hypotheses:

- Ependymomas occurring in different anatomical locations within the CNS are biologically distinct.
- Biological differences exist between ependymomas from younger children (under three years of age) and older children (over three years of age).
- Molecular events in paediatric ependymoma pathogenesis are predictive of clinical and/or prognostic behaviour.

3.2 Materials and methods

3.2.1 500K SNP array analysis of 63 paediatric ependymomas

3.2.1.1 The sample cohort

Following tumour and blood DNA extraction as described in Chapter 2, sections 2.1.3 – 2.1.4, 500K SNP array assays were performed on 63 paediatric ependymomas from 45 patients (42 primary and 21 recurrent tumours) in two collaborative institutions with appropriate facilities. Both centres adopted the same manufacturer's protocol, an overview of which is given in Chapter 2, section 2.2.1. The DNA of 39 tumours from 26 patients, together with matching blood DNA samples were processed at the Hartwell Centre, St. Jude's Research Hospital, Memphis, USA under the auspices of Dr. Richard Gilbertson. Allied to this, the DNA of 24 ependymomas from 19 children, in addition to their blood DNA samples were processed at the Almac Biotechnology Centre, Belfast, Northern Ireland, UK under the guidance of project manager Dr. Andrea McCulla.

SNP array signal intensity data from both centres was processed as outlined in Chapter 2, sections 2.2.2 - 2.2.5. Data on the 39 tumours (but not the blood specimens) processed by the Gilbertson group has subsequently been incorporated into a SNP array study of 204 adult and paediatric ependymomas discussed elsewhere in this work (Johnson, Wright et al. 2010). Initially it was hoped to analyse the .chp files from all 63 paediatric ependymomas against those generated from patient matched blood DNA (45 cases). However, five blood samples were excluded from further analysis as they demonstrated a pattern of copy number aberration deemed artefactual when analysed against all other blood samples in the cohort, causing the percentage of SNPs with a

diploid copy number for these samples to fall below a pre-determined satisfactory threshold of 99 % (Appendix 2). As a consequence, 10 ependymomas matching to these five blood samples (four primary and six recurrent tumours) were individually re-analysed against a pool of the remaining 40 constitutional DNA specimens. This conformed with manufacturer's recommendations that for copy number analysis of unmatched patient samples, a minimum of 30 reference samples should be used (Affymetrix 2007). SNP probes within the X chromosome were thereby not analysed since both genders contributed DNA to the constitutional reference pool required for this small proportion of the cohort.

Clinical data for the 45 patients that constitute the 500K SNP array cohort is summarised in Figure 3.1. The comprehensive data set is detailed in Table 3.1. The term 'paediatric' defined a patient aged below 21 years. Of the 63 ependymomas analysed, 57 were from 42 patients who had contributed either a primary tumour alone, or primary and subsequent recurrent tumours. Indeed, the cohort incorporated eight sets of primary and recurrent ependymomas from the same patient.
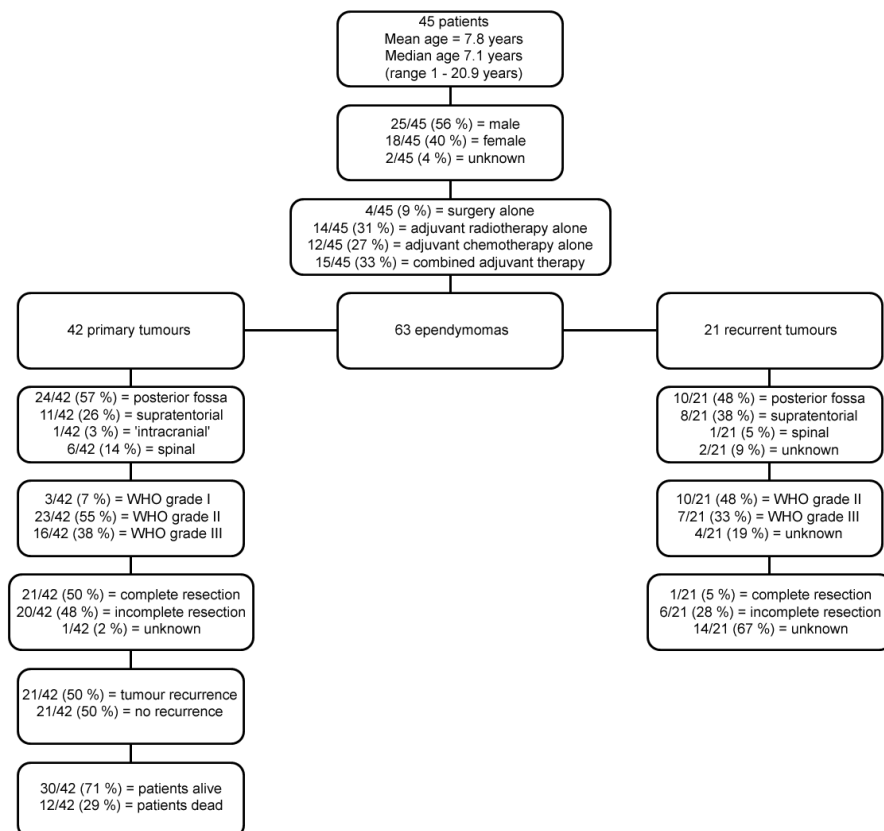
Figure 3.1: Clinical data summary for the 45 patients that constitute the 500K SNP array cohort. Of the 63 tumours analysed, 57 were from 42 patients who had contributed either a primary tumour alone, or primary and subsequent recurrent tumours. The remaining six ependymomas from three patients contributing only recurrent tumours.

94

Table 3.1: Clinical parameters of the 500K SNP microarray cohort.

Sample ID	Age at Diagnosis (yrs)	Sex	Tumour Location	WHO Grade	Primary Tumour Resection Status	Adjuvant therapy	Status	EFS (yrs) (> if censored)	OS (yrs) (> if censored)
1P	1.0	M	PF	II	Complete	C	D	1.4	2.1
2P	1.3	F	PF	III	Incomplete	C	D	1.5	1.6
3P	1.5	M	/	III	Complete	C	ADF	1.8	1.8
4P	1.8	M	ST	II	Complete	C	ADF	0.8	0.8
5P	1.8	F	PF	II	Complete	C	ADF	8.3	8.3
6P	1.8	M	PF	II	Incomplete	C	A (palliative)	1.3	1.3
7P	2.0	F	PF	III	Complete	C & RT	D	8.8	9.6
8P	2.4	M	PF	III	Incomplete	C	ADF	7.3	7.3
9P	2.5	M	PF	III	Complete	C, RT at relapse	D	2.2	6.9
9R2	as above	as above	/	/	/	RT	as above	as above	as above
9R3	as above	as above	PF	III	/	RT	as above	as above	as above
9R4	as above	as above	/	/	/	RT	as above	as above	as above
9R5	as above	as above	PF	III	/	RT	as above	as above	as above
10P	2.5	/	/	II	Incomplete	C	D	1.2	1.3
11P	2.6	M	PF	II	Complete	C	ADF	1.7	1.7
12P	2.6	M	PF	III	/	C	ADF	2.8	2.8
13P	2.8	M	PF	II	Complete	C & RT	ADF	12.8	12.8
14P	2.9	M	PF	II	Incomplete	C, RT at relapse	A (r)	0.3	10.1
15P	3.5	F	PF	II	Complete	C	ADF	1.7	1.7
16P	3.8	F	PF	II	Complete	RT	A (r)	2.9	8.2
16R1	as above	as above	PF	II	Complete	RT	as above	as above	as above
17P	3.8	F	ST	II	Incomplete	C & RT	A (r)	2.3	21.0
17R1	as above	as above	ST	II	Incomplete	C & RT	as above	as above	as above
17R2	as above	as above	ST	II	Incomplete	C & RT	as above	as above	as above
18P	4.2	M	ST	III	Incomplete	C	D	2.8	3.5
18R1	as above	as above	ST	III	Incomplete	C	as above	as above	as above
19R1	4.5	M	PF	/	Incomplete	RT	D	1.4	5.8
19R2	as above	as above	PF	/	Incomplete	RT	as above	as above	as above
20P	5.6	M	ST	II	Incomplete	RT	D	0.8	1.0
20R1	as above	as above	ST	II	Incomplete	RT	as above	as above	as above
21P	6.5	/	PF	II	Incomplete	Surgery alone	D	0.6	1.4
22P	6.7	F	ST	II	Complete	C & RT	ADF	5.8	5.8
23P	7.1	F	ST	III	Complete	RT	D	1.5	5.6
24P	7.1	M	SP	II	Incomplete	RT	ADF	3.1	3.1

Sample ID	Age at Diagnosis (yrs)	Sex	Tumour Location	WHO Grade	Primary Tumour Resection Status	Adjuvant therapy	Status	EFS (yrs) (> if censored)	OS (yrs) (> if censored)
25P	8.0	F	PF	II	Complete	RT	ADF	2.5	2.5
26P	8.8	M	PF	II	Incomplete	C & RT	D	4.0	5.7
26R1	as above	as above	PF	II	/	C & RT	as above	as above	as above
26R2	as above	as above	PF	II	/	C & RT	as above	as above	as above
26R3	as above	as above	PF	II	/	C & RT	as above	as above	as above
27P	8.9	F	SP	I	Complete	Observation only	ADF	1.1	1.1
28P	9.8	M	PF	III	Complete	RT	ADF	3.5	3.5
29P	10.0	F	PF	II	Incomplete	C & RT	ADF	0.5	0.5
30P	10.0	M	PF	II	Complete	C & RT	D	2.7	5.7
31P	10.3	M	SP	I	Incomplete	RT at relapse	A (r)	0.8	2.3
32P	11.4	F	ST	III	Complete	RT	ADF	2.2	2.2
33R1	12.0	F	ST	III	/	RT	A (r)	0.5	8.5
34P	12.0	F	ST	III	Incomplete	C, RT at relapse	D	0.3	3.0
35P	12.2	F	PF	III	Incomplete	C & RT	A (r)	2.2	5.8
35R1	as above	as above	PF	III	/	C & RT	as above	as above	as above
35R2	as above	as above	PF	III	/	C & RT	as above	as above	as above
36P	12.7	M	ST	III	Incomplete	C & RT	A (r)	0.3	9
37P	13.8	F	ST	II	Incomplete	RT	ADF	0.7	0.7
38P	14.6	M	SP	II	Incomplete	Surgery alone	A (r)	/	/
39P	14.7	M	PF	III	Complete	RT	ADF	6.6	6.6
40P	14.7	M	PF	III	Complete	RT, CRT at relapse	A (r)	1.0	3.3
40R1	as above	as above	SP	III	/	C & RT	as above	as above	as above
41P	14.8	F	ST	II	Complete	Surgery alone	ADF	7.8	7.8
42P	15.3	M	SP	I	Incomplete	RT	ADF	0.7	0.7
43R1	15.5	M	ST	II	/	C & RT	D	7.9	9.3
43R2	as above	as above	ST	II	/	C & RT	as above	as above	as above
43R3	as above	as above	ST	II	/	C & RT	as above	as above	as above
44P	16.3	M	PF	III	Complete	C & RT	A (r)	3.1	3.2
45P	20.9	F	SP	II	Incomplete	RT	ADF	1.1	1.1

Samples are listed in age order, with the term 'paediatric' defining a patient aged below 21 years. P = primary, R1 – R5 = 1st – 5th recurrence, M = male, F = female, PF = posterior fossa/infratentorial, ST = supratentorial, SP = spinal, C = chemotherapy, RT = radiotherapy, A (r) = alive but relapsed, ADF = alive and disease free, D = dead of disease, EFS = event-free survival, OS = overall survival.

3.2.1.2 Global genomic imbalance data analysis

While chromosome arm aberrations could be visualised through Spotfire Decision Site[®] generated heatmaps (Chapter 2, section 2.2.5), the precise designation of chromosome arm imbalance was performed using Microsoft Excel 2007. For a chromosome arm to be defined as gained or lost, 80 % or more of the SNP probes located on that arm had to have a copy number above or below two respectively. Specifically, a copy number of three to six represented gain, with a copy number above five signifying amplification. In contrast, a copy number of one represented hemizygous loss, whereas a value of zero constituted homozygous deletion. While removal of the X chromosome from analysis has been discussed, chromosome 21p was also excluded due to the paucity of SNP probes on this arm.

The same probe aberration threshold was adopted to identify smaller regional imbalances within 783 cytobands (Shaffer 2009) across the 22 autosomes. Gain of a region was thereby defined by a copy number of three or above, while loss was signified by a copy number of one or below, in at least 80 % of interrogated SNP probes. A complete list of cytoband copy number data for all 63 ependymomas in the SNP array cohort is included in Appendix 10D. Unsupervised hierarchical clustering of this genomic cytoband data was performed in Spotfire Decision Site[®] where both UPGMA (unweighted average) and euclidean distance settings were applied. The observed cluster patterns were then validated by principal component analysis (PCA). Via the 'rgl', 'pvclust' and 'cluster' packages in the R software environment, a 3-dimensional PCA cluster plot was generated. A computer script was devised for this aspect of the data analysis by Dr. Edward Schwalbe and Dr. Steve Clifford at the Northern Institute for Cancer Research, Newcastle (Appendix 10E).

Tumours with a balanced genome were defined as those without chromosome arm or cytoband imbalance, where at least 95 % of all SNP probes across the tumour genome had a diploid copy number.

3.2.1.3 SNP array call rates

As described in Chapter 2, section 2.2.3, the SNP call rate for an array was generated by GTYPE as a percentage marker of probe, and thereby array, reliability. While values of 93 % or above were deemed successful, lower values were included because of the genomic aberrations present in tumour DNA. Values below 85 % were deemed unacceptable for the purposes of subsequent analysis. The SNP call rate results for the 63 paediatric ependymoma samples and the 40 constitutional bloods analysed were of a satisfactory quality (Appendices 3 and 4). The Affymetrix[®] NspI 250K array results for the tumour DNA samples had a mean SNP call rate of 96.38 ± 0.31 %, a median of 97.01 % and a range of 87.76 – 99.2 %, while the Affymetrix[®] StyI 250K array results had a mean SNP call rate of 95.26 ± 0.3%, a median of 95.87 % and a range of 87.27 % – 97.97 %. Likewise, the Affymetrix[®] NspI 250K array results for the blood DNA samples had a mean SNP call rate of 96.4 ± 0.43 %, a median of 97.2 % and a range of 88.27 – 99.5 %, while the Affymetrix[®] StyI 250K array results had a mean SNP call rate of 96.5 ± 0.34 %, a median of 97.3 % and a range of 88.28 – 99.11 %.

3.2.1.4 FISH validation of SNP array results

The protocol adhered to and the method of scoring used is described in Chapter 2, section 2.5.

3.2.2 Methylation array analysis of 98 paediatric ependymomas

3.2.2.1 The sample cohort

DNA from 110 paediatric ependymomas (84 primary and 26 recurrent tumours) was processed using the Illumina[®] GoldenGate[®] Cancer Panel I array for methylation at the Wellcome Trust Centre for Genetics, Oxford, UK under the guidance of project manager Dr. Joseph Trakolo. An overview of the protocol adopted and data processing methods are described in Chapter 2, sections 2.3.1 – 2.3.4. Two SAMs were used to analyse the 110 ependymoma samples. Surplus well spaces on the second SAM were occupied by paediatric high grade glioma DNA samples, which were being assessed in

an independent methylation analysis. Technical replicates from the high grade glioma cohort were used to assess intra-assay variability. After performing the quality control measures described in Chapter 2, section 2.3.3, 12 unsatisfactory ependymoma samples were excluded from further analysis, resulting in the analysis of 98 tumours from 82 patients (73 primary and 25 recurrent ependymomas). SNP array analysis was performed on 51 of these ependymomas (33 primary and 18 recurrent).

Clinical data for the 82 patients comprising the methylation array cohort is summarised in Figure 3.2. The comprehensive data set is detailed in Table 3.2. The term 'paediatric' again defined a patient aged below 21 years.
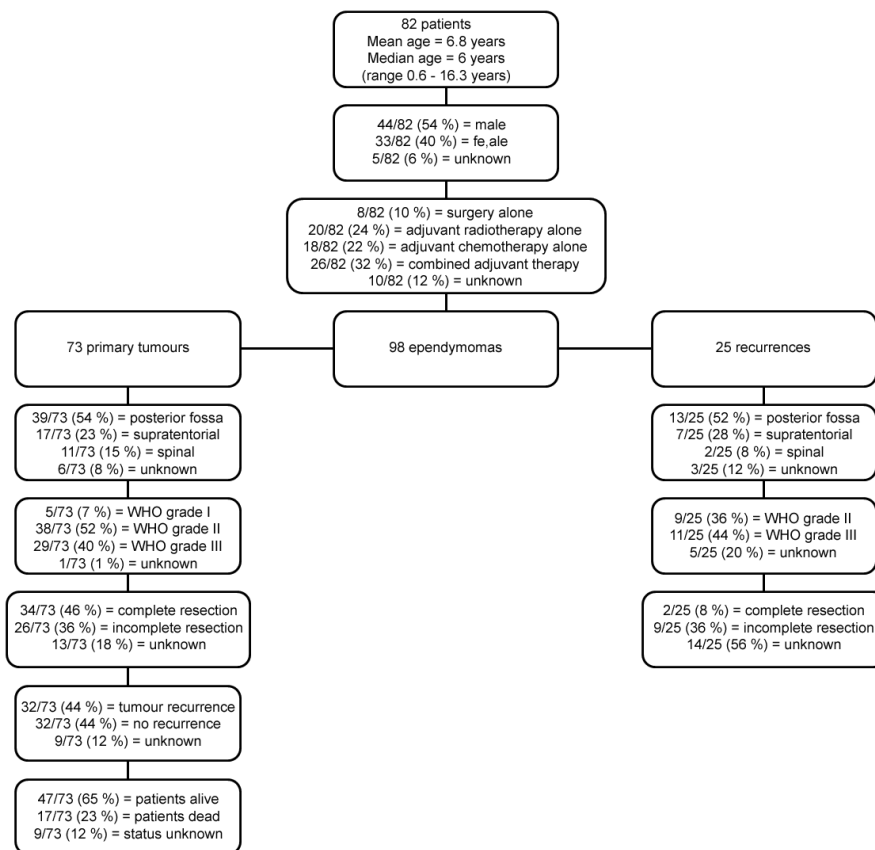
Figure 3.2: Clinical data summary for the 82 patients that constitute the methylation array cohort. Of the 98 ependymomas analysed, 86 were from 73 patients who had contributed either a primary tumour alone, or primary and subsequent recurrent tumours. The cohort also contained 12 ependymomas from nine patients contributing only recurrent tumours.

Table 3.2: Clinical parameters of the methylation array cohort.

Sample ID	Age at Diagnosis	Sex	Tumour Location	WHO Grade	Primary Tumour Resection Status	Adjuvant therapy	Status	EFS (yrs) (> if censored)	OS (yrs) (> if censored)
M 1P	0.6	F	ST	III	Incomplete	C	D	0.3	0.6
M 2P	1.0	M	PF	II	Complete	C	A(r)	1.4	2.1
M 3P	1.3	F	PF	III	Incomplete	C	D	1.5	1.6
M 4P	1.3	M	PF	II	Incomplete	C & RT	D	0.6	2.9
M 5R1	1.3	M	PF	II	Incomplete	C & RT	D	5.7	6.0
M 6P	1.5	M	/	III	Complete	C	ADF	1.8	1.8
M 7P	1.6	M	ST	III	Complete	C	ADF	5.3	5.3
M 8P	1.7	F	PF	II	Incomplete	C, RT at relapse	A(r)	1.2	8.1
M 9P	1.8	M	ST	II	Complete	C	ADF	0.8	0.8
M 10P	1.8	F	PF	II	Complete	C	ADF	8.3	8.3
M 11R1	1.8	M	PF	III	/	C, RT at relapse	A(r)	1.4	2.2
M 12P	1.9	M	PF	III	Incomplete	C	ADF	0.0	0.0
M 13P	2.0	F	/	III	Complete	C, RT at relapse	A(r)	2.3	2.6
M 14P	2.0	F	PF	III	Complete	C	ADF	15.2	15.2
M 15P	2.0	F	PF	III	Complete	C & RT	D	8.8	9.6
M 16P	2.0	M	PF	II	/	/	/	/	/
M 17R1	2.2	F	PF	II	Complete	C, RT at relapse	A(r)	5.8	10.6
M 18P	2.3	M	PF	II	Complete	C, RT at relapse	D	5.6	6.4
M 18R1	as above	as above	/	/	Complete	RT	D	as above	as above
M 19P	2.3	M	/	III	Incomplete	C	A(r)	1.3	1.3
M 20P	2.3	M	PF	II	/	/	/	/	/
M 21P	2.4	M	PF	III	Incomplete	C	ADF	7.3	7.3
M 22P	2.5	M	PF	III	Complete	C, RT at relapse	D	2.2	6.9
M 22R2	as above	as above	/	/	/	RT	D	as above	as above
M 22R3	as above	as above	PF	III	/	RT	D	as above	as above
M 22R5	as above	as above	PF	III	/	RT	D	as above	as above
M 23P	2.5	/	PF	III	/	/	/	/	/
M 24P	2.6	M	PF	II	/	C	ADF	1.7	1.7
M 25P	2.6	M	PF	II	Incomplete	C	D	1.8	1.9
M 26P	2.6	M	PF	III	/	C	ADF	2.8	2.8
M 27P	2.8	M	PF	II	Complete	C	ADF	1.4	1.4
M 28P	3.5	F	PF	II	Complete	C	ADF	1.7	1.7
M29P	3.6	M	PF	II	Incomplete	/	/	/	/
M30P	3.7	M	PF	II	Complete	RT	ADF	5.5	5.5

Sample ID	Age at Diagnosis (yrs)	Sex	Tumour Location	WHO Grade	Primary Tumour Resection Status	Adjuvant therapy	Status	EFS (yrs) (> if censored)	OS (yrs) (> if censored)
M 31P	3.8	F	PF	II	Complete	RT	A(r)	2.9	8.2
M 31R1	as above	as above	/	/	/	/	as above	as above	as above
M 32P	3.8	F	ST	II	Incomplete	C & RT	A(r)	2.3	21.0
M 32R1	as above	as above	ST	II	Incomplete	/	as above	as above	as above
M 32R2	as above	as above	ST	II	Incomplete	/	as above	as above	as above
M 33P	4.0	F	PF	II	Complete	C	ADF	2.3	2.3
M 34P	4.2	M	ST	III	Incomplete	C	D	2.8	3.5
M 34R1	as above	as above	ST	III	Incomplete	C	D	as above	as above
M 35P	4.3	F	PF	II	Complete	RT	/	/	/
M 36R1	4.5	M	PF	/	Incomplete	RT	D	1.4	5.8
M 36R2	as above	as above	PF	/	Incomplete	RT	as above	as above	as above
M 37P	4.8	F	PF	II	Incomplete	RT	D	2.3	3.2
M 38P	5.0	F	PF	II	Complete	RT	ADF	1.1	1.1
M 39P	5.4	M	PF	III	/	RT	A(r)	2.3	4.3
M 40R1	5.6	M	ST	II	Incomplete	RT	D	0.8	1.0
M 41P	6.0	M	/	II	/	/	/	/	/
M 42R1	6.0	F	PF	III	/	/	/	/	/
M 43P	6.1	M	PF	III	Complete	RT	D	0.9	1.8
M 44P	6.3	F	ST	III	Complete	C & RT	D	1.9	5.1
M 45R1	6.5	/	PF	III	Incomplete	RT	A(r)	1.1	2.8
M 46P	6.5	/	PF	II	Complete	RT	ADF	2.4	2.4
M 47P	6.5	/	PF	II	/	/	D	0.6	1.4
M 48P	6.7	F	ST	II	Complete	C & RT	ADF	7.1	7.1
M 49P	7.1	F	ST	III	Complete	/	D	1.5	5.6
M 50P	7.1	M	SP	II	Incomplete	RT	ADF	3.1	3.1
M 51P	7.1	F	ST	III	Complete	RT	A(r)	1.5	3.8
M 52P	7.3	F	SP	III	Incomplete	Surgery alone	ADF	2.3	2.3
M 53P	8.2	M	SP	II	Incomplete	C & RT	ADF	0.9	0.9
M 54P	8.8	M	PF	II	/	C & RT	D	4.0	5.7
M 54R2	as above	as above	PF	III	/	C & RT	as above	as above	as above
M 54R3	as above	as above	PF	III	/	C & RT	as above	as above	as above
M 55P	9.3	F	PF	II	/	RT, C at relapse	D	0.8	1.5
M 56P	9.8	F	ST	III	Complete	C & RT	ADF	8.0	8.0
M 57P	9.8	M	PF	III	Complete	RT	ADF	4.8	4.8

Sample ID	Age at Diagnosis (yrs)	Sex	Tumour Location	WHO Grade	Primary Tumour Resection Status	Adjuvant therapy	Status	EFS (yrs) (> if censored)	OS (yrs) (> if censored)
M 58P	10.0	M	PF	II	Complete	C & RT	D	2.5	5.6
M 59P	10.0	F	PF	II	Incomplete	C & RT	ADF	0.5	0.5
M 60P	10.0	M	SP	I	/	/	/	/	/
M 61P	10.3	M	SP	I	Incomplete	RT at relapse	A(r)	0.8	2.3
M 62P	10.9	F	ST	III	Incomplete	C & RT	ADF	1.8	1.8
M 63P	11.0	M	SP	II	Complete	Surgery alone	ADF	0.9	0.9
M 64P	11.4	F	ST	III	Complete	RT	ADF	2.2	2.2
M 65P	11.6	F	SP	II	Incomplete	Surgery alone	ADF	1.6	1.6
M 66P	12.0	F	ST	III	Incomplete	C & RT	D	0.3	3.0
M 67P	12.2	F	PF	III	Incomplete	C & RT	A(r)	2.2	5.8
M 67R1	as above	as above	PF	III	/	C & RT	as above	as above	as above
M 67R2	as above	as above	PF	III	/	C & RT	as above	as above	as above
M 68R1	12.5	M	SP	II	/	RT, C&RT at relapse	A(r)	3.3	5.0
M 69P	12.7	M	ST	III	Incomplete	C & RT	A(r)	0.3	9.0
M 70P	13.2	M	SP	II	Complete	Surgery alone	ADF	2.3	2.3
M 71P	13.8	F	ST	II	Incomplete	RT	ADF	0.7	0.7
M 72P	14.6	M	SP	II	Incomplete	Surgery alone	A(r)	3.1	3.3
M 73P	14.7	M	PF	III	Complete	C, C&RT at relapse	A(r)	1.0	3.3
M 73R1	as above	as above	SP	III	/	C & RT	as above	as above	as above
M 74P	14.8	F	ST	II	Complete	Surgery alone	ADF	7.8	7.8
M 75P	15.3	M	SP	I	Incomplete	RT	ADF	0.7	0.7
M 76R1	15.5	M	ST	II	/	C & RT	D	7.9	9.3
M 76R2	as above	as above	ST	II	/	C & RT	as above	as above	as above
M 76R3	as above	as above	ST	II	/	C & RT	as above	as above	as above
M 77P	15.7	M	SP	I	/	Surgery alone	ADF	2.0	2.0
M 78P	15.9	F	PF	II	Incomplete	RT	ADF	3.1	3.1
M 79P	16.0	M	SP	I	Complete	Surgery alone	/	/	/
M 80P	16.3	M	PF	III	Complete	C & RT	A(r)	3.1	3.2
M 81P	16.3	F	ST	II	Complete	RT	A(r)	1.1	1.1
M 82P	/	/	/	/	/	/	/	/	/

Samples are listed in age order, with the term 'paediatric' defining a patient aged below 21 years. P = primary, R1 – R5 = 1st – 5th recurrence, M = male, F = female, PF = posterior fossa/infratentorial, ST = supratentorial, SP = spinal, C = chemotherapy, RT = radiotherapy, A (r) = alive but relapsed, ADF = alive and disease free, D = dead of disease, EFS = event-free survival, OS = overall survival.

3.2.2.2 Methylation array data analysis

Once corrected Beta methylation scores for the 1,505 probes had been calculated, as discussed in Chapter 2, section 2.3.4, 84 probes on the X chromosome were removed from the analysis as their inclusion in this mixed gender cohort resulted in clustering of samples according to patient sex. The corrected Beta scores across 1,421 GoldenGate® Cancer Panel I CpG probes, mapping to 768 genes, for all 98 tumours comprising the ependymoma methylation cohort is shown in Appendix 10F. Bootstrapped unsupervised hierarchical clustering was performed using the R package 'pvclust' (Suzuki and Shimodaira 2006) using Euclidean distance, average agglomeration and 10,000 replications. Subgroups with an approximate unbiased probability (p) value of greater than 95 % were deemed significant. The observed clustering patterns were assessed using principle component analysis via the R programmes 'rgl', 'pvclust' and 'cluster'. A computer script was devised for this aspect of the data analysis by Dr. Edward Schwalbe and Dr. Steve Clifford at the Northern Institute for Cancer Research, Newcastle (Appendix 10G).

3.2.2.3 Assessment of methylation intra-assay reproducibility

Four superfluous wells were used to assess intra-assay variability, using DNA from two high grade glioma samples as technical replicates (Figure 3.3). The correlation of the replicates for both samples was high (correlation coefficients 0.868 and 0.812, $p < 0.01$, Kendall tau correlation), indicating low intra-assay variability.
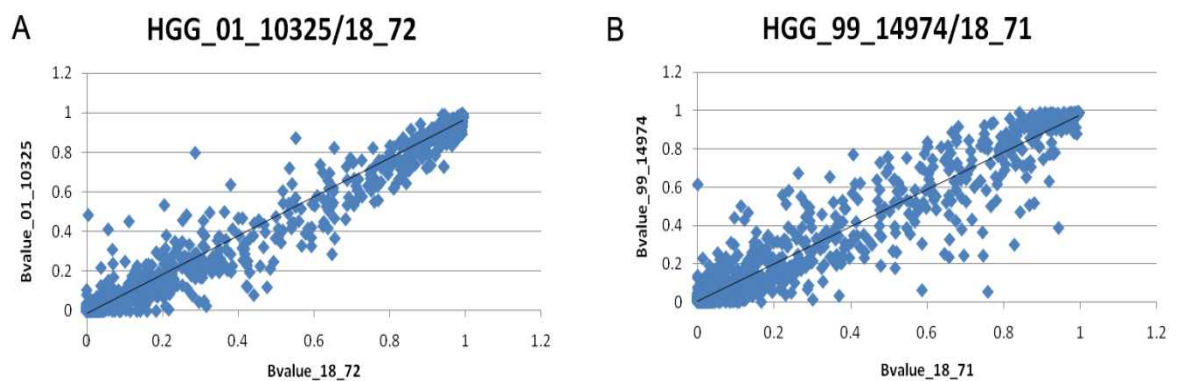
Figure 3.3: Kendall tau correlation of corrected Beta scores from the GoldenGate® Cancer Panel I methylation array for two sets of high grade glioma technical replicates. (A) Replicate sample A was divided and labelled HGG_01_10325 for the first assay and 18_72 for the second assay. (B) Replicate sample B was divided and labelled HGG_99_14974 for the first assay and 18_71 for the second assay. Replicate sample A had a correlation coefficient of 0.868, $p < 0.01$, while replicate sample B had a correlation coefficient of 0.812, $p < 0.01$.

3.2.3 Statistical analysis

The statistical tests performed in this chapter are described in Chapter 2, section 2.7. The evaluation of genomic imbalances between primary ependymomas categorised according to a patient age threshold of three years was restricted to the posterior fossa cohort of 24 ependymomas, where tumours were obtained from 12 children aged above three years and 12 patients aged below three years. This enabled a comparative assessment of sufficient numbers of ependymomas from a standardised CNS location.

Multivariate event-free survival analysis for the SNP and methylation array cohorts included the clinical variables of patient age, tumour location, WHO histological grade and degree of surgical resection. Since adjuvant therapy commencement dates were unavailable for the many patients in both cohorts, radiotherapy and chemotherapy were removed as variables for event-free survival as the timing of their use relative to any relapses that occurred in these patients could not be determined accurately. Nevertheless radiotherapy and chemotherapy remained as variables for overall survival.

3.3 Results

3.3.1 Clinical associations: the SNP array primary tumour cohort

Statistical analysis of the primary tumour cohort from the SNP array study was undertaken to initially identify any associations between the clinical variables of tumour location, patient age at diagnosis, tumour grade and tumour resection status. This revealed posterior fossa ependymomas were associated with patients aged below three years ($p = 0.005$, two-tailed Fisher's exact test). The association of posterior fossa with young patients and spinal ependymoma with older age groups was confirmed by one-way ANOVA. This revealed a significant difference in mean patient age between primary posterior fossa tumours (6 ± 1 years), supratentorial tumours (8.5 ± 1.4 years) and spinal tumours (12.8 ± 2.1 years) ($p = 0.01$). Post-hoc comparisons revealed this was the result of the patient age difference between posterior fossa and spinal ependymomas ($p = 0.009$, eta 0.2), while the age difference between intracranial

ependymoma patients was not significant. No other associations between the variables detailed above were observed.

The estimated mean overall survival time for patients of the primary tumour cohort was 12 ± 2 years (range $0.5 - 21$ years). The mean time for follow up was 4.5 ± 0.6 years (range $0.5 - 21$ years). Disease progression occurred in half of the cases (21/42) with a mean time to relapse of 2.4 ± 0.5 years (range $0.3 - 8.8$ years). The percentage of the primary cohort achieving five year event-free survival was 35.7 ± 9.4 %, while five year overall survival was achieved by 77.4 ± 8 %. Univariate survival analysis revealed incomplete resection to be associated with a worse estimated mean event-free survival (6.8 ± 1.4 years versus 2.4 ± 0.6 years, $p = 0.007$). Resection status had no significant effect on overall survival, however. Other variables including patient age, sex, tumour location, tumour grade and adjuvant therapy had no association with patient outcome.

3.3.2 Clinical associations: the methylation array primary tumour cohort

Statistical analysis of the primary tumour cohort from the methylation array study was also undertaken to initially identify any associations between the clinical variables of tumour location, patient age at diagnosis, tumour grade and tumour resection status. This revealed supratentorial ependymomas were associated with female patients and an anaplastic histology ($p = 0.005$ and 0.019 respectively, two-tailed Fisher's exact test). Posterior fossa ependymomas were again associated with patients aged below three years ($p = 0.005$, two-tailed Fisher's exact test). One-way ANOVA once more confirmed significant differences in mean patient age for ependymomas from different CNS locations (posterior fossa tumours: 5.2 ± 0.7 years; supratentorial tumours: 8.3 ± 1.1 years; spinal tumours: 11.6 ± 0.9 years, $p = 0.0004$). As seen in the SNP array cohort, post-hoc comparisons identified the patient age difference between posterior fossa and spinal ependymomas was the most significant ($p = 4 \times 10^{-5}$, eta 0.26), although the difference between posterior fossa and supratentorial tumours was also significant ($p = 0.036$).

The estimated mean overall survival time was 10.8 ± 2 years (range $0 - 21$ years). All spinal cases remained alive. The mean time for follow up was 4 ± 0.5 years (range $0 - 21$ years). Recurrence occurred in 32/73 (44 %) patients with a mean time to relapse of

2 ± 0.3 years (range 0.3 – 7.1 years). The percentage of the cohort achieving five year event-free survival was 37.2 ± 7.6 %, while five year overall survival was achieved by 76.8 ± 6.7 %. As was the case for the primary SNP array cohort, survival analysis revealed incomplete resection to be associated with a worse estimated mean event-free survival (6.7 years ± 1.4 years versus 2.9 years ± 0.6 years, p = 0.032). Again, resection status had no significant effect on overall survival. Patient age, sex, tumour location, tumour grade and adjuvant therapy also had no association with patient outcome.

3.3.3 Chromosomal arm imbalance in 63 paediatric ependymomas

To enable a global visualisation of chromosome arm imbalance across the entire paediatric ependymoma SNP array cohort, a copy number heatmap was created in Spotfire Decision Site® (Figure 3.4). A table summarising the chromosome arm anomalies for each tumour sample in the cohort was also generated (Table 3.3). Net gain of chromosomal material occurred more frequently than loss. The most frequent aberration was gain of the long arm of chromosome 1 (1q) which was seen in 13/63 (21 %) cases. Three of these cases (21P, 26R2, 44P) demonstrated a level of 1q gain represented by a copy number of four or higher. Gain of either 9p or all of chromosome 9 was another common feature, seen in 12/63 (19 %) of cases. The most frequent loss was of chromosome 22q, present in 5/63 (8 %) tumours, while loss of either 6q or the whole of chromosome 6 occurred in 3/63 (5 %) cases.

Table 3.3: Chromosome arm alterations in 63 paediatric ependymomas, analysed using the Affymetrix® 500K SNP array platform (80 % imbalance threshold).

Tumour sample	Chromosome/arm gain	Chromosome/arm loss
1P	-	-
2P	-	-
3P	-	-
4P	-	-
5P	-	-
6P	-	-
7P	-	-
8P	-	-
9P	-	-
9R2	19, 22q	-
9R3	-	-
9R4	-	6q
9R5	-	-
10P	1q	16q
11P	-	-

Tumour sample	Chromosome/arm gain	Chromosome/arm loss
12P	-	-
13P	-	-
14P	-	-
15P	-	-
16P	-	-
16R1	1q	-
17P	-	-
17R1	-	-
17R2	-	-
18P	-	-
18R1	-	-
19R1	8,9	22q
19R2	1q, 2, 8, 9p	-
20P	-	-
20R1	-	-
21P	1q*, 7	-
22P	-	-
23P	-	-
24P	9, 18, 20	-
25P	9	-
26P	1q	-
26R1	1q, 2p, 8, 9	-
26R2	1q*, 2p, 8, 9	-
26R3	1q, 2p, 8, 9	-
27P	4, 7, 11, 12, 16, 17, 18, 21q, 22q	-
28P	4, 5, 9, 11, 12, 18, 19, 20	21q
29P	4, 14q, 15q, 18	-
30P	1q	10q
31P	4, 5, 7p, 9, 11, 12, 16, 17, 18, 19, 20, 21q	-
32P	-	-
33R1	-	-
34P	19p	-
35P	1q	-
35R1	-	-
35R2	-	-
36P	-	-
37P	-	-
38P	20p	22q
39P	1q, 19	-
40P	9, 13q, 14q	6, 22q
40R1	9, 13q, 14q	6, 22q
41P	-	-
42P	9, 16, 17, 20	-
43R1	1q	-
43R2	-	-
43R3	-	-
44P	1q*, 18	-
45P	2, 7, 12q, 15q, 17	22q

Genomic gain is represented by copy number three and loss is represented by copy number one unless stated (see *).
P = primary, R1 – R5 = 1st – 5th recurrence, * = copy number ≥ 4.
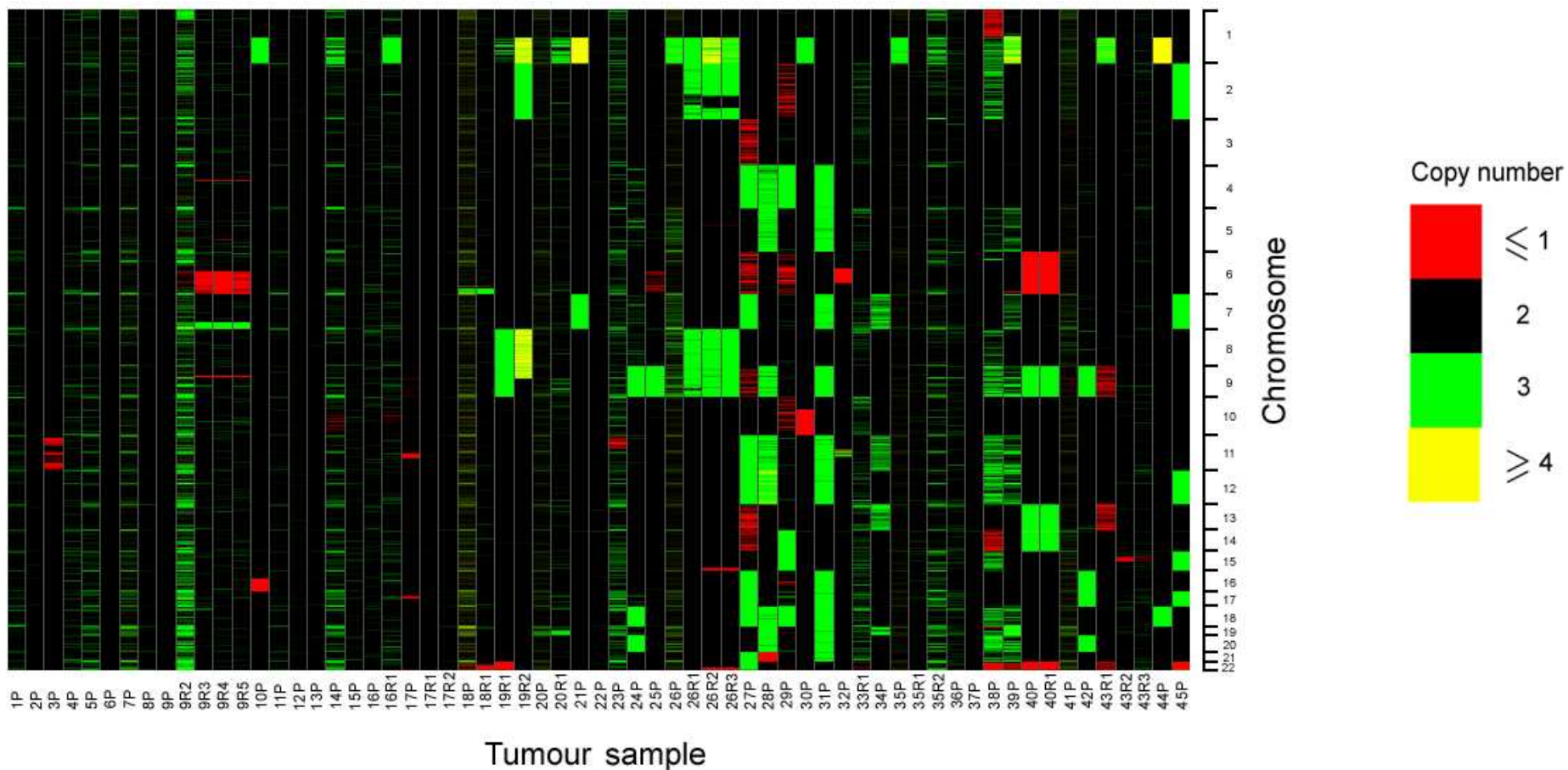
Figure 3.4: Spotfire Decision Site® copy number heatmap demonstrating Affymetrix® 500K SNP array results across the genome for 63 paediatric ependymomas, ordered by patient age. Diploid genomic regions are coloured black. Regions exhibiting genomic loss are coloured red, while regions demonstrating gain are coloured green or yellow, depending on whether the gain represents a copy number of three or greater respectively. P = primary, R1 – R5 = 1st – 5th recurrence.
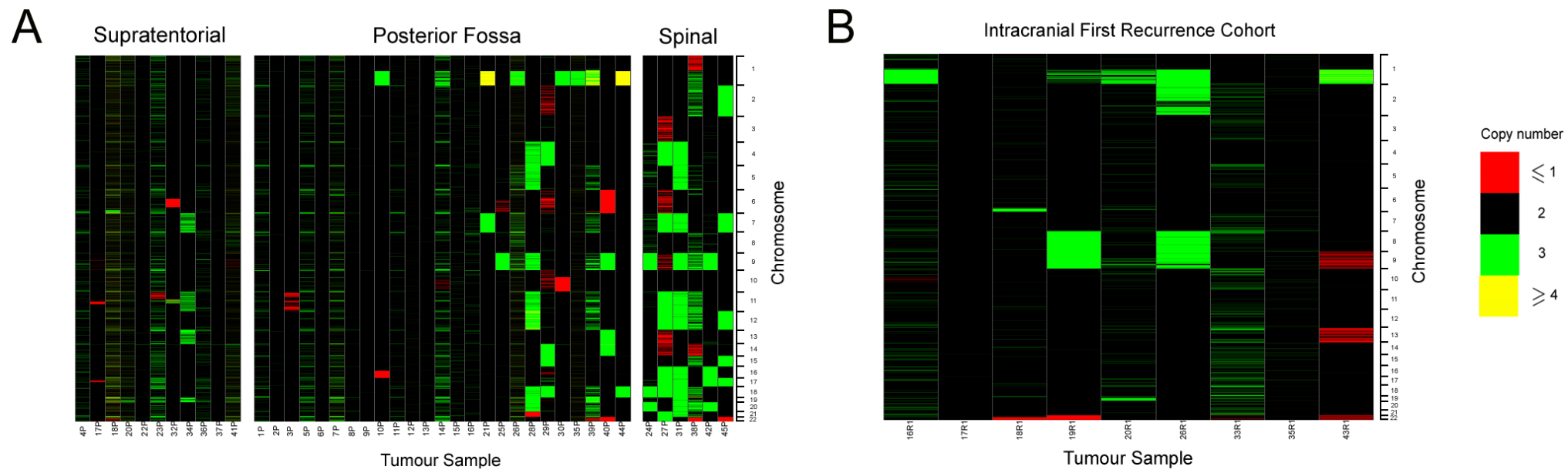
Figure 3.5: Spotfire Decision Site® copy number heatmaps demonstrating Affymetrix® 500K SNP array results across the genome for (A) 42 primary paediatric ependymomas, categorised according to the three locations of these tumours within the central nervous system; the supratentorial region, the posterior fossa and the spinal cord and (B) nine intracranial recurrent ependymomas. Diploid genomic regions are coloured black. Regions exhibiting genomic loss are coloured red, while regions demonstrating gain are coloured green or yellow, depending on whether the gain represents a copy number of three or greater respectively. P = primary, R1 = 1st recurrence.
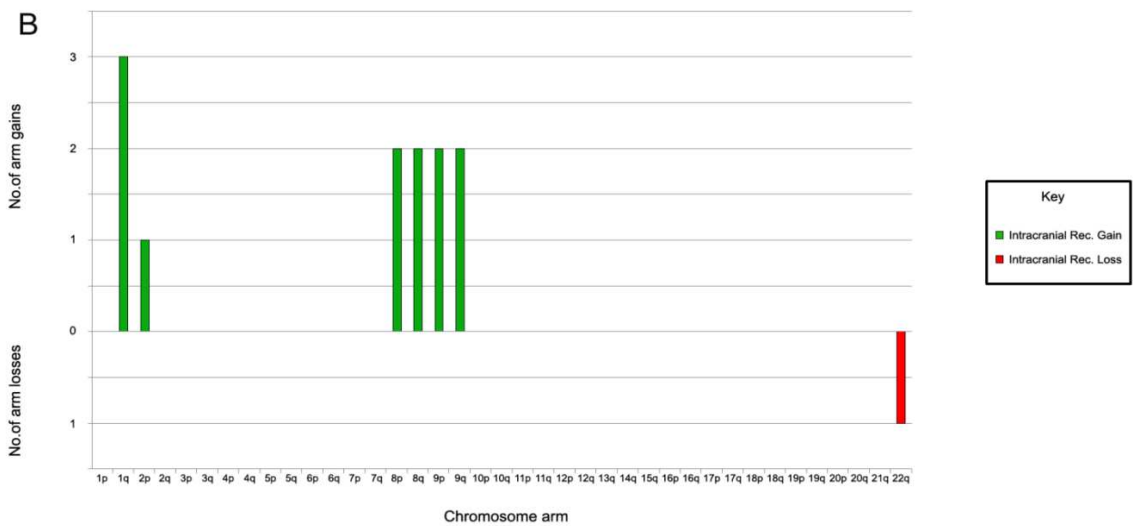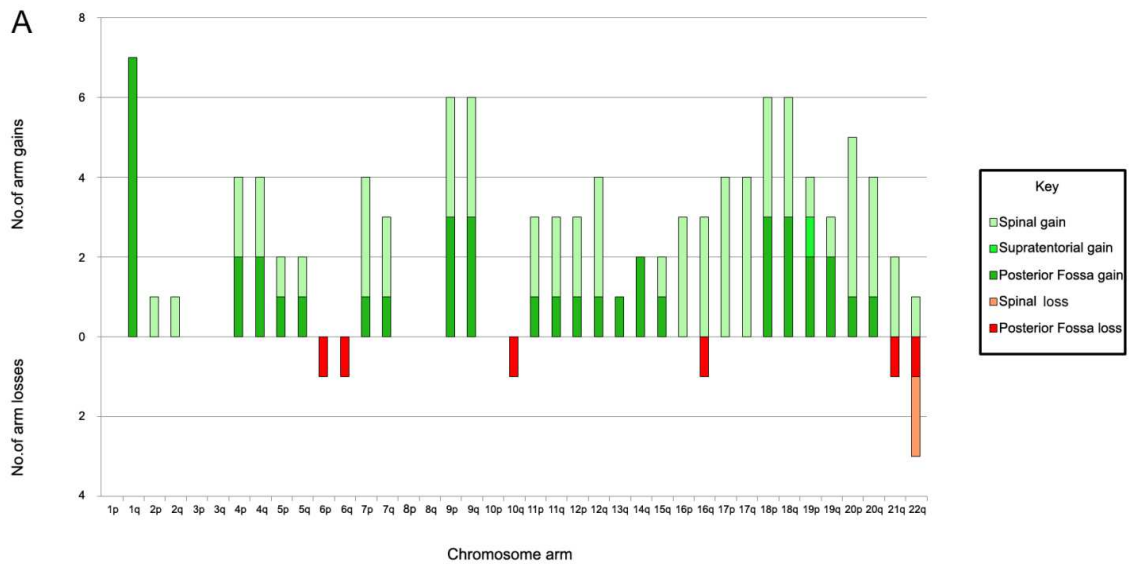
Figure 3.6: Frequency plot of the number of chromosome arm gains and losses in (A) 42 primary paediatric ependymomas and (B) nine intracranial recurrent ependymomas. Gains are coloured green and losses are coloured red. Furthermore, where relevant, the total number of chromosomal arm imbalances in the primary cohort is segregated by the frequency with which they occurred in each of the three CNS tumour locations – supratentorial, posterior fossa or spinal.

3.3.4 Association of chromosomal arm imbalances with clinical subgroups

Since 10 patients contributed more than one tumour sample to the 500K SNP array sample cohort, a more precise method of identifying associations between genomic imbalances and clinical subgroups was established by analysing only the primary tumours, followed by the intracranial first recurrent ependymomas. Consequently, Figure 3.5 illustrates Spotfire Decision Site® generated copy number heatmaps for the 42 paediatric primary ependymomas (separated according to tumour location within the CNS) and nine intracranial first recurrent tumours within the sample cohort, while Figure 3.6 provides a summarised view of the arm aberrations present in these two groups. Tables 3.4 – 3.7 order the most frequent chromosome arm imbalances in primary supratentorial, posterior fossa, spinal and first recurrent intracranial ependymomas.

Table 3.4: Chromosome arm alterations in 11 primary supratentorial ependymomas.

Chromosome arm gain		Chromosome arm loss	
Chromosome arm	Number (%) of tumours demonstrating gain	Chromosome arm	Number (%) of tumours demonstrating loss
19p	1 (9 %)	Nil	/

Table 3.5: Chromosome arm alterations in 24 primary posterior fossa ependymomas.

Chromosome arm gain		Chromosome arm loss	
Chromosome arm	Number (%) of tumours demonstrating gain	Chromosome arm	Number (%) of tumours demonstrating loss
1q	7 (29 %)	6p	1 (4 %)
9p	3 (13 %)	6q	1 (4 %)
9q	3 (13 %)	10q	1 (4 %)
18p	3 (13 %)	16q	1 (4 %)
18q	3 (13 %)	21q	1 (4 %)
4p	2 (8 %)	22q	1 (4 %)
4q	2 (8 %)		
14q	2 (8 %)		
19p	2 (8 %)		
19q	2 (8 %)		
5p	1 (4 %)		
5q	1 (4 %)		
7p	1 (4 %)		
7q	1 (4 %)		
11p	1 (4 %)		
11q	1 (4 %)		
12p	1 (4 %)		
12q	1 (4 %)		
13q	1 (4 %)		
15q	1 (4 %)		
20p	1 (4 %)		
20q	1 (4 %)		

Table 3.6: Chromosome arm alterations in six primary spinal ependymomas.

Chromosome arm gain		Chromosome arm loss	
Chromosome arm	Number (%) of tumours demonstrating gain	Chromosome arm	Number (%) of tumours demonstrating loss
17p	4 (67 %)	22q	2 (33 %)
17q	4 (67 %)		
20p	4 (67 %)		
7p	3 (50 %)		
9p	3 (50 %)		
9q	3 (50 %)		
12q	3 (50 %)		
16p	3 (50 %)		
16q	3 (50 %)		
18p	3 (50 %)		
18q	3 (50 %)		
20q	3 (50 %)		
4p	2 (33 %)		
4q	2 (33 %)		
7q	2 (33 %)		
11p	2 (33 %)		
11q	2 (33 %)		
12p	2 (33 %)		
21q	2 (33 %)		
2p	1 (17 %)		
2q	1 (17 %)		
5p	1 (17 %)		
5q	1 (17 %)		
15q	1 (17 %)		
19p	1 (17 %)		
19q	1 (17 %)		
22q	1 (17 %)		

Table 3.7: Chromosome arm alterations in nine intracranial first recurrent ependymomas.

Chromosome arm gain		Chromosome arm loss	
Chromosome arm	Number (%) of tumours demonstrating gain	Chromosome arm	Number (%) of tumours demonstrating loss
1q	3 (33 %)	22q	1 (11 %)
8p	2 (22 %)		
8q	2 (22 %)		
9p	2 (22 %)		
9q	2 (22 %)		
2p	1 (11 %)		

Distinct patterns of genomic imbalance between primary ependymomas from different CNS locations were evident. Spinal ependymomas were characterised by numerous arm and whole chromosomal aberrations when compared to intracranial tumours, particularly involving gain of chromosomes 17 ($p = 1 \times 10^{-4}$), 20p ($p = 7 \times 10^{-4}$), 16 ($p = 0.0019$), 12q ($p = 0.007$), 20q ($p = 0.007$), 21q ($p = 0.018$), 9 and 18 ($p = 0.03$, two-tailed Fisher's exact test). In contrast, a relative paucity of arm imbalances (19p gain in 1/11 samples) was a feature of supratentorial tumours. Gain of chromosome 1q was associated with primary posterior fossa ependymomas ($p = 0.0295$, two-tailed Fisher's exact test) and was the most frequent imbalance seen in the recurrent cohort, while gain of chromosome 8 was exclusive to intracranial recurrent tumours ($p = 0.036$, two-tailed Fisher's exact test). No statistical associations were found between particular chromosome arm imbalances and the clinical parameters of patient sex, tumour grade and degree of surgical tumour resection.

In this study, 15/42 (36 %) primary ependymomas were designated balanced (defined as $\geq$ 95% of the analysed SNP probes being assigned a diploid copy number), and were associated with children aged below three years ($p = 0.041$, two-tailed Fisher's exact test). Age-related categorical analyses of genomic imbalance was performed exclusively on the primary posterior fossa ependymoma cohort, where tumours were obtained from 12 children aged above three years and 12 patients aged below three years. While no association was found with a specific arm imbalance in either age category, the number of arm imbalances between the two subgroups was significant (one arm imbalance in the under three year group, 46 imbalances in the over three year group; $p = 1.2 \times 10^{-4}$, Mann-Whitney test).

The impact of chromosome 1q gain on patient outcome within the primary ependymoma cohort was assessed. While a reduced event-free survival (EFS) and overall patient survival (OS) was noted for ependymomas with 1q gain, neither reached univariate statistical significance (estimated mean EFS: 2.9 ± 0.7 years versus 5.8 ± 1.1 years, $p = 0.29$, estimated mean OS: 4.8 ± 0.8 years versus 13.1 ± 2.1 years, $p = 0.142$, Kaplan-Meier analysis). However, 1q gain showed a detrimental effect on overall survival from multivariate analysis incorporating the variables of patient age, adjuvant radiotherapy and chemotherapy, tumour histology, location and resection status (hazards ratio 37.1 (95 % CI 1.712 – 804.5, $p = 0.021$) (Table 3.8).

113

Table 3.8: Multivariate overall survival analysis incorporating SNP array gain of chromosome 1q.

Cox regression multivariate analysis (n=42)	Overall Survival		
Factor	**Hazards ratio**	**95 % CI**	**P value**
Histology (WHO grade II vs III)	1.986	0.532 – 7.413	0.308
Tumour location (PF vs ST)	0.032	0.001 – 1.422	0.075
Resection status (complete vs incomplete)	0.849	0.203 – 3.561	0.823
Patient age (below 3 years vs above 3 years)	13.032	0.353 – 480	0.163
Radiotherapy (received vs not received)	0.609	0.112 – 3.296	0.565
Chemotherapy (received vs not received)	2.405	0.257 – 22.501	0.442
Chromosome 1q gain (gain vs no gain)	37.117	1.712 – 805	0.021

The primary SNP array cohort of 42 ependymomas was analysed. ST = supratentorial, PF = posterior fossa, 95 % CI = 95 % confidence interval. Significant results are highlighted in yellow. Results with a trend towards significance are highlighted in pale yellow.

3.3.5 Association of cytoband imbalances with clinical subgroups

Initial cytoband analysis revealed the telomeric region 7p22.3 to be the most frequently gained in the primary cohort (14/42 cases, 33 %). The most frequently lost cytobands in this group included 6q11.2, 6q12 and 6q22.2 (4/42, 10 %). The most frequently gained and lost regions in the intracranial first recurrent cohort were 1q24.1 (5/9, 56 %) and 22q13.33 (3/9, 33 %) respectively. Fisher's exact testing revealed associations between genomic cytoband imbalances and particular clinical tumour groups, in addition to refining the cytobands within chromosome 1q that demonstrated significant gain in posterior fossa ependymomas (Table 3.9). These associations included gain of 9q34.11-34.4 in grade 3 intracranial primary ependymomas and gain of 1q24.1, loss of 22q13.1-13.31 and loss of 22q13.33 in intracranial first recurrent tumours.

Table 3.9: Statistically significant associations between cytoband genomic imbalances and particular clinical subgroups (as determined by two-tailed Fisher's exact testing).

Clinical Group	Cytoband imbalance	p-value
Male patient primary tumours	Gain of 20q11.22	0.01
	Gain of 9p11.1	0.02
	Gain of 9q13, 9q21.2, 9q33, 20q13.13	0.029
Posterior fossa primary tumours	Gain of 1q21.3, 1q22, 1q25.3, 1q32.2, 1q42.12-44	0.013
	Gain of 1q21.1, 1q23.1, 1q23.3-24.2, 1q25.1-25.2, 1q31.1-31.3, 1q41	0.03
Intracranial grade III primary tumours	Gain of 9p11.2	0.005
	Gain of 9p11.1, 9p12, 9p13.2, 9q34.11-12, 19p13.3	0.014
	Gain of 11p15.5, 19q13.41, 9q34.13-34.4, 11q13.2, 12q13.12, 13q11, 18p11.21, 18q11.1, 19q13.12, 19q13.13, 19q13.32, 19q13.42	0.035
Intracranial first recurrent tumours	Gain of 1q24.1	0.04
	Loss of 22q13.33	0.006
	Loss of 22q13.1 – 13.31	0.036

Note: Statistically significant associations already discovered during chromosome arm imbalance analysis (such as those seen with spinal ependymomas) are not shown unless cytoband analysis refined the region associated with the clinical subgroup.

Unsupervised hierarchical clustering and PCA was also performed on the cytoband imbalance data of the primary paediatric SNP array cohort, so that genetically distinct subgroups of tumours could be identified and linked with clinical and/or pathological parameters (Figures 3.7A – C).
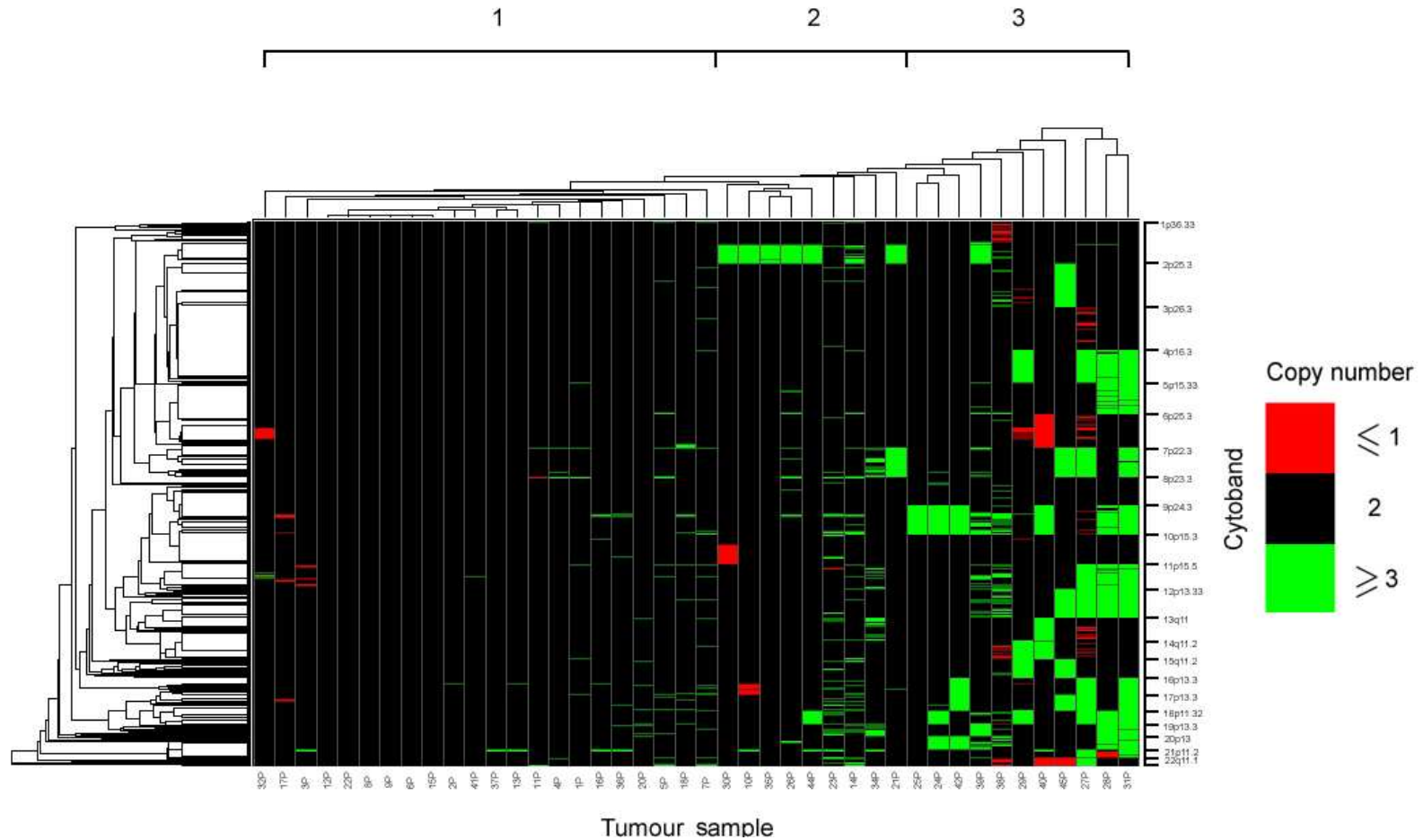
Figure 3.7A: Unsupervised hierarchical cluster heatmap of copy number imbalances in 783 cytobands across the 22 autosomes for 42 primary paediatric ependymomas (Spotfire Decision Site®). Cytobands were imbalanced if ≥ 80 % of the encompassed SNP probes had a copy number of ≤ 1 (loss – red) or ≥ 3 (gain – green). Tumour samples divided into three broad groups. The first group contained samples with few cytoband imbalances or, in six cases, complete absence of imbalance. The second group was characterised by gain of cytobands within the long arm of chromosome one (1q). The third group comprised primary tumours with numerous imbalances, often involving large genomic regions covered by consecutive cytobands.

116

B
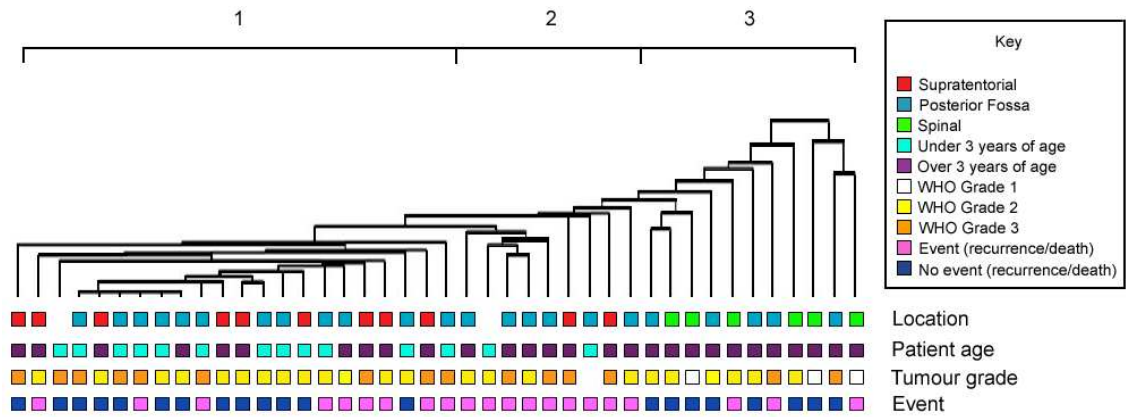
Figure 3.7B**:** The unsupervised hierarchical cluster dendrogram from Figure 3.7A, with clinical information plotted underneath each sample branch. Cluster group one included ependymomas from younger patients compared to the rest of the cohort, reflected in 12/14 (86 %) samples from children under three years of age falling into this cluster. The mean patient age for cluster group one was 4.8 $\pm$ 0.9 years, compared to 10.2 $\pm$ 1.1years for the rest of the cohort. Every group two patient had either relapsed or died, which was reflected in a worse event-free (p = 0.001) and overall survival (p = 0.04) for this group. Group three contained ependymomas from older children and included all of the spinal tumours in the cohort.

C
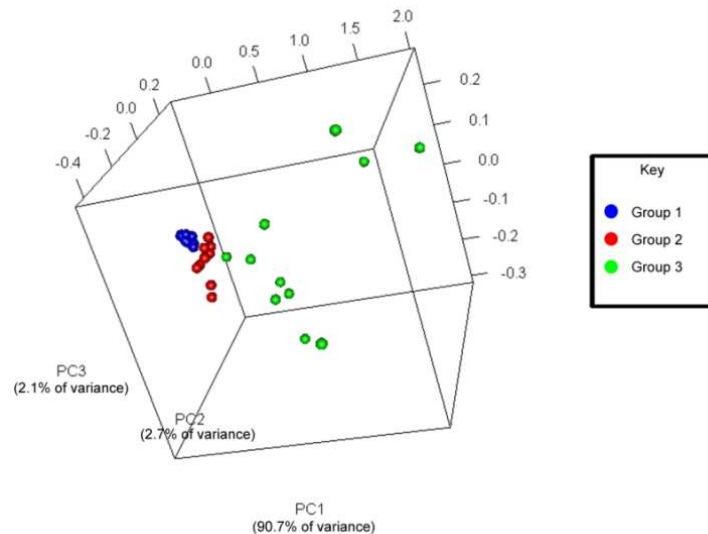
Figure 3.7C: PCA cluster plot of the cytoband imbalance data on 42 primary paediatric ependymomas. The clustering result from unsupervised hierarchical clustering of Figure 3.7A was reinforced by the PCA cluster, demonstrating the three groups previously described in the text and in Figures 3.7A and B.

117

Three groups were identified by both methods (also termed 'group one', 'group two' and 'group three'). The first cluster group contained 22 ependymomas with few genomic aberrations. Indeed, all 15 tumours with a balanced genomic profile were found in this cluster group, as were the majority (9/11) of supratentorial tumours. The patient age of this cluster was significantly lower than the rest of the primary cohort (mean age of 4.8 ± 0.9 years versus 10.2 ± 1.1 years, $p = 0.001$, independent t-test). The second group consisted of nine tumours and was associated with the gain of chromosome 1q or cytobands within chromosome 1q, the most frequently gained region being 1q21.2 in 8/9 cases ($p = 1 \times 10^{-4}$, two-tailed Fisher's exact test). Re-evaluation of SNP probe location using the web-based genomic database Ensembl (http://www.ensembl.org) found that a small proportion of probes, previously denoted as landing within 1q21.2 according to the accompanying annotation file (Netaffx file build 07.12.07) actually extended into the 1q21.3 cytoband. Therefore, for improved precision, the 1q21.2 cytoband was re-defined as the 1q21.2 − 21.3 region. Univariate survival analysis revealed a significantly worse event-free and overall survival for children in this second group compared with the other primary ependymoma patients (estimated mean EFS: 1.8 ± 0.4 years versus 6.6 ± 1.2 years, $p = 0.001$; estimated mean OS: 5.6 ± 1.1 years versus 13.5 ± 2.7 years, $p = 0.041$) (Figure 3.8).
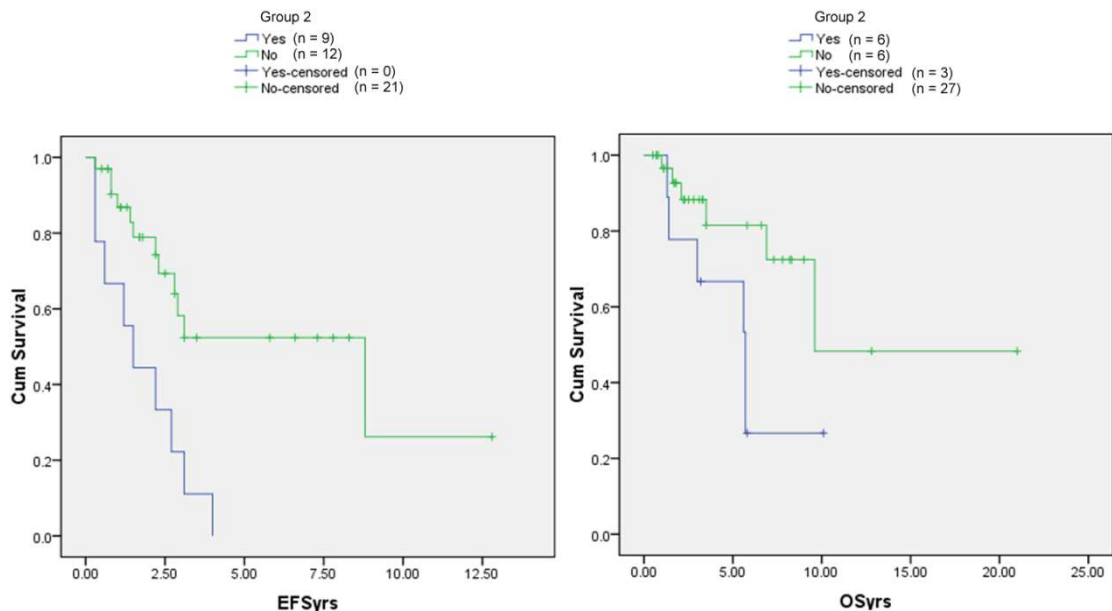
Figure 3.8: Kaplan-Meier EFS and OS curves for nine ependymoma patients from cluster analysis group two (blue lines) compared to the rest of the primary SNP array cohort (green lines). Patients from this group had a significantly worse EFS (percentage from each group attaining five year EFS = 0 % (group two) versus 52.4 ± 11.4 % (remaining primary cohort), $p = 0.001$) and OS (percentage from each group attaining five year OS = 66.7 ± 15.7 % (group two) versus 81.5 ± 8.8 % (remaining primary cohort), $p = 0.041$).

Multivariate survival analysis was also performed. Since commencement dates were unavailable for 13/27 (48 %) patients receiving radiotherapy and almost all patients receiving chemotherapy in the primary cohort, these were removed as variables for event-free survival as the influence on any relapses that occurred in these patients could not be determined accurately. Nevertheless radiotherapy and chemotherapy remained as variables for overall survival. The multivariate analysis showed that patients from group two had a reduced EFS (hazards ratio 5.843 (95 % CI 1.832 – 18.641) p = 0.003), although the adverse effect on overall survival became a trend towards significance (hazards ratio 3.55 (95 % CI 0.971 – 13) p = 0.055) (Tables 3.10 and 3.11). The 1q21.2 – 1.2q1.3 region was also the most frequently gained 1q locus across the entire primary ependymoma cohort (9/42, 21 %). Unsurprisingly, an unfavourable event-free patient survival finding was again observed when comparing all cases with gain against the remaining primary tumours (hazards ratio 4.487 (95 % CI 1.059 – 19) p = 0.042).

Table 3.10: Multivariate event-free survival analysis incorporating group two from the Affymetrix$^{®}$ 500K SNP array analysis.

Cox regression multivariate analysis (n=42)	Event-Free Survival		
Factor	Hazards ratio	95 % CI	P value
Histology (WHO grade II vs III)	1.471	0.607 – 3.564	0.393
Tumour location (PF vs ST)	0.344	0.096 – 1.232	0.101
Resection status (incomplete vs complete)	2.370	0.924 – 6.080	0.073
Patient age (below vs above three years)	2.131	0.544 – 8.341	0.277
Group two (yes vs no)	5.843	1.832 – 18.641	0.003

The primary SNP array cohort of 42 ependymomas was analysed. Radiotherapy and chemotherapy were not included as variables for event-free survival due to missing clinical data regarding commencement dates, as discussed in section 3.2.3. 95 % CI = 95 % confidence interval. ST = supratentorial, PF = posterior fossa. Statistically significant results are highlighted in yellow. Results with a trend towards statistical significance are highlighted in pale yellow.

Table 3.11: Multivariate overall survival analysis incorporating group two from the Affymetrix$^{®}$ 500K SNP array analysis.

Cox regression multivariate analysis (n=42)	Overall Survival		
Factor	Hazards ratio	95 % CI	P value
Histology (WHO grade II vs III)	1.764	0.491 – 6.343	0.385
Tumour location (PF vs ST)	0.595	0.083 – 4.251	0.605
Resection status (incomplete vs complete)	1.024	0.246 – 4.264	0.974
Patient age (below versus above three years)	1.201	0.099 – 14.573	0.886
Radiotherapy (received vs not received)	0.431	0.085 – 2.202	0.312
Chemotherapy (received vs not received)	1.822	0.181 – 18.359	0.611
Group two (yes vs no)	3.554	0.971 – 13.004	0.055

The primary SNP array cohort of 42 ependymomas was analysed. ST = supratentorial, PF = posterior fossa, 95 % CI = 95 % confidence interval. Results with a trend towards statistical significance are highlighted in pale yellow.

The third group comprised of 11 ependymomas characterised by more numerous imbalances, often encompassing large genomic regions. This group was associated with tumours from older children (mean age of 12.2 $\pm$ 1.3 years versus 5.9 $\pm$ 0.8 years, p = 2.9×10^{-4}, independent t-test) and contained all of the spinal ependymomas (p = 8.8×10^{-5}, two-tailed Fisher's exact test).

Survival analysis did not reveal significant results for either group one or three when compared against the remaining tumour cohort. In the case of group three, this was despite all 11 patients being alive, although the mean follow-up time was relatively short at 2.3 years (Figure 3.9).
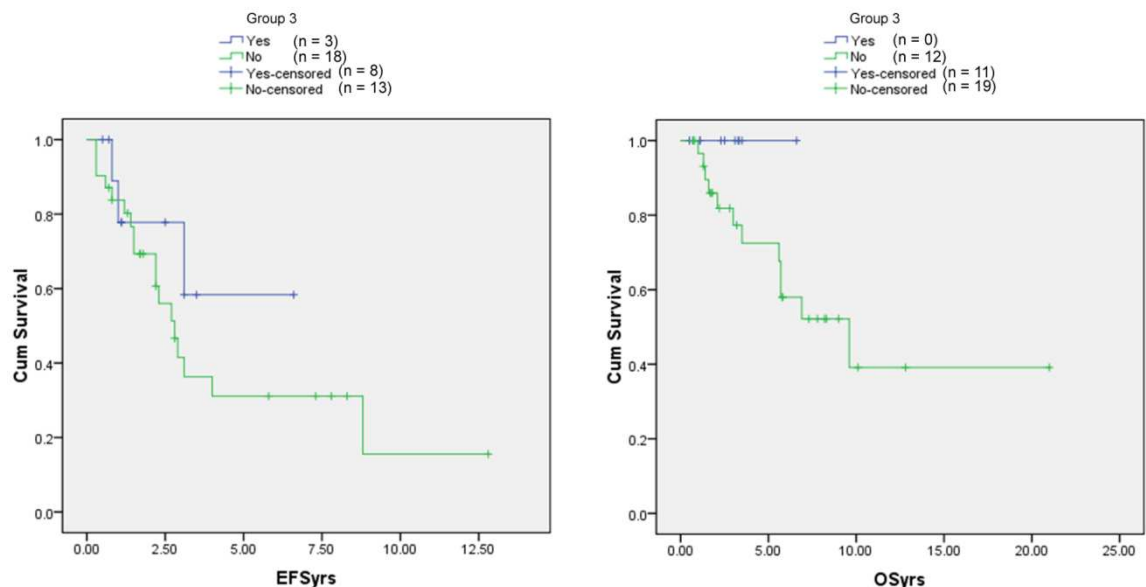
Figure 3.9: Kaplan-Meier EFS and OS curves for 11 ependymoma patients from cluster analysis group three (blue lines) compared to the rest of the primary SNP array cohort (green lines). Although improved, patients from group three did not have a significantly enhanced EFS (percentage from each group attaining five year EFS = 58.3 $\pm$ 19.8 % (group three) versus 31.1 $\pm$ 10 % (remaining primary cohort), p = 0.307). All patients in group three were alive but censored after a relatively short mean follow-up time of 2.3 years. The resulting OS data for this group was thereby not significantly different from the remaining cohort (percentage from each group attaining five year OS = 100 % (group three) versus 72.5 % (remaining primary cohort), p = 0.151). Cum survival = cumulative survival.

3.3.6 FISH validation of chromosome 1q gain results from the SNP array

To validate the chromosome 1q gain result from the 500K SNP array analysis, FISH was performed on 21 primary ependymoma tumour samples that had also been processed on the 500K SNP array, using a commercial LSI 1p36/LSI 1q25 dual colour probe (Vysis, USA) on FFPE based tissue microarrays. The 1q25 FISH probe covered a genomic region of approximately 400kb located within cytoband 1q25.2, encompassing

35 of the SNP probes on the 500K array (SNP_A-2119283 – SNP_A-1800896). During the FISH analysis, gain of 1q25 was defined by the presence of three or more copies of the green (1q25) probe signal per nucleus.

For each of the 21 samples, the percentage of nuclei counted revealing 1q25 gain by FISH was initially correlated with the SNP array derived copy number for the 400kb FISH probe region (Figure 3.10) and subsequently the presence of entire chromosome 1q arm gain, as defined by the 80% chromosome arm SNP probe threshold described in section 3.2.1.2 (Table 3.12). This analysis found that the presence of 1q25 gain in over 30 % of nuclei counted by FISH correlated strongly with SNP array gain of the corresponding FISH probe region (R = 0.79, p = 2.49×10^{-5}, Spearman's correlation), although a positive correlation also existed with SNP derived gain of the whole chromosome 1q arm (Table 3.12).

Table 3.12: Correlation of chromosome 1q SNP array data and 1q25 FISH scores for 21 primary paediatric ependymomas.

Tumour sample ID	SNP copy number of region on 1q25.2 covered by FISH probe	Imbalance of chromosome 1q derived from SNP data based on 80% threshold (actual percentage)	Percentage of nuclei counted demonstrating 1q25 gain by FISH
1P	2	No imbalance	8 %
2P	2	No imbalance	10 %
3P	2	No imbalance	13 %
8P	2	No imbalance	11 %
9P	2	No imbalance	10 %
13P	2	No imbalance	9 %
22P	2	No imbalance	6 %
23P	2	No imbalance	13 %
27P	2	No imbalance	8 %
29P	2	No imbalance	18 %
34P	2	No imbalance	7 %
36P	2	No imbalance	11 %
38P	2	No imbalance	27 %
40P	2	No imbalance	3 %
42P	2	No imbalance	10 %
10P	3	Gain	33 %
14P	3	No imbalance	38 %
26P	3	Gain	56 %
30P	3	Gain	31 %
35P	3	Gain	40 %
39P	3	Gain	52 %
Spearman's correlation coefficient in relation to FISH score	R = 0.79, p = 2.49×10^{-5}	R = 0.7, p = 3.76×10^{-4}	

Note: Sample 14P is highlighted as, although it did not exceed the chromosome arm imbalance threshold of 80 %, 69 % of the SNP probes within chromosome 1q were gained for this sample. Key: P = primary, R = Spearman's correlation coefficient.
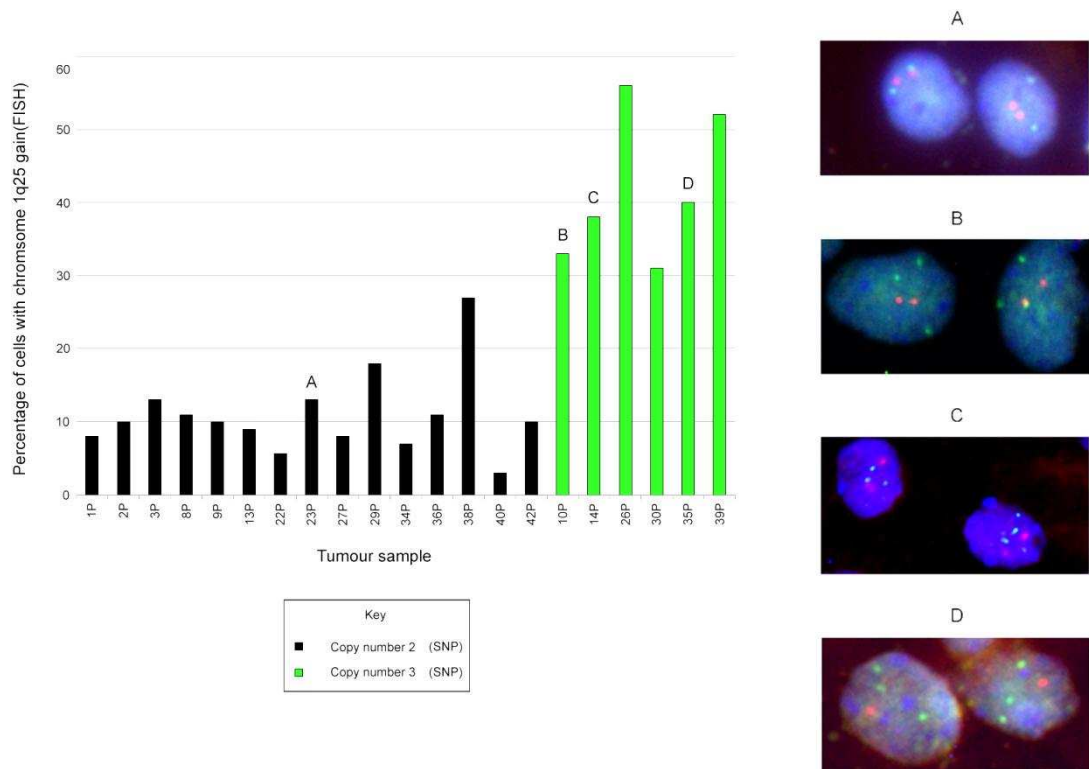
Figure 3.10: Chromosome 1q25 FISH validation of the 500K SNP array findings for 21 primary ependymomas. The colour of each sample bar represents the SNP array derived copy number for a region of 35 SNP probes on chromosome 1q25.2 corresponding to the region covered by the LSI 1p36/LSI 1q25 dual colour FISH probe (Vysis). Samples with (green) or without (black) copy number gain of the region can be seen. Each bar's height represents the percentage of nuclei with copy number gain from the 1q25 FISH analysis of that sample, where gain was defined by the presence of three or more copies of the 1q25 probe signal per nucleus. A demonstration of 1q25 gain in over 30 % of nuclei scored using FISH correlated with SNP derived copy number gain of the probe region (Spearman's Rank = 0.79, p= 2.49x10^{-5}). Examples of 1q25 FISH for one diploid sample (A) and 3 gained samples (B – D) are shown (magnification x 40), where green signal = 1q25 probe and red signal = 1p36 probe.

3.3.7 Methylation array cluster analysis identifies clinically relevant subgroups

The corrected methylation scores for the 1,421 GoldenGate® Cancer Panel I CpG probes for all 98 tumours comprising the ependymoma methylation cohort are shown in Appendix 10F. Since normal, control tissue (such as paediatric brain tissue) was not available for this analysis, only comparisons between tumour groups were possible. Bootstrapped unsupervised hierarchical clustering and PCA were performed on the methylation array primary tumour cohort of 73 ependymomas, in order to identify distinct molecular tumour subgroups and potentially associate them to known clinical and/or pathological demographic factors (Figures 3.11A – C).

122

Bootstrapped clustering revealed two main sample clusters with an approximate unbiased probability (p) value of 99 % (orange and yellow shaded groups in Figures 3.11A and 3.11B). The first cluster of 13 tumours (shaded orange) contained predominantly spinal ependymomas, while the second cluster of 59 tumours (shaded yellow) was composed mainly of intracranial ependymomas ($p < 0.0001$, two-tailed Fisher's exact test) (Figure 3.11B). The mean patient age of this first cluster was older than the second cluster (10.3 ± 4.1 years versus 5.9 ± 4.5 years, $p = 0.003$, independent t-test). The large intracranial cluster also contained numerous subgroups, although these did not attain approximate unbiased probability values above 95 % on bootstrapped cluster analysis. For example, a large cluster of 34 posterior fossa ependymomas demonstrated a similar methylation profile (Figure 3.11B), yet contained a distinct subgroup of nine tumours from children aged under three years ($p = 0.01$, two-tailed Fisher's exact test), the majority of which (6/9, 67% cases) had an anaplastic histology (circled red in Figure 3.11B). In addition, a cluster of five supratentorial ependymomas (circled green in Figure 3.11B) demonstrated a distinct methylation profile to that of the other supratentorial tumours in the cohort, suggesting biological sub-classification may also exist for supratentorial ependymomas, despite sharing the same locality within the central nervous system. Survival analysis revealed no significant differences in patient outcome between the clusters identified above, while ependymomas with a balanced genomic profile from the SNP array analysis did not cluster independently on the methylation analysis. Genes found to be differentially methylated between particular clusters identified in this chapter, or between selected clinical subgroups are subsequently detailed in Chapter 4, section 4.3.4.

A
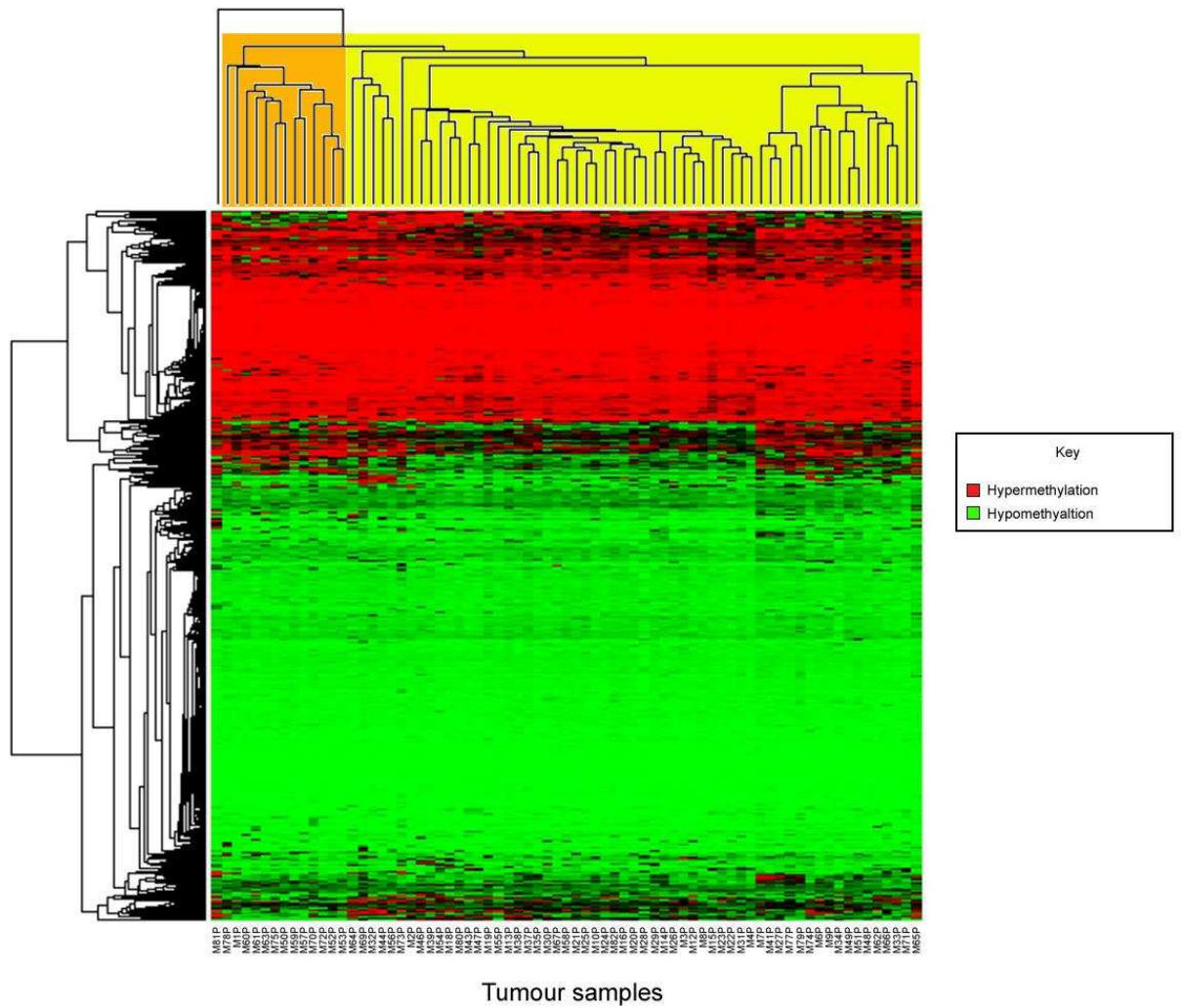
Figure 3.11A: Bootstrapped, unsupervised hierarchical cluster heatmap of the corrected Beta methylation scores across 1,421 GoldenGate® Cancer Panel I CpG probes for 73 primary paediatric ependymomas (10,000 replications). Red on the heatmap represents hypermethylated probes while green represents hypomethylation. Two main clusters revealed an approximate unbiased probability (p) value greater than the 95 % significance threshold and are shaded in orange and yellow for discrimination. The clinical demographics of the samples in these clusters are demonstrated in Figure 3.11B.

B

Clinical aspects of tumour samples
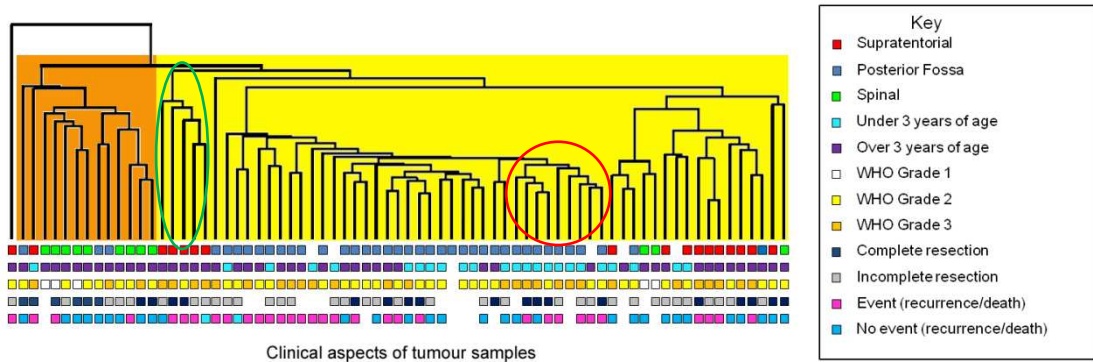
Figure 3.11B: The unsupervised hierarchical cluster dendrogram from Figure 3.11A, with clinical information plotted underneath each sample branch. Cluster one (orange) included predominantly spinal ependymomas from older children while cluster two (yellow) was comprised mainly of intracranial tumours (p < 0.0001, two-tailed Fisher's exact test). Within the intracranial methylation cluster were discrete subclusters including a group of nine posterior fossa tumours from patients aged below three years (p = 0.01, two-tailed Fisher's exact test), two-thirds of which had an anaplastic histology (circled red). In addition, a discreet subcluster of five supratentorial tumours (circled green) was evident, with a different methylation profile to the majority of other supratentorial tumours in the primary cohort.

C
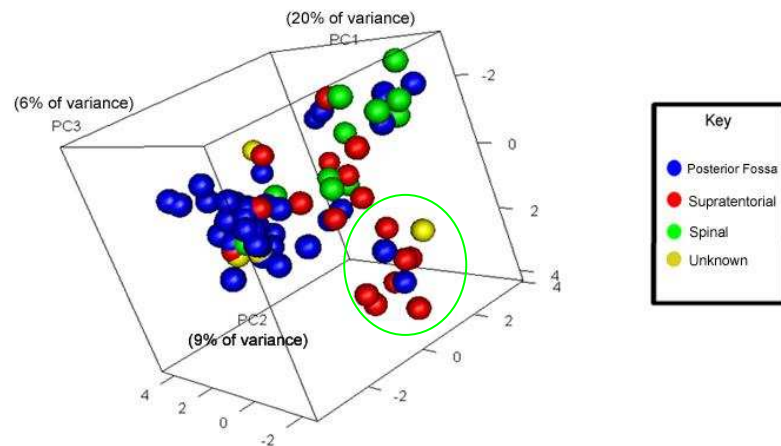
Figure 3.11C: PCA cluster plot of the corrected Beta methylation scores across the 1,421 GoldenGate® Cancer Panel I CpG probes for 73 primary paediatric ependymomas. Samples are labelled according to tumour location within the central nervous system (see figure key). The methylation cluster profiles demonstrated by the bootstrap analysis were reinforced by PCA, with the majority of spinal ependymomas grouping separately to their intracranial counterparts. The clustering of posterior fossa tumours is evident, while subtle methylation profile differences between supratentorial ependymomas results in a supratentorial tumour predominant subcluster formation (circled again in green), reflecting differences already noted for this tumour group from the bootstrapped cluster analysis.

125

3.4 Discussion

Our current inability to predict the clinical sequelae of paediatric ependymomas based on traditional histological and clinical information underlines the distinct management challenge presented by these tumours. An improved understanding of ependymoma biology in children will help to redress this issue. Two different array-based analyses were utilised in this study. The 500K Affymetrix[®] SNP array was used to assess genomic imbalance and loss of heterozygosity in 42 primary and 21 recurrent paediatric ependymomas. This is the first 500K SNP array study of ependymomas to normalise tumour data against patient matched constitutional DNA in order to identify tumour-specific alterations and account for genomic alterations already present in the patient germline. The Illumina[®] GoldenGate[®] Cancer Panel I assay for methylation analysed 73 primary and 25 recurrent paediatric ependymomas, reporting the methylation status at 1,421 CpG sites mapping to 768 genes associated with DNA methylation and/or cancer. It is the first reported epigenetic array study performed on paediatric ependymomas.

For both array platforms, results from the tumour cohorts were initially analysed for broad genomic and epigenetic changes which were correlated with clinical and histological factors, together with patient outcome. The results demonstrated distinct differences between paediatric ependymomas, particularly relating to tumour location within the CNS, patient age at diagnosis and tumour recurrence. Moreover, the SNP array work also identified genomic gain involving chromosome 1q to be associated with an adverse patient prognosis. Higher resolution gene-level analysis was subsequently performed on both ependymoma cohorts, with the results discussed in later chapters.

The two sets of paediatric ependymomas used for both microarray platforms, and the patients from which they were obtained, generally reflected the clinical profile reported for this tumour group in the literature (Kilday, Rahman et al. 2009; Wright and Gajjar 2009). The vast majority of paediatric ependymomas in both cohorts were of intracranial origin, with primary posterior fossa tumours accounting for the majority of the total intracranial cohort. Almost half of all the patients analysed were below five years of age at diagnosis, while approximately one third of children from both cohorts (33 % of the SNP cohort, 30 % of the methylation cohort) were aged below 36 months.

The predilection of ependymomas from these very young patients for a posterior fossa locality and the association of spinal tumours with older children (Dyer, Prebble et al. 2002; Kilday, Rahman et al. 2009) were also evident in both groups. Most of the primary ependymomas in each cohort were of classic histology, in keeping with several other series (Rousseau, Habrand et al. 1994; Perilongo, Massimino et al. 1997; Horn, Heideman et al. 1999; Grundy, Wilne et al. 2007), while almost 50 % were deemed to have had a complete surgical resection. From literature, the reported resection rate ranges from 31 – 85 %, reflecting diverse surgical techniques, inconsistent criteria for defining complete tumour excision and variable surgical accessibility of ependymomas from different CNS locations (Rousseau, Habrand et al. 1994; Bouffet, Perilongo et al. 1998; Merchant and Fouladi 2005). A slight predominance of male patients was present (male to female ratios of 1.4:1 and 1.3:1 for SNP and methylation cohorts respectively) which concurred with several other sizeable studies (Goldwein, Leahy et al. 1990; Horn, Heideman et al. 1999; Grundy, Wilne et al. 2007).

While the five year overall survival for both cohorts of 77 % ($\pm$ 8 % (SNP cohort); $\pm$ 7 % (methylation cohort)) was higher than the 39 – 71 % generally reported (Robertson, Zeltzer et al. 1998; Paulino, Wen et al. 2002; Zacharoulis, Levy et al. 2007; Wright and Gajjar 2009), the five year event-free survival rates (36 $\pm$ 9 % (SNP cohort); 37 $\pm$ 8 % (methylation cohort)) and the mean time to relapse (2.4 $\pm$ 0.5 years (SNP cohort); 2 $\pm$ 0.3 years (methylation cohort)) replicated the findings from current literature (Pollack, Gerszten et al. 1995; Perilongo, Massimino et al. 1997; Robertson, Zeltzer et al. 1998; Horn, Heideman et al. 1999; Agaoglu, Ayan et al. 2005; Zacharoulis, Levy et al. 2007). The deleterious effect of incomplete surgical resection on event-free survival was evident from analysis of both SNP and methylation array patient groups, supporting complete resection as a consistently reported favourable clinical prognostic factor (Bouffet, Perilongo et al. 1998; Robertson, Zeltzer et al. 1998; Horn, Heideman et al. 1999; Merchant, Li et al. 2009). However, this effect did not translate into a worse overall survival in either patient cohort.

3.4.1 Aberrations in paediatric ependymomas from different CNS locations

Both SNP and methylation array analyses highlighted distinct differences in the genomic and epigenetic profiles of paediatric ependymomas originating from different sites within the CNS, often corroborating the findings of preceding lower resolution, low throughput studies (sections 3.3.4, 3.3.5 and 3.3.7).

Paediatric spinal ependymomas were characterised by numerous arm or whole chromosome genomic anomalies including gains of chromosome 17p, 12q, 21q, 16, 9, 18 and 20. Diverse and extensive chromosome imbalances were also a feature of paediatric spinal ependymomas from the CGH meta-analysis discussed previously (Chapter 1, section 1.5.2) (Kilday, Rahman et al. 2009). While several imbalances were different between the two spinal tumour cohorts, the frequent gain of chromosomes 18, 20 and 9 were common to both. The importance of chromosome 9 gain in spinal ependymomas has been highlighted by a recent integrated genome-wide analysis of RNA expression and DNA copy number in adult and paediatric ependymomas, as copy number gain correlated with a general increase in the level of gene expression across the chromosome (Johnson, Wright et al. 2010). This is in keeping with studies of CDKN2A (p14ARF), a putative tumour suppressor gene located at 9p21.3 that regulates neural stem cell proliferation and whose deletion has been shown to rapidly expand progenitor cell numbers in developing neural tissue (Taylor, Poppleton et al. 2005). Whilst FISH and microsatellite analysis have demonstrated that CDKN2A deletion is virtually exclusive to supratentorial ependymomas (Taylor, Poppleton et al. 2005; Schneider, Monoranu et al. 2009), gene expression analyses have revealed downregulation of CDKN2A in intracranial tumours, yet not spinal counterparts (Korshunov, Neben et al. 2003; Mendrzyk, Korshunov et al. 2006).

Deletion of chromosome 22q was the most frequent genomic loss in the SNP array spinal cohort (2/6 cases), a finding that requires confirmation in a larger cohort of paediatric spinal ependymomas. Nevertheless, the aforementioned integrated genomic and expression analysis performed by Johnson et al. found that spinal ependymomas frequently lost chromosome 22q and had an expression signature enriched with genes deleted and underexpressed on chromosome 22q such as TNRC6B (22q13.1) (Johnson,

Wright et al. 2010). A sub-region implicated in spinal ependymoma pathogenesis is 22q12, harbouring the Protein 4.1 superfamily member NF2. Mutation analysis has revealed that, in association with LOH for chromosome 22, somatic NF2 mutations exist but in a proportion of ependymomas restricted exclusively to spinal cases (Ebert, von Haken et al. 1999; Lamszus, Lachenmayer et al. 2001). In addition, NF2 mutations are seen in patients with Neurofibromatosis type II, a cancer predisposition syndrome regularly associated with the occurrence of spinal ependymomas (Kulkarni 2004). Indeed, one of the two spinal ependymomas with 22q loss from this analysis, sample 38P, was obtained from a patient diagnosed with neurofibromatosis type II, suggesting a tumour suppressor role for this gene in ependymomas of the spinal cord. Genetic alterations of other Protein 4.1 family members have also been reported for extracranial ependymomas, such as 4.1B deletion and 4.1R underexpression (Singh, Gutmann et al. 2002; Rajaram, Gutmann et al. 2005). This SNP array study also revealed however, that deletion of chromosome 22q was not exclusive to spinal ependymomas, lending support to the CGH meta-analysis which revealed chromosome 22 loss was a common feature of paediatric intracranial ependymomas (Chapter 1, section 1.5.2.2) (Kilday, Rahman et al. 2009).

The SNP array analysis additionally highlighted location specific genomic differences between primary paediatric intracranial ependymomas. Posterior fossa tumours were associated with gain of chromosome 1q, particularly the regions 1q21.3, 1q22, 1q25.3, 1q32.2 and 1q42.12 – 44. Supratentorial tumours only demonstrated gain of chromosome 19p in a single sample, a region gained in other cerebral tumours such as CNS PNETs (Pfister, Remke et al. 2007). This contrasts with the CGH meta-analysis of paediatric ependymomas which confirmed 1q gain as the most frequent genomic imbalance in both posterior fossa and supratentorial ependymomas (Chapter 1, section 1.5.2.2) (Kilday, Rahman et al. 2009). The lack of 1q gain in the supratentorial tumours of the SNP array cohort may be explained by relatively small sample numbers, although the 500K integrated SNP analysis of 204 ependymomas also observed that supratentorial tumours were characterised by numerous focal copy number aberrations, rather than broad imbalances (Johnson, Wright et al. 2010). The CGH meta-analysis further revealed that posterior fossa ependymomas in children often demonstrate loss of chromosomes 22, 6 and 17 while at recurrence, the loss of both 6q and 10q appeared exclusive to infratentorial tumours (Chapter 1, section 1.5.2.2) (Kilday, Rahman et al.

2009). This latter finding was not reflected in the SNP array analysis where chromosome 6 loss was seen in a spinal recurrence, albeit of a primary posterior fossa ependymoma (samples 40P and 40R1).

The epigenetic array analysis also demonstrated a clear segregation of spinal from intracranial ependymomas. This categorisation reflects the results of several gene expression profile analyses which have shown differential expression profiles between intracranial and extracranial ependymomas (Chapter 4, section 4.1). (Korshunov, Neben et al. 2003; Taylor, Poppleton et al. 2005; Lukashova-v Zangen, Kneitz et al. 2007; Palm, Figarella-Branger et al. 2009; Johnson, Wright et al. 2010). The methylation study further revealed distinct profile differences both between and within posterior fossa and supratentorial ependymoma tumour groups, again supporting previous gene expression work. For instance, in addition to finding clearly defined expression signatures for supratentorial and posterior fossa ependymomas, one study of 103 tumours identified three infratentorial expression subgroups including one characterized by numerous DNA amplifications and one where the samples had recurrent genomic gains of chromosome 1q (Taylor, Poppleton et al. 2005). Another analysis of 83 ependymomas identified eight intracranial ependymoma subgroups based on their expression profiles (Johnson, Wright et al. 2010). The findings from this study therefore suggest that the distinct expression profiles identified in such subgroups may be a manifestation of dysregulated gene methylation mechanisms. Differentially methylated genes between ependymomas from the different CNS locations are examined in the next chapter.

This analysis has contributed further to evidence demonstrating biological distinctions between ependymomas from different locations within the CNS. One explanation for this is that the individual tissue environment of each CNS location allows certain genetic aberrations to prosper whilst restricting the development of others (Hirose, Aldape et al. 2001). Evidence that the signature genes of supratentorial and spinal ependymomas are reciprocally expressed by murine embryonic radial glia in the corresponding CNS compartments (Taylor, Poppleton et al. 2005; Poppleton and Gilbertson 2007), has also led to the view that ependymomas from different locations within the CNS are derived from regionally (and developmentally) specified neural progenitor cells, presumed to be radial glia, that have undergone malignant

transformation (Taylor, Poppleton et al. 2005; Johnson, Wright et al. 2010) as discussed previously in Chapter 1, section 1.1.1. This concept is examined again in the final chapter.

3.4.2 Aberrations in ependymomas from younger and older children

This work also identified biological disparity between ependymomas from young and old paediatric patients (sections 3.3.4, 3.3.5 and 3.3.7). The comparison of genomic imbalances between primary ependymomas categorised according to a patient age threshold of three years was performed exclusively on the posterior fossa cohort of 24 ependymomas, since tumour location was standardised and half of this subgroup fell either side of the age cut-off. Three years was selected as the threshold since this is the age at which craniospinal radiotherapy is routinely introduced as a therapeutic regime for paediatric ependymoma in the United Kingdom. While neither age category was associated with a particular broad imbalance, the number of chromosome arm imbalances was significantly fewer in the younger posterior fossa cohort. This finding refined the numerical difference in imbalances found between 116 paediatric and 187 adult primary ependymomas from the CGH meta-analysis (707 paediatric versus 869 adult anomalies; Chapter 1, section 1.5.2.1) (Kilday, Rahman et al. 2009).

Unsupervised clustering of cytoband copy number data from the primary tumour cohort of the SNP array analysis elucidated this age discrepancy further and identified three broad genomic subgroups similar to those found in preceding lower resolution analyses (Dyer, Prebble et al. 2002; Mendrzyk, Korshunov et al. 2006; Puget, Grill et al. 2009). The first cluster group contained samples with very few or, in six cases, absolutely no cytoband imbalances present. All 15 tumours (36 %) with a defined balanced SNP profile and the majority of ependymomas from children aged below three years were found in this cluster group, which had a mean patient age of 4.8 ± 0.9 years, compared to 10.2 ± 1.1 years for the rest of the primary cohort. This supported the conventional and array CGH finding that a balanced genomic profile without chromosomal gain or loss, whilst uncommon in adult ependymomas, could be seen in $36 - 58$ % of paediatric cases and was particularly associated with children under three years of age (Reardon, Entrekin et al. 1999; Hirose, Aldape et al. 2001; Ward, Harding et al. 2001; Carter,

Nicholson et al. 2002; Dyer, Prebble et al. 2002; Modena, Lualdi et al. 2006). The second group was characterised by infrequent imbalances and gain of cytobands within chromosome 1q. In contrast to the first cluster group, the third subgroup showed frequent, large genomic anomalies, similar to those seen in the CGH meta-analysis of adult and spinal ependymomas (Chapter 1, section 1.5.2) (Kilday, Rahman et al. 2009). Indeed, all of the spinal tumours in the SNP analysis were located in this third group which had a relatively older mean patient age of 12.2 $\pm$ 1.3 years compared to the remaining primary cohort.

Bootstrapped unsupervised hierarchical cluster analysis of the methylation array data also grouped the ependymoma samples into two cohorts with significantly different mean patient ages, although this may have reflected the high proportion of spinal tumours in the older cluster and intracranial tumours in the younger cluster. Nevertheless, such clustering further isolated a small subgroup of nine ependymomas within a larger cluster of 34 posterior fossa tumours which were from children aged below three years, suggesting ependymoma methylation profiles may differ according to patient age.

These global array findings demonstrate clear genomic and potential epigenetic differences between ependymomas from younger and older children. Ependymomas with a balanced genomic profile are often seen in children below three years of age, while the number of genomic imbalances that occur escalates with increasing patient age at diagnosis. At present, the explanations for this remain unclear. A difference in the number of genomic aberrations between ependymomas from both age groups may reflect the increasing genomic instability seen during the ageing process (Lombard, Chua et al. 2005) and appears consistent with a canonical multi-step cancer initiation/progression model for ependymomas from older children and adults. It also suggests that fewer defects in cell regulating processes are needed to initiate ependymoma in young children, since the behaviour of cells from immature normal tissue is similar to that of cancer cells with respect to differentiation, survival and self-renewal (Knudson 1971; Scotting, Walker et al. 2005). Paediatric ependymomas demonstrating balanced genomic profiles or subtle imbalances substantiate this theory. This would imply that the genes affected by mutation in the younger age group are potent oncogenes or tumour suppressor genes, or are genes responsible for regulating

cell differentiation and self renewal that only possess an oncogenic effect restricted to a developmentally determined temporal window, environment, or indeed cell type as purported by comparative human and murine ependymoma expression work (Johnson, Wright et al. 2010).

Alternatively, epigenetic phenomena may be affecting the expression profiles of an unknown number of genes in ependymomas from young children by altering their transcription and expression without a corresponding detectable genomic imbalance. The observation that tumours with a balanced SNP genome did not cluster independently on the GoldenGate® Cancer Panel I analysis reduces the possibility of methylation being a responsible mechanism, although the relatively low resolution and design of the current array for pre-selected genes cannot preclude it. Novel methylation arrays may help to address this uncertainty by interrogating over 27,000 CpG sites to provide a more comprehensive map of the tumour epigenome. Furthermore, the involvement of alternative epigenetic mechanisms such as histone deacetylation remains plausible and warrants further consideration (Peyre, Commo et al. 2010).

3.4.3 Aberrations in potential prognostic groups

3.4.3.1 Cytoband imbalance group two / chromosome 1q gain

The most frequent chromosome arm imbalance seen in the primary SNP array cohort was gain of chromosome 1q (7/42 tumours, 17 %), where it was associated with a posterior fossa location and a worse overall patient survival on multivariate analysis. It was also the most frequent aberration seen in the intracranial first recurrent cohort (3/9 tumours, 33 %) (section 3.3.4). Moreover, the unsupervised hierarchical clustering of cytoband copy number data across the primary tumour cohort identified a set of ependymomas characterised by gain of either the entire arm of chromosome 1q or cytobands confined to this region (group two) (section 3.3.5). Such tumours were associated with adverse outcome on univariate survival analysis and tumour progression on multivariate analysis. Whilst mirroring the CGH meta-analysis findings which demonstrated 1q gain as the most frequent anomaly in paediatric primary and recurrent ependymomas (Chapter 1, section 1.5.2) (Kilday, Rahman et al. 2009), the SNP array

results also imply a role for this aberration as a marker of both disease progression and ultimately, patient survival in this tumour group. This thereby lends support to the proposal of chromosome 1q gain, in conjunction with CDKN2A deletion, as detected by aCGH and FISH, being an adverse risk stratification marker that could be introduced into forthcoming clinical trials for ependymomas regardless of patient age (Korshunov, Witt et al. 2010).

Conventional cytogenetic analyses have tried to decipher the mechanism for 1q gain in paediatric ependymoma (Kramer, Parmiter et al. 1998; Mazewski, Soukup et al. 1999) (Chapter 1, section 1.5.1). They have suggested that chromosome 1q gain results from a variety of unbalanced rearrangements with material from numerous partner chromosomes rather than a single, recurrent translocation. Therefore, a dosage effect for genes on 1q could be more important for initiating and propagating tumour growth than the influence of a translocated regulatory gene from a second chromosome. Furthermore, it has been hypothesized that the association between partial chromosomal rearrangement, such as chromosome 1q gain and poor prognosis in paediatric ependymoma could be anticipated since the acquisition of partial imbalances may have biological effects that exceed those preserving a broad genomic balance across individual chromosomes (Dyer, Prebble et al. 2002).

As chromosome 1q spans a relatively large genomic area, previous cytogenetic and genomic studies have attempted to localise important regions of imbalance within chromosome 1q in paediatric ependymoma, associating gain of the region 1q21 – 32 with an anaplastic histology, tumour recurrence and adverse prognosis (Carter, Nicholson et al. 2002; Dyer, Prebble et al. 2002; Mendrzyk, Korshunov et al. 2006). Aberrations within this genomic sub-region were also observed by the SNP array cytoband analysis. In particular, gain of the region 1q21.2 – 21.3 was identified as the most frequent regional imbalance within cluster analysis group two which was associated with an adverse patient outcome (section 3.3.5). Indeed, it was the most frequently gained chromosome 1q locus across the primary cohort (9/42, 21 %), being associated independently with an adverse event-free survival. It was also the most common region of gain from cytoband analysis of the entire recurrent cohort (9/21 cases, 43 %) and one of the most frequent aberrations in the intracranial first recurrent cohort (4/9 cases, 44 %). This suggests that gain of 1q21.2 – 21.3 could be implicated in

ependymoma progression and thus patient prognosis, warranting further analysis to establish candidate oncogenes. This is explored in subsequent chapters. Literature supports this theory, with gain of 1q21 associated with adverse prognosis in haematological malignancies (Fonseca, Van Wier et al. 2006) and solid cancers such as ovarian cancer (Kudoh, Takano et al. 1999), where it has been associated with disease recurrence in tumours resistant to cisplatin, a chemotherapeutic agent used regularly in paediatric ependymoma management.

3.4.3.1.1 Refining the SNP definition of 1q gain

In this SNP array analysis, an ependymoma was defined as having gain of chromosome 1q if 80 % or more of the 19, 252 SNP probes mapping to this arm demonstrated a copy number greater than two. This imbalance threshold was chosen as an unambiguous marker of true regional genomic gain or loss. FISH validation of chromosome 1q gain was performed on a cohort of 21 tumours using a LSI 1p36/1q25 dual colour probe (section 3.3.6). The presence of chromosome 1q gain as denoted by SNP analysis was found to correlate with the presence of 1q25 gain in over 30 % of nuclei counted by FISH ($R = 0.7$, $p = 3.76 \times 10^{-4}$) which validated the 5/21 ependymomas with SNP array evidence of chromosome 1q gain. However, one tumour sample (14P) revealed 1q25 FISH gain in 38 % of nuclei counted, yet did not exhibit SNP array evidence of arm imbalance. Retrospective analysis of this sample revealed that 69 % of the SNP probes confined to chromosome 1q had a copy number of three or above, below the designated 80 % threshold but still relatively high. Consequently, a higher correlation was achieved if the FISH percentage scores were compared to the copy number of the SNP probes confined to the actual region on 1q25.2 covered by the FISH probe (where gain in sample 14P was accounted for), or alternatively, if the arm imbalance SNP threshold was dropped from 80 % to 50 %. ($R = 0.79$, $p = 2.49 \times 10^{-5}$). This collectively suggests that the accuracy of defining chromosome arm imbalance in SNP array analysis could be improved by lowering the proportion of encompassed SNP probes demonstrating copy number gain to 50 %. Indeed, such a threshold has been used recently in other large SNP array analysis of paediatric and adult high grade gliomas and ependymomas (Johnson, Wright et al. 2010; Paugh, Qu et al. 2010).

3.4.3.2 Cytoband imbalance group three

Survival analysis was performed on all three primary ependymoma groups established from unsupervised cluster analysis of cytoband copy number data from the SNP array study (section 3.3.5). As with the first cluster, the 11 older patients with ependymomas comprising group three, incorporating all the spinal cases and characterised by numerous and often large genomic imbalances, did not demonstrate a statistically significant difference in event-free or overall survival compared to the remaining cohort ($p = 0.307$ and 0.151 respectively). This contrasted with the results of CGH and DNA cytometry studies which have suggested that a high number of genomic anomalies in ependymoma are associated with tumours of a lower histological grade and a favourable patient outcome (Scheil, Bruderlein et al. 2001; Carter, Nicholson et al. 2002; Gilhuis, van der Laak et al. 2004; Louis, Ohgaki et al. 2007).

It has been argued that the association of numerous, sizeable copy number aberrations with improved patient outcome could reflect intermediate ploidy involving specific whole chromosomal imbalance patterns (Carter, Nicholson et al. 2002). This phenomenon has been associated with favourable outcome in other paediatric neoplasms such as acute lymphoblastic leukaemia (Chessels, Swansbury et al. 1997) in addition to spinal ependymomas, which predominate in group three of this SNP array analysis (Ebert, von Haken et al. 1999). Studies have also suggested that cells of intermediate ploidy exhibit increased chemosensitivity compared to cells of a near diploid state (Synold, Relling et al. 1994; Kaspers, Veerman et al. 1995). Indeed, one particular CGH analysis of 53 paediatric intracranial ependymomas found that 'numerical' tumours characterised by 13 or more chromosomal imbalances had an improved overall survival when compared to tumours with a balanced genomic profile (Dyer, Prebble et al. 2002) (Chapter 1, section 1.5.2.1).

There are explanations for the discrepancy in results between this study and the work of Dyer et al. The average number of arm or whole chromosome imbalances seen in the third group from the SNP array analysis was five, below the threshold of 13 set in the CGH study. Indeed, none of the ependymomas in the SNP array cohort would have had enough broad imbalances to be classified as 'numerical' in the Dyer study, placing them

in an inferior prognostic category. The categorisation adopted by an aCGH analysis of 68 ependymomas (Mendrzyk, Korshunov et al. 2006), where 'numerical' samples were those with greater than two imbalances, may therefore be more appropriate for the SNP array cohort examined. Moreover, all of the patients in group three of the present study remained alive although several of the cases were relatively new diagnoses, with a short mean follow up time of 2.3 years. This implies that a significant improvement in overall survival for this patient subgroup may be realised if their follow up period was extended to match the rest of the primary tumour cohort, in keeping with the findings from literature noted above.

3.4.3.3 Other aberrations in recurrent intracranial ependymomas

Although the SNP array analysis found that gain of either the entire chromosome 1q arm or confined sub-regions within it were associated with paediatric ependymoma recurrence and a worse relapse free patient survival, other genomic imbalances with a potential role in disease progression were also noted (sections 3.3.4 and 3.3.5).

The gain of chromosome 9 was a frequent occurrence in the SNP array cohort, including the entire recurrent cohort where it was observed in 5/21 cases (24 %). Gain of chromosome 8 was observed exclusively in intracranial recurrent tumours of the SNP array cohort, being significantly associated with the first recurrent cohort when compared to primary tumours. In one posterior fossa ependymoma patient (samples 26P – R3), gain of both chromosomes 8 and 9 was acquired at recurrence and maintained through three subsequent local recurrences. In another case (sample 40P and 40R1), gain of chromosome 9 was present in the primary posterior fossa tumour and maintained in the spinal recurrence.

Deletion of chromosome 22q was the most frequent chromosome arm loss in the primary cohort (3/42 cases, 7 %), occurring in both intracranial and extracranial tumours. However, 22q deletion was also present in the intracranial first recurrent cohort (1/9 cases, 11 %), with 22q13.1 – 13.31 and 22q13.33 loss being significantly associated with this recurrent group. Furthermore, chromosome 22q loss was identified as another maintained aberration in the spinal recurrence (40R1) of the primary

posterior fossa sample 40P, along with chromosome 6 loss and the aforementioned chromosome 9 gain.

These findings indicate that imbalances involving chromosomes 1q, 6, 9 and 22 may be implicated in primary ependymoma tumourigenesis and subsequent tumour maintenance. Furthermore these chromosomal anomalies, along with the acquired gain of chromosome 8, may also be implicated in disease progression and therapeutic resistance. Many of these genomic aberrations are examined at a higher resolution and discussed further in the fifth chapter which explores maintained and acquired genomic alterations in primary and recurrent paediatric ependymoma paired samples.

3.5 Summary

Paediatric ependymomas continue to present a management challenge, largely due to the difficulty in predicting a clinical course from current histopathological and putative prognostic markers. Indeed, present treatment regimens for childhood ependymomas in the United Kingdom are stratified according to patient age and degree of tumour resection, both of which have been inconsistently reported as markers of outcome. As a consequence, attention has focussed on an improved understanding of ependymoma biology to improve patient survival.

Initially, this study has shown clear genomic and epigenetic disparity between paediatric ependymomas, particularly with respect to tumour location, patient age, tumour recurrence and patient outcome. Underpinning this biological diversity is the purported concept of the ependymoma-initiating cell.

The SNP array analysis found that spinal ependymomas were characterised by numerous and often broad genomic imbalances including gains of chromosome 17p, 12q, 21q, 16, 9, 18 and 20. In contrast supratentorial tumours demonstrated few large anomalies, while posterior fossa tumours were associated with gain of chromosome 1q. Likewise, the differentiation of spinal and intracranial ependymomas was evident from the methylation array analysis, which also observed distinct profiles both between and within posterior fossa and supratentorial groups. The SNP array data also revealed

ependymomas from younger children had fewer broad genomic aberrations than older children, while anomalies found at tumour recurrence included gains of chromosome 1q, 6, 8 and 9 and loss of chromosome 22. Moreover, clustering of imbalances enabled paediatric ependymomas to be categorised into three genomic groups. One such group, characterised by gain of either the entire arm of chromosome 1q or cytobands confined to this region, was associated with adverse patient survival thereby suggesting that this aberration warrants further consideration as a prognostic marker in paediatric ependymoma.

Many of the global aberrations detected in this work were subsequently analysed at an increased genomic resolution, the results of which are presented in subsequent chapters.

CHAPTER 4

PUTATIVE CANDIDATE GENES IN THE PATHOGENESIS
OF PAEDIATRIC EPENDYMOMA

4.1 Introduction

Previous expression analyses have implicated a variety of genes in ependymoma pathogenesis, particularly when considering tumour location, grade or genomic imbalance. Intracranial ependymomas have been shown to upregulate RAF1 (3p25) and members of the Notch, EPHB-EPHRIN, Hedgehog and bone morphogenetic protein pathways. In contrast, spinal ependymomas overexpress genes responsible for peptide production and activity including PLA2G5 (1p36.1) and ITIH2 (10p14), together with homeobox (HOX) family members involved in normal anteroposterior tissue development (Korshunov, Neben et al. 2003; Taylor, Poppleton et al. 2005; Modena, Lualdi et al. 2006; Lukashova-v Zangen, Kneitz et al. 2007; Palm, Figarella-Branger et al. 2009). The overexpression of EVI1 is reported to be exclusive to infratentorial ependymomas (Koos, Bender et al. 2011), while the disparate expression of CDKN2A (9p21.3) between spinal and supratentorial ependymomas and the association of spinal tumours with mutations of NF2 (22q12.2) and other Protein 4.1 family members have already been highlighted in Chapter 3. Differential gene expression profiles between classic and anaplastic intracranial ependymomas have been identified (Korshunov, Neben et al. 2003; Palm, Figarella-Branger et al. 2009), while the dysregulated expression of certain genes within regions of frequent genomic alteration such as chromosome 1q and 22q has also been demonstrated, including the overexpression of DUSP12 (1q23.3), PRELP (1q32) and HSPA6 (1q23) and the underexpression of G22P1 (22q13.2) and MCM5 (22q13.1) (Almeida, Zhu et al. 1998; Korshunov, Neben et al. 2003; Suarez-Merino, Hubank et al. 2005; Mendrzyk, Korshunov et al. 2006; Modena, Lualdi et al. 2006). However, almost all of the cohorts examined in these analyses were either wide-ranging in age or comprised entirely of adults, questioning the relevance of such findings for ependymomas in children.

Where feasible, some studies have detected age-related differences in the biological profile of particular genes in ependymomas from mixed-age cohorts. These include the overexpression of cell proliferation regulator genes LDHB (12p12.1) and STAM (10p12.3), upregulation of TNC (9q33), NTRK2 (9q21.33), ASS (9q34.1), KIAA0368 (9q31.3), hypomethylation of CDKN2A and CDKN2B (9p21.3) and hypermethylation of HIC1 (17p13.3) in younger patients (Korshunov, Neben et al. 2003; Rousseau, Ruchoux

et al. 2003; Waha, Koch et al. 2004; Modena, Lualdi et al. 2006). Although uncommon, exclusive expression analyses of paediatric ependymomas have been performed, finding supratentorial tumours differ from posterior fossa counterparts in the overexpression of neuronal markers NEFL (8p21.2), LHX2 (9q33.3), FOXG1 (14q12), TLX1 (10q24.3), and NPTXR (22q13.1) yet underexpression of RELN (7q22.1) and TNC (Andreiuolo, Puget et al. 2010), whilst also revealing aberrant expression of further genes on chromosomes 1q and 22q such as laminin (1q31), GAC1 (1q32) and CBX7 (22q13.1) (Suarez-Merino, Hubank et al. 2005). However, whether these findings are representative of the wider paediatric ependymoma population remains uncertain as the examined cohort sizes were small.

Until recently, the limited resolution of CGH to examine the ependymoma genome for copy number alterations meant that candidate genes within identified regions of gain or loss could only be postulated. Two subsequent array CGH studies refined this approach by assimilating genomic data with gene expression profiling for a subset of their cohorts, identifying genes with potentially copy number driven dysregulated expression. These included the putative tumour suppressor genes PTGDS (9q34), SULT4A1 (22q13.3) and oncogenes RXRA (9q34), TNC (9q33), ANAPC2 (9q34), PTGES (9q34), NTRK2 (9q21), COL27A1 (9q32) and NOTCH pathway members JAG1 (20p12) and NOTCH1 (9q34) (Modena, Lualdi et al. 2006; Puget, Grill et al. 2009). However, both studies analysed small cohorts and could not exclude other genes encompassed within these regions of imbalance from involvement in ependymoma pathogenesis.

The resolution of SNP array technology addresses the latter issue, allowing a more precise identification of copy number alteration boundaries. Using this platform, Johnson and colleagues integrated copy number data for 204 ependymomas with expression array profiles for a subset of 83 tumours (Johnson, Wright et al. 2010). This identified 107 putative ependymoma oncogenes including THAP11 (16q22.1), PSPH (7p11.2), EPHB2 (1p36.12), RAB3A (19p13.11), PCDH family members, PTPRN2 (7q36.3) and again, NOTCH1. Likewise, 130 potential tumour suppressor genes were detected including PTEN (10q23.31), CDKN2A (9p21.3), STAG1 (3q22.3) and TNRC6B (22q13.1). Moreover, overexpression of EPHB2 in CDKN2A/Ink4a/Arf $^{-/-}$ mice generated highly penetrant tumours that modelled the histology and expression profile

of certain human supratentorial ependymomas, supporting the suggestion that these two genes are implicated in the pathogenesis of a particular ependymoma subtype. However, whilst the largest reported study of its kind, Johnson's analysis did not exclusively assess a paediatric population. The cohort also comprised primary and recurrent ependymomas, often from the same patient, leading to the potential over-representation of certain anomalies, while the blood reference DNA used for the SNP analysis was obtained from a pool of adult leukaemia patients in remission post treatment.

The objective of this 500K SNP array analysis was to refine the copy number aberrations of ependymomas occurring in a paediatric age range, by normalising tumour DNA predominantly against patient-matched constitutional DNA. After an initial global analysis, the subsequent aim was to look with higher resolution for frequent and novel alterations both across the cohort and within clinical subgroups, then associate particular focal regions of imbalance in intracranial ependymoma with patient survival. The ability to generate copy number and LOH data for each SNP probe also enabled a unique opportunity to look for regions of acquired uniparental disomy (aUPD), a mechanism resulting in mutational homozygosity despite copy number neutrality as described in Chapter 1, section 1.3.2. While aUPD has been identified in other paediatric cancers (Fitzgibbon, Smith et al. 2005; George, Attiyeh et al. 2007; Raghavan, Smith et al. 2008), this was the first analysis attempting to establish its existence in childhood ependymoma.

As with the SNP array work, methylation array analysis performed as part of this study has already demonstrated distinct epigenetic profiles between different paediatric ependymoma groups, such as those according to tumour location within the CNS. Subsequent high resolution analysis of this epigenetic work aimed to identify differentially methylated genes responsible for such differences. Of interest were putative tumour suppressor genes demonstrating hypermethylation and thereby potential epigenetic inactivation or, conversely, candidate oncogenes conceivably upregulated through hypomethylation.

The following hypotheses were explored:

- High resolution genomic analysis will identify genes and/or pathways potentially involved in paediatric intracranial ependymoma pathogenesis.

- Ependymomas occurring in different anatomical locations within the central nervous system (CNS) harbour genes demonstrating distinct genomic and/or epigenetic aberrations.

- Intracranial ependymomas from younger children (under three years of age) harbour genes exhibiting different genomic and epigenetic alterations to intracranial ependymomas from older children (over three years of age).

- The genomic copy number alteration or methylation status of particular genes can define other groups with a clinical and/or prognostic relevance in paediatric intracranial ependymoma.

- Acquired uniparental disomy is a tumourigenic mechanism in paediatric ependymoma.

4.2 Materials and methods

4.2.1 500K SNP array analysis of 45 paediatric ependymomas

4.2.1.1 The sample cohort

Tumour and blood DNA extraction is outlined in Chapter 2, sections 2.1.3 − 2.1.4. An overview of the 500K SNP array protocol and data processing procedures followed are described in Chapter 2, sections 2.2.1 − 2.2.5. Clinical data for the entire SNP array cohort is summarised in Chapter 3, Figure 3.1 and the comprehensive data set is detailed in Chapter 3, Table 3.1. Of the cohort, 39 tumours (62 %) were incorporated into the SNP array study by Johnson and colleagues (Johnson, Wright et al. 2010). For the high resolution genomic imbalance analysis presented in this chapter, 36 primary and nine first recurrent intracranial ependymomas were assessed. Subsequent recurrences (2nd − 5th) were examined but are presented and discussed in Chapter 5. High resolution analysis of spinal ependymomas was not performed. It was deemed uninformative since the aberrations seen in these tumours predominantly involved whole chromosomal changes. As in Chapter 3, probes on chromosome X were excluded from the analysis (Chapter 3, section 3.2.1.1), while the analysis of genomic imbalance

between patients aged above and below three years was again performed exclusively on the primary posterior fossa cohort of 24 ependymomas.

Acquired UPD (aUPD) analysis requires the comparison of blood and tumour DNA from the same patient. Since five erroneous blood DNA samples had been removed from the study (Chapter 3, section 3.2.1.1), ten ependymomas corresponding to these five patients were also excluded from this particular analysis (samples 14P, 17P – R2, 19R1 – R2, 35P – R2 and 39P). Primary spinal tumours were included in this aspect of the analysis which resulted in an aUPD cohort of 38 primary and six intracranial first recurrent ependymomas.

4.2.1.2 Genomic imbalance data analysis – gene list formation

Annotated copy number data for each SNP probe within a tumour sample was available. Using the Excel 2007 program (Microsoft, USA), this information was collated and ordered according to the frequency with which a particular SNP probe demonstrated copy number gain or loss within a given collection of tumours. Genomic loss was defined by a copy number value below two, which was further divided into hemizygous (copy number one) and homozygous (copy number zero) deletion. Lower level copy number gain was defined by a copy number of three or four, while amplification was achieved with a copy number of five or six. At least five consecutive SNPs spanning regions greater than 10kb had to reveal the same imbalance before the corresponding region, established from the annotation data, could be incorporated into a final 'gene list' for the particular tumour group assessed. These parameters were in keeping with other SNP array analyses (Baker, Preisinger et al. 1990; Northcott, Nakahara et al. 2009; Johnson, Wright et al. 2010) and thereby provided a means of identifying genes with the most common copy number aberrations across a chosen number of ependymomas. Gene lists generated for particular clinical subgroups were also compared against each other to establish group specific genomic aberrations.

The annotation file (Netaffx file build 07.12.07) used to generate the gene lists only accounted for genes assigned to SNPs on the array and thereby not all genes within a given focal region. However, as the lists were prepared manually, only SNP-assigned

genes were created in order to avoid the presentation of exhaustive lists and satisfy the inherent time constraints on the project. This was adhered to except for focal regions encompassing the genes of interest selected for subsequent qPCR validation. All genes within these regions were identified using the web-based genomic database Ensembl (http://www.ensembl.org) and included in the gene lists presented. Established gene functions were derived from the Genecard database (www.genecards.org) unless referenced.

4.2.1.3 aUPD data analysis and visualisation

Genotyping data was also available for each probe on the 500K array, as a result of processing through GTYPE (Chapter 2, section 2.2.3). By comparing probe genotype calls between patient-matched tumour and blood DNA through CNAG, regions of LOH throughout the genome were identified. Combining LOH results with the copy number data for each SNP probe in Microsoft Excel 2007 enabled sites of copy number neutral LOH (aUPD) to be identified across the analysed cohort (Appendix 10H). A true region of focal aUPD was defined when five or more consecutive SNP probes covering a distance greater than 10kb had demonstrated copy number neutral LOH. The aUPD data from Excel 2007 was exported to Spotfire Decision Site$^{\circledR}$ for data visualisation, whilst aUPD regions within genes were used to formulate subsequent gene lists. An overview of the method used to calculate aUPD for each tumour sample is shown in Appendix 5.

4.2.1.4 Real time qPCR validation of SNP array results

The SNP derived copy number for selected genes of interest was validated using real time qPCR as detailed in Chapter 2, section 2.4. Where possible, tumour and blood DNA processed on the SNP microarray was also used for qPCR analysis. However in 42/63 ependymomas from the SNP array cohort (67 %; 26 primary and 16 recurrent tumours), the DNA stock had been exhausted, either from the microarray analysis itself or subsidiary collaborative projects. Consequently, the DNA from 31 ependymomas from the SNP array cohort (58 %; 24 primary and seven recurrent tumours) was used for qPCR validation. Within this sample set, 18 tumours (12 primary and six recurrent) required DNA re-extraction from a different cut of tumour (performed as detailed in

Chapter 2, section 2.1.2 and 2.1.3). Furthermore, in certain qPCR experiments, control DNA (Promega, UK) was used as an alternative to three of the erroneous samples of blood DNA (17P – R2 bl, 19R – R2 bl and 35P – R2 bl) discussed previously.

For validated target genes, up to eight tumours demonstrating imbalance of that gene from the microarray analysis were evaluated with between three and eight tumours exhibiting a normal SNP copy number for that gene. The precise number of samples used varied depending on the target gene being analysed (section 4.3.5). The gene copy numbers derived from the SNP array and real time qPCR methods were correlated using Spearman's correlation (Chapter 2, section 2.7.3). Tables documenting all qPCR results for the genes of interest are included in Appendix 10I.

4.2.2 Methylation array analysis of 98 paediatric ependymomas

An overview of the methylation array protocol adopted and subsequent data processing methods are described in Chapter 2, sections 2.3.1 – 2.3.4. Clinical data for the entire methylation array cohort is summarised in Chapter 3, Figure 3.2 and the comprehensive data set is detailed in Chapter 3, Table 3.2. The methylation data for all tumours of the array cohort is included in Appendix 10F.

4.2.2.1 Differential gene methylation analysis

This was performed as described in Chapter 2, section 2.7.7. Probes on the X chromosome were removed from the analysis as detailed in Chapter 3, section 3.2.2.2. The clinical subgroups of patient sex (male versus female), age (below three years of age versus above three years of age), tumour resection status (complete versus incomplete), tumour grade (WHO grade II versus WHO grade III), tumour location (spinal versus intracranial, supratentorial versus alternative locations, posterior fossa versus alternative locations) and patient survival status (alive versus dead) were assessed on the primary cohort of 73 tumours. In addition, the intracranial primary cohort of 56 ependymomas was initially compared against 11 first recurrent intracranial tumours and then the entire intracranial recurrent cohort.

Genes differentially methylated between ependymomas from different CNS locations were cross-checked for matches against a list of signature genes of ependymomas from the three principal CNS sites as identified by the expression array work of Taylor and colleagues (Taylor, Poppleton et al. 2005). This was performed to establish whether such characteristic expression profiles could be a manifestation of dysregulated gene methylation mechanisms.

4.2.3 Other statistical analysis

Gene lists from particular clinical subgroups were compared statistically to identify group-specific focal imbalances. An overview of the formulaic method used to perform this is shown in Appendix 6. Imbalances associated with a particular group were ranked in order of statistical significance as deemed by two-tailed Fisher's exact testing (SPSS 16). Only associations with a p-value below 0.05 were considered. The clinical subgroups were derived from the variables of tumour location (supratentorial versus posterior fossa), patient age (posterior fossa ependymomas: patients under three years versus over three years) and tumour grade (WHO grade II versus grade III).

Other statistical tests performed in this chapter are described in Chapter 2, section 2.7. The variables included in multivariate survival analysis are as described in Chapter 3, section 3.2.3.

4.3 Results

4.3.1 Gene copy number imbalance in paediatric intracranial ependymoma

4.3.1.1 Common focal regions of genomic gain in primary intracranial ependymomas

In keeping with the lower genomic resolution analysis (Chapter 3, section 3.3), copy number gain was a more frequent event than loss. Focal regions of genomic gain in the primary intracranial cohort of 36 ependymomas were identified and ranked according to frequency (Figure 4.1 and Table 4.1). The most common regions were predominantly

located on chromosome 1q and chromosome 18, specifically within 1q44 (18/36 samples; 50 %), 1q32.3 (up to 17/36; 47 %), 18p11.21 (16/36; 44 %) and 18q23 (16/36; 44 %). Genes found to be encompassed within the regions on chromosome 1q included the transcriptional regulators ZNF672 and ZNF692 (1q44) and NSL1 (1q32.3), a gene encoding part of a protein complex required for kinetochore formation Within the 18p11.21 region was the SEH1L gene which encodes another protein complex component responsible for promoting kinetochore function. Other common sites of genomic gain included foci within 4q35.2, 11q24.3 and 9q31.3. These encompassed genes such as DNAJC25 (9q31.3) which encodes a member of the DNAJ/HSP40 heat shock protein family.
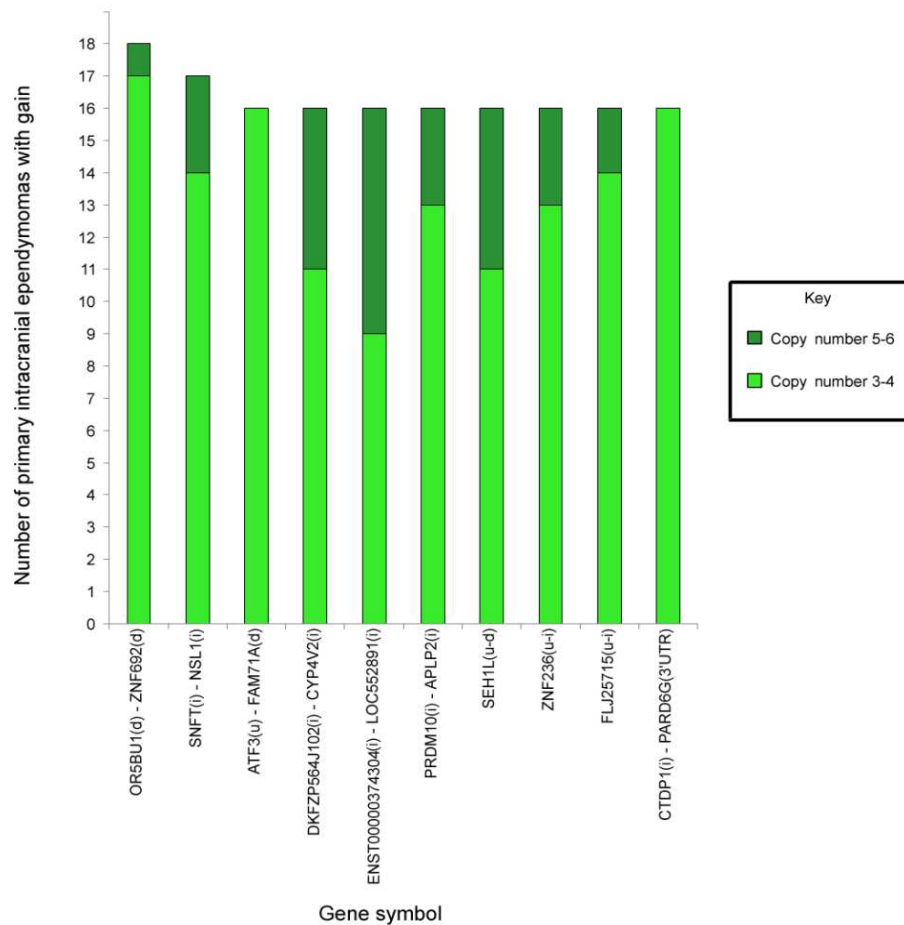
Figure 4.1: The most frequent focal regions of increased copy number in 36 primary intracranial ependymomas. For each gene region, samples with low level copy number gain (copy number 3 or 4) are differentiated from those with amplification (copy number 5 or 6) by light and dark green shading respectively. bp = base pair, i = intronic, u = upstream of annotated gene, d = downstream of annotated gene, 3'UTR = 3' untranslated region, CDS = coding sequence.

Table 4.1: The most frequent focal regions of increased copy number in 36 primary intracranial ependymomas.

Locus	Start (bp)	End (bp)	Gene symbol	Additional genes in this region	Total	CN 3 or 4	CN 5 or 6
1q44	247069146	247135059	OR5BU1 (d) – ZNF692 (d)	SH3BP5L, ZNF672, ZNF692	18	17	1
1q32.3	210937704	211013542	SNFT (i) – NSL1 (i)		17	14	3
1q32.3	210801617	210925420	ATF3 (u) – FAM71A (d)	FAM71A	16	16	0
4q35.2	187326127	187353221	DKFZP564J102 (i) – CYP4V2 (i)		16	11	5
9q31.3	113426420	113463254	ENST00000374304 (i) – LOC552891 (i)	DNAJC25	16	9	7
11q24.3	129371863	129473754	PRDM10 (i) – APLP2 (i)		16	13	3
18p11.21	12918541	12979722	SEH1L (u – d)		16	11	5
18q23	72613722	72670749	ZNF236 (u – i)		16	13	3
18q23	75474284	75507191	FLJ25715 (u-i)		16	14	2
18q23	75602382	76017006	CTDP1 (i) – PARD6G (3'UTR)	PQLC1,TXNL4A, C18orf22, ADNP2	16	16	0

Note: The gene symbols representing the start and end of each specified genomic region are shown, together with additional encompassed genes identified from the Affymetrix[®] annotation file (Netaffx file build 07.12.07). The total number of samples within the cohort demonstrating genomic gain of the region is also shown. This total is then split according to numbers exhibiting low level gain (CN 3 or 4) or amplification (CN 5 or 6). CN = copy number, bp = base pair, i = intronic, u = upstream of annotated gene, d = downstream of annotated gene, 3'UTR = 3' untranslated region, CDS = coding sequence. Genes highlighted in red had their copy number validated by real time qPCR. Focal regions encompassing these highlighted genes were additionally examined using the web-based genomic database Ensembl to identify all other interrogated genes which are duly included.

4.3.1.2 Common focal regions of genomic loss in primary intracranial ependymomas

Focal regions of copy number loss in the primary intracranial cohort of 36 tumours were identified and ranked according to frequency (Figure 4.2 and Table 4.2). Several of the most common regions were located across chromosome 6q, specifically within 6q13 – 14.1 (4/36 samples; 11 %), 6q16.1 (up to 5/36; 14 %), 6q22.1 – 22.31 (4/36; 11 %) and 6q27 (4/36; 11 %). Other foci of deletion were found at 11q12.2, 16q24.1, 22q11.21, 22q12.3 and 22q13.31 (4/36 samples; 11 %). Putative tumour suppressor genes reported from other studies were encompassed by certain focal regions detected on chromosome 6q. These included FILIP1 (6q14.1), a gene encoding a filament interacting protein (Shimada, Shiratori et al. 2009) and the tyrosine kinase encoding FRK/RAK (6q22.1) (Yim, Peng et al. 2009).
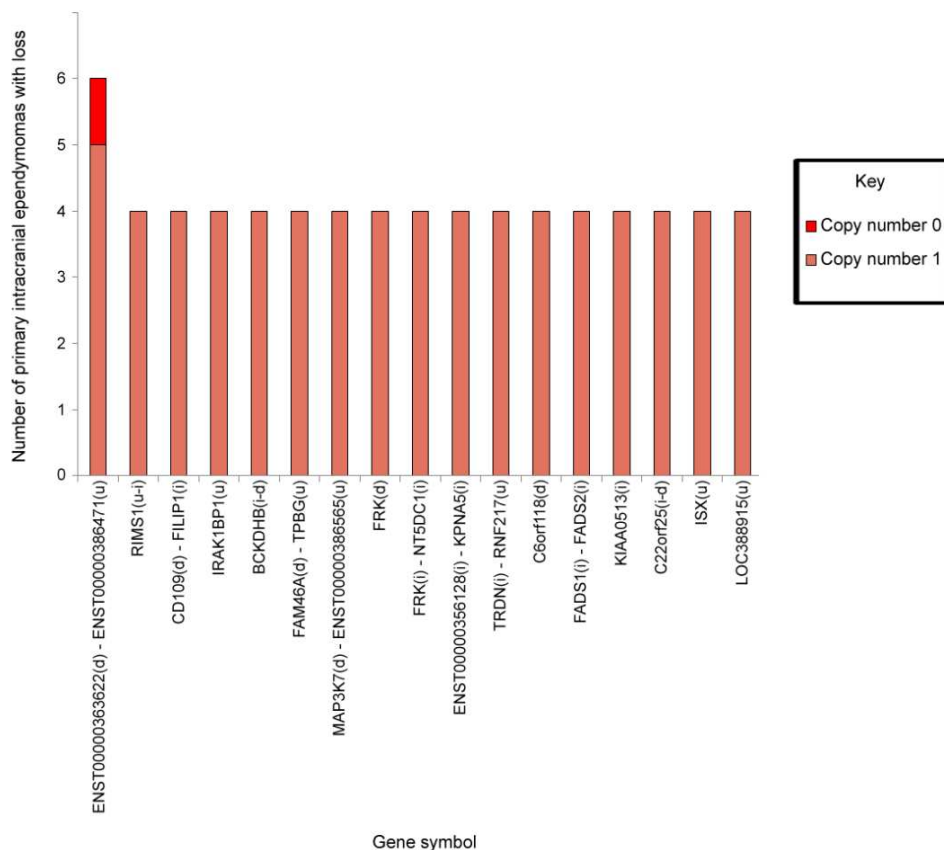
Figure 4.2: The most frequent focal regions of decreased copy number in 36 primary intracranial ependymomas. For one gene region, samples with hemizygous deletion (copy number 1) are differentiated from the sample with homozygous deletion (copy number 0) by light and dark red shading respectively. bp = base pair, i = intronic, u = upstream of annotated gene, d = downstream of annotated gene.

Table 4.2: The most frequent focal regions of decreased copy number in 36 primary intracranial ependymomas.

Locus	Start(bp)	End(bp)	Gene symbol	Additional Genes in this region	Total	CN 1	CN 0
6q16.1	92913726	92951742	ENST00000363622 (d) – ENST00000386471 (u)		6	5	1
6q13	72646417	72967122	RIMS1 (u-i)		4	4	0
6q13-14.1	75809165	76249718	CD109 (d) – FILIP1 (i)	TMEM30A, COL12A1, COX7A2	4	4	0
6q14.1	79252979	79544001	IRAK1BP1 (u)		4	4	0
6q14.1	80951883	81916795	BCKDHB (i-d)		4	4	0
6q14.1	82398726	83098352	FAM46A (d) – TPBG (u)	FAM46A, IBTK	4	4	0
6q16.1	92449499	92760201	MAP3K7 (d) – ENST00000386565 (u)		4	4	0
6q22.1	115407082	115764663	FRK (d)		4	4	0
6q22.1	116381423	116642944	FRK (i) – NT5DC1 (i)	COL10A1	4	4	0
6q22.1	116926752	117126249	ENST00000356128 (i) – KPNA5 (i)	FAM26D, FAM26E, RWDD1, RSHL3	4	4	0
6q22.31	123841038	125314016	TRDN (i) – RNF217 (u)	ENST00000334268, TCBA1	4	4	0
6q27	164971699	165411821	C6orf118 (d)		4	4	0
11q12.2	61327359	61365899	FADS1 (i) – FADS2 (i)		4	4	0
16q24.1	83628627	83640211	KIAA0513 (i)		4	4	0
22q11.21	18411964	18433554	C22orf25 (i-d)		4	4	0
22q12.3	32986814	33103088	ISX (u)		4	4	0
22q13.31	46653061	46744848	LOC388915 (u)		4	4	0

Note: The gene symbols representing the start and end of each specified genomic region are shown, together with additional encompassed genes identified from the Affymetrix® annotation file (Netaffx file build 07.12.07). The total number of samples within the cohort demonstrating genomic loss of the region is also shown. This total is then split according to numbers exhibiting hemizygous (CN 1) or homozygous genomic deletion (CN 0). CN = copy number, bp = base pair, i = intronic, u = upstream of annotated gene, d = downstream of annotated gene. Genes highlighted in red had their copy number validated by real time qPCR. Focal regions encompassing these highlighted genes were additionally examined using the web-based genomic database Ensembl to identify all other interrogated genes which are duly included.

4.3.1.3 Common focal regions of genomic gain in recurrent intracranial ependymomas

Focal regions of copy number gain in the first recurrent intracranial cohort of nine ependymomas were identified and ranked according to frequency (Table 4.3 and Figure 4.3). Genomic gain was again observed more often than loss and was seen particularly on chromosome 1q. The most common regions were located within 1q23.3, 1q25.3 and 1q32.1 (6/9 samples; 67 %). This latter site encompassed the gene NAV1 which encodes a member of the neuron navigator protein family.

Other common regions of gain spanning chromosome 1q were identified in over half of the cohort (5/9 samples 55 %). Numerous genes were encompassed by these locations including CHIT1 and CHI3L1/YKL40 (1q32.1), both members of the chitinase family and NES (1q23.1) which encodes the neuronal stem cell marker Nestin. Genes involved in a variety of cellular processes required for tumourigenesis and recurrence were also evident. For instance, genes reported to be involved in cell migration and invasion, neurite growth and the inhibition of apoptosis were identified such as BCAN/BEHAB (1q23.1), ASTN1, TNN (1q25.1), ABL2 (1q25.2), LAMC1 (1q25.3), MDM4 (1q32.1), PIK3C2B, NFASC and CDC42BPA (1q42.13) (Jaworski, Kelly et al. 1996; Volkmer, Leuschner et al. 1996; Fink, Hirsch et al. 1997; Migliorini, Lazzerini Denchi et al. 2002; Neidhardt, Fehr et al. 2003; Wiksten, Liebkind et al. 2003; Wilkinson, Paterson et al. 2005; Katso, Pardo et al. 2006; Srinivasan and Plattner 2006). Genes encoding kinetochore associated proteins were also detected, including PMF1 (1q23.1), NUF2 (1q23.3), KIF14 (1q32.1) and CENPF (1q41).

Table 4.3: The most frequent focal regions of increased copy number in nine first recurrent intracranial ependymomas.

Locus	Start (bp)	End (bp)	Gene symbol	Additional genes in this region	Total	CN 3 or 4	CN 5 or 6
1q23.3	163712154	163733164	RXRG (u)		6	5	1
1q25.3	180381192	180429300	ZNF648 (u)		6	5	1
1q32.1	199876870	199934485	NAV1 (i)		6	6	0
1q21.1	142756696	144232761	ENST00000360154 (i) – PEX11B (i)	PDE4DIP, ENST00000369338, HFE2, LIX1L	5	5	0
1q21.3	151712798	152059582	S100A7 (d) – GATAD2B (i)	S100A1-6/A13,14,16, C1orf77, SNAPAP, NPR1, INTS3	5	5	0
1q21.3	152834710	153070269	ADAR (i) – KCNN3 (i)		5	5	0
1q23.1	154245657	158107790	SSR2 (3'UTR) – SLAMF8 (d)	MAPBPIP, LMNA, SEMA4A, SLC25A44, PMF1, SMG5, CCT3, C1orf182, C1orf161, RHBG, MEF2D, IQGAP3, APOA1BP, GPATCH4, BCAN, NES, ISG20L2, MRPL24, PRCC, INSRR, SH2D2A, NTRK1, ARHGEF11, FLJ16478, ETV3, FCRL1-5, CD5L, KIRREL, SLAMF8, ENST00000368172, CD1C-E, OR10T2, CADM3, ENST00000368153, OR10R2, OR6Y1, DARC, SPTA1, ENST00000368153/361688/334632, OR6N1, OR6N2, MNDA, ENST00000359709, IFI16, FCER1A, FCRL6, PYHIN1, APCS, ENST00000328408/289731/356104/368102	5	5	0
1q23.3	161675875	161768950	NUF2 (u) – ENST00000385703 (d)	NUF2	5	5	0
1q23.3	163423053	163710084	LMX1A (u) – RXRG (d)	LMX1A, RXRG	5	5	0
1q23.3-24.1	163776914	165356012	LRRC52 (u) - ENST00000361200 (i)	LRRC52, MGST3, ENST00000367888, ALDH9A1, TMCO1, UCK2, FAM78B, ENST00000367877,ENST00000336247, POGK, TADA1L, C1orf32, MAEL, GPA33	5	5	0
1q24.2	167194079	167400429	NME7 (u-i)	ATP1B1	5	5	0

154

Locus	Start (bp)	End (bp)	Gene symbol	Additional genes in this region	Total	CN 3 or 4	CN 5 or 6
1q25.1	173357037	176182097	TNN (i) – SEC16B (i)	KIAA0040, TNR, ENST00000367674, RFWD2, PAPPA2, ASTN1, FAM5B	5	5	0
1q25.2	176881837	179132968	RALGPS2 (u) – STX6 (u)	RALGPS2, ANGPTL1, FAM20B, TOR3A, ABL2, SOAT1, C1orf125, TDRD5, C1orf76, TOR1AIP1, XPR1, TOR1AIP2, CEP350, QSOX1, LHX4, ACBD6,	5	5	0
1q25.3	179453720	179830712	IER5 (d) – CACNA1E (i)	ENST00000367573	5	5	0
1q25.3	180249013	180378729	ZNF648 (u-d)		5	5	0
1q25.3	180571770	181488636	ENST00000361845 (CDS) – NMNAT2 (i)	TEDDM1, C1orf120, RGSL1, RGSL2, RNASEL, RGS8, NPL, DHX9, ENST00000367557, ENST00000384741, LAMC1, LAMC2	5	5	0
1q25.3	182920164	183870395	C1orf21(d) – HMCN1(u)	EDEM3, FAM129A, RNF2, C1orf26, IVNS1ABP	5	5	0
1q32.1	198509033	199328441	ENST00000367356 (u) – CACNA1S (i)	ENST00000367356, KIF14, DDX59, CAMSAP1L1, C1orf106, KIF21B	5	5	0
1q32.1	199741167	199873895	CSRP1(i) – ENST00000367302 (i)	ENST00000362085	5	5	0
1q32.1	200080574	200298931	IPO9 (i) – ELF3 (d)	LMOD1, TIMM17A, RNPEP, ELF3	5	5	0
1q32.1	201414285	201590259	CHI3L1 (u) – FMOD (d)	CHI3L1, CHIT1, BTG2, FMOD	5	5	0
1q32.1	201910943	201930848	ATP2B4 (i)		5	5	0
1q32.1	202096179	202785465	SNRPE (u) – MDM4 (3'UTR)	SNRPE, SOX13, ETNK2, REN, KISS1, GOLT1A, PLEKHA6, PIK3C2B,	5	5	0
1q32.1	202922030	203238108	LRRN2 (d) – NFASC (i)	ENST00000367173	5	5	0
1q32.1	206609806	206620808	PLXNA2 (u)		5	5	0
1q32.2	207669983	208414613	ENST00000367032 (5'UTR) – SERTAD4 (u)	CAMK1G, LAMB3, HSD11B1, ENST00000367026, TRAF3IP3, C1orf74, ENST00000367021, IRF6, C1orf107, SYT14	5	5	0
1q32.3	209708573	210010455	RD3 (d) – LPGAT1 (i)	ENST00000367000, SLC30A1, NEK2	5	5	0
1q41	211087402	212921197	LOC149643(i) – KCNK2 (u)	FLVCR1, VASH2, ANGEL2, RPS6KC1, PROX1, SMYD2, PTPN14, CENPF	5	5	0

Locus	Start (bp)	End (bp)	Gene symbol	Additional genes in this region	Total	CN 3 or 4	CN 5 or 6
1q42.13	224703718	225258279	ENST00000363938 (d) – CDC42BPA (i)	C1orf95, ITPKB, PSEN2, ENST00000313259, ENST00000366779, CABC1	5	5	0
1q42.13	228770808	229925804	COG2 (u) – DISC1 (i)	COG2, AGT, CAPN9, C1orf198, TTC13, ARV1, FAM89A, ENST00000366656, TRIM67, C1orf131, GNPAT, EGLN1, TSNAX	5	5	0
1q42.2	231650194	231813713	ENST00000366656 (i)		5	5	0
1q42.3	233488558	234652165	ARID4B (i) – EDARADD (i)	TBCE, B3GALNT2, GNG4, LYST, NID1, GPR137B, ERO1LB	5	5	0
1q43	240824595	241218126	PLD5 (u)		5	5	0
1q44	243485054	243810042	ENST00000366519 (d) – ENST00000366518 (i)	ENST00000329504	5	5	0
8p21.3	21344076	21350938	ENST00000387608 (u)		5	4	1
8q24.23	139139534	139368001	FAM135B (u-i)		5	5	0

Note: The gene symbols representing the start and end of each specified genomic region are shown, together with additional encompassed genes identified from the Affymetrix® annotation file (Netaffx file build 07.12.07). The total number of samples within the cohort demonstrating genomic gain of the region is also shown. This total is then split according to numbers exhibiting low level gain (CN 3 or 4) or amplification (CN 5 or 6). CN = copy number, bp = base pair, i = intronic, u = upstream of annotated gene, d = downstream of annotated gene, 3'UTR = 3' untranslated region, CDS = coding sequence. The gene highlighted in red had its copy number validated by real time qPCR.
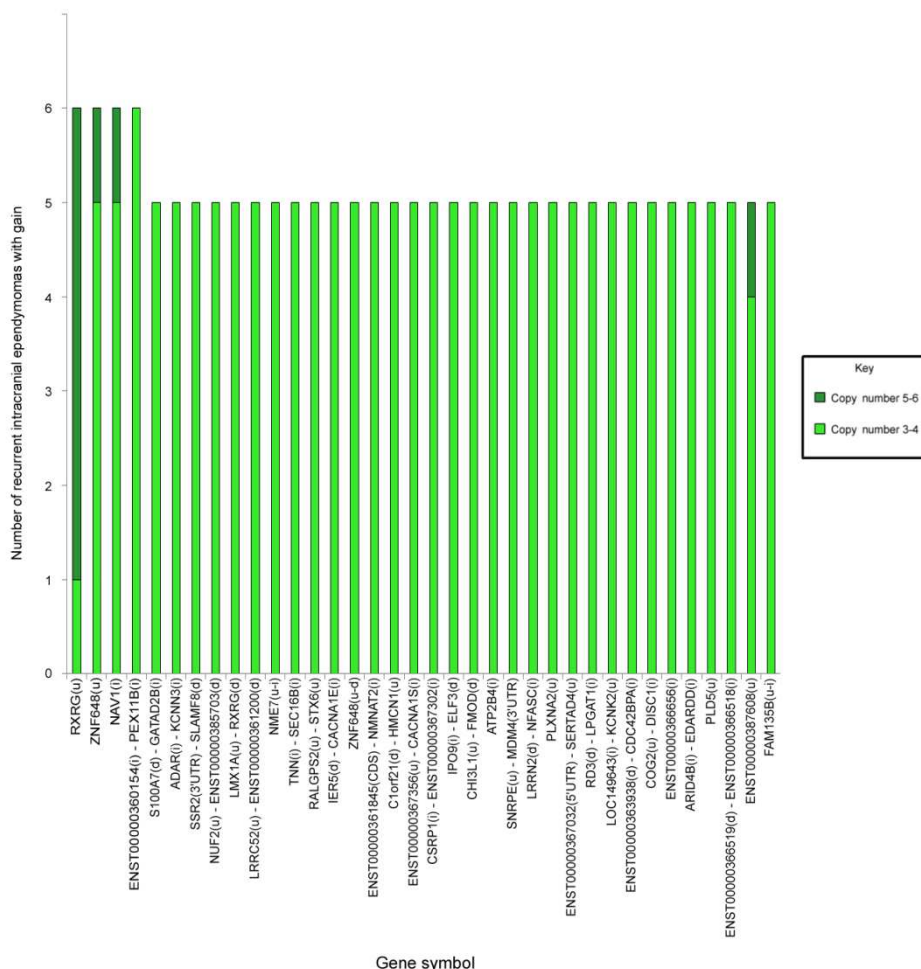
Figure 4.3: The most frequent focal regions of increased copy number in nine first recurrent intracranial ependymomas. For each gene region, samples with low level copy number gain (copy number 3 or 4) are differentiated from those with amplification (copy number 5 or 6) by light and dark green shading respectively. bp = base pair, i = intronic, u = upstream of annotated gene, d = downstream of annotated gene, 3'UTR = 3' untranslated region, CDS = coding sequence.

4.3.1.4 Common focal regions of genomic loss in recurrent intracranial ependymomas

Focal regions of copy number loss in the first recurrent intracranial cohort of nine ependymomas were identified and ranked according to frequency (Table 4.4 and Figure 4.4). The most common regions spanned the length of chromosome 22q, particularly within 22q12.3, 22q13.1 and 22q13.31 – 13.33 (3/9 samples 33 %). Putative tumour suppressor genes reported from other studies were encompassed within these regions, such as the chromobox protein encoding CBX7 (22q13.1) (Suarez-Merino, Hubank et al. 2005) and PPARA (22q13.31) which encodes a member of the peroxisome proliferator receptor family (Niho, Takahashi et al. 2003; Grabacka, Plonka et al. 2006; Urbanska, Pannizzo et al. 2008).

157

Table 4.4: The most frequent focal regions of decreased copy number in nine first recurrent intracranial ependymomas.

Locus	Start(bp)	End(bp)	Gene symbol	Additional genes in this region	Total	CN 1	CN 0
22q12.3	34247255	34362917	MCM5 (d) – LOC284912 (u)	RASD2, MB	3	3	0
22q13.1	37451984	38282044	GTPBP1 (CDS) – RPS19BP1 (u)	UNC84B, DNAL4, NPTXR, APOBEC3 family, CBX7, PDGFB, SYNGR1, MAP3K7IP1, MGAT3, SMCR7L	3	3	0
22q13.31	43444530	43664982	LOC553158 (u) – PHF21B (i)	PRR5, LOC553158, ARHGAP8,	3	3	0
22q13.31	44861612	45167944	FLJ27365 (i) – CELSR1(i)	PPARA, LOC150383, PKDREJ, FLJ20699, GTSE1	3	3	0
22q13.31	45441994	45481189	DIP (i) – CERK (i)		3	3	0
22q13.32	47241222	47347102	FAM19A5 (u) – ENST00000336769 (i)		3	3	0
22q13.32	47404665	47446381	FAM19A5 (i)		3	3	0
22q13.33	48960405	49576671	PANX2 (d) – MGC70863 (i)	TRABD, TUBGCP6, HDAC10, SAPS2, ADM2, NCAPH2, CPT1B, ARSA	3	3	0

Note: The gene symbols representing the start and end of each specified genomic region are shown, together with additional encompassed genes identified from the Affymetrix® annotation file (Netaffx file build 07.12.07). The total number of samples within the cohort demonstrating genomic loss of the region is also shown. This total is then split according to numbers exhibiting hemizygous (CN 1) or homozygous genomic deletion (CN 0). CN = copy number, bp = base pair, i = intronic, u = upstream of annotated gene, d = downstream of annotated gene.
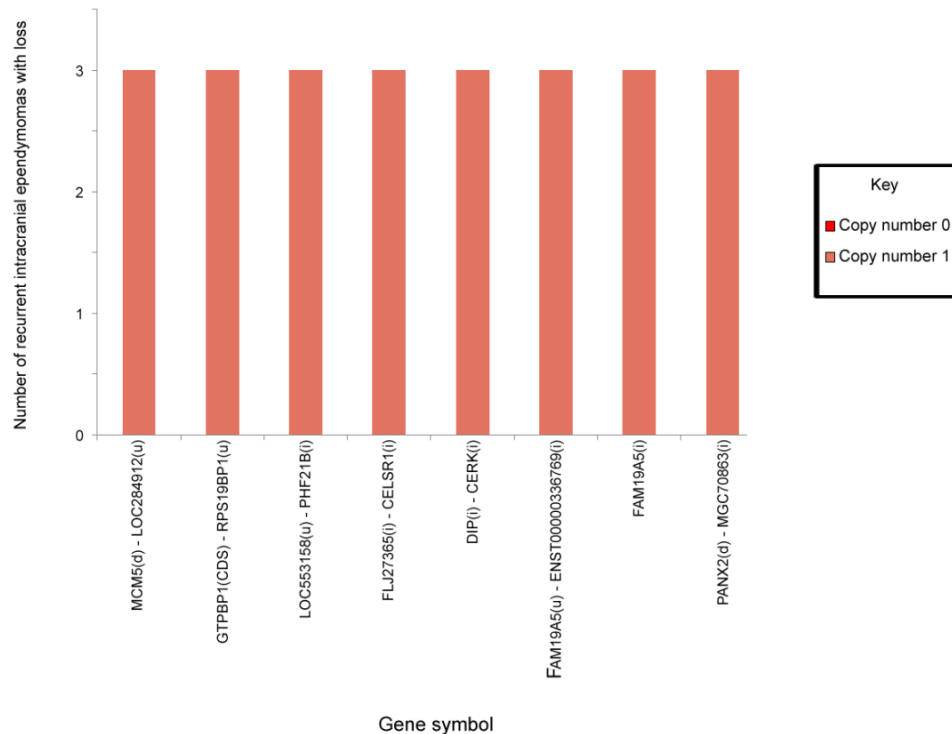
Figure 4.4: The most frequent focal regions of decreased copy number in nine first recurrent intracranial ependymomas. Samples with hemizygous deletion (copy number 1) are shown, coloured in light red. bp = base pair, i = intronic, u = upstream of annotated gene, d = downstream of annotated gene.

4.3.1.5 Common focal regions of amplification in intracranial ependymomas

Focal regions of amplification (copy number five or six) in the primary and first recurrent intracranial tumour cohorts were identified and ranked according to frequency (Tables 4.5A and B). The most common regions in the primary cohort were located within 1q41 and 7p15.2 (8/36 samples; 22 %). The latter region encompassed members of the HOX transcription factor family which, intriguingly, are reported to be relatively underexpressed in intracranial ependymomas when compared to corresponding extracranial tumours (Korshunov, Neben et al. 2003; Taylor, Poppleton et al. 2005; Palm, Figarella-Branger et al. 2009; Johnson, Wright et al. 2010). This finding is discussed later in this chapter. Less common regions of amplification were found on 9q31.3, 10q21.1, 13q31.1 and 18q22.3 (7/36, 19 %). The DNAJC25 gene within 9q31.3 had already been identified as one of the most frequently gained genes in the cohort (Table 4.1). Within the recurrent cohort, the most frequent regions of high level gain encompassed genes such as PLXNC1 (12q22) (2/9 samples; 22 %) which encodes a member of the plexin receptor family.

159

Table 4.5A: The most frequent focal regions of amplification in 36 primary intracranial ependymomas.

Locus	Start (bp)	End (bp)	Gene symbol	Additional genes in this region	Frequency
1q41	219025582	219031421	MOSC2 (d) – MOSC1 (i)		8/36
7p15.2	27147941	27208884	HOXA5 (3'UTR) – EVX1 (u)	HOXA6, HOXA7, HOXA9, HOXA10, HOXA1, HOXA13	8/36
9q31.3	113434096	113466466	DNAJC25 (i)		7/36
10q21.1	59937923	59951522	TFAM (d) – ENST00000373886 (i)		7/36
13q31.1	79810102	79825947	SPRY2 (i-d)		7/36
18q22.3	67022770	67037233	CBLN2 (d)		7/36

Table 4.5B: The most frequent focal regions of amplification in nine first recurrent intracranial ependymomas.

Locus	Start (bp)	End (bp)	Gene symbol	Additional genes in this region	Frequency
3p21.31	45331970	45350074	LARS2 (u)		3/9
1p22.2	90012111	90026999	LRRC8D (u)		2/9
2p24.1	20688172	20704649	HS1BP3 (i)		2/9
5p15.33	1789711	1802215	MRPL36 (d)		2/9
6p22.1	26065088	26125150	TRIM38 (u) – HIST1H1A (u)	TRIM38	2/9
6p21.1	42388855	42417178	TRERF1 (i)		2/9
6q25.3	159831048	159845349	FNDC1 (d)		2/9
8p23.1	11389950	11403101	BLK (i)		2/9
10p15.2	3045000	3046044	PFKP (u)		2/9
11p12	41059394	41075935	ENST00000386948 (d)		2/9
12q22	93148834	93173234	PLXNC1 (i)		2/9
15q23	67367475	67384434	PAQR5 (u)		2/9

Note: The gene symbols representing the start and end of each specified genomic region are shown, together with additional encompassed genes identified from the Affymetrix® annotation file (Netaffx file build 07.12.07). The total number of samples within the cohort demonstrating amplification of the region is also shown. CN = copy number, bp = base pair, i = intronic, u = upstream of annotated gene, d = downstream of annotated gene. The genes highlighted in red had their copy number validated by real time qPCR. Focal regions encompassing these highlighted genes were additionally examined using the web-based genomic database Ensembl to identify all other interrogated genes which are duly included.

4.3.1.6 Focal regions of homozygous loss in intracranial ependymomas

Only two focal regions of homozygous deletion (copy number zero) were identified in the cohort of 36 primary and nine first recurrent intracranial paediatric ependymomas. The regions were detected only once in separate tumour samples. One such site, detected in ependymoma 23P, was located within chromosome 6q16.1. It did not encompass any genes, mapping downstream of ENST00000363622 and upstream of ENST00000386471. This was verified from the Ensembl database. Hemizygous deletion of this region was also detected in five other primary intracranial tumours. The other region of homozygous deletion was located within chromosome 9p21.3 of recurrent sample 18R1, encompassing the recognised ependymoma tumour suppressor genes CDKN2A and CDKN2B (Taylor, Poppleton et al. 2005; Johnson, Wright et al. 2010).

4.3.1.7 Focal regions of genomic imbalance associated with posterior fossa ependymomas

Within the primary intracranial cohort, focal regions of copy number gain associated with posterior fossa ependymomas (compared to supratentorial tumours) were identified and ranked according to statistical significance (Table 4.6). The most significant regions were all located on chromosome 1q. Numerous genes were encompassed, including genes already noted as being among the most frequently gained in the intracranial recurrent cohort (Table 4.3). Such genes included the aforementioned CHIT1 (1q32.1; $p = 0.033$), CHI3L1/YKL40 (1q32.1; $p = 0.015$), CDC42BPA (1q42.13; $p = 0.033$) and ASTN1 (1q25.2; $p = 0.033$), a gene encoding a neuronal adhesion molecule which has been reported to mediate the migration of primitive neurons in the cerebellum (Edmondson, Liem et al. 1988). Other regions of 1q gain associated with posterior fossa tumours encompassed genes such as the kinase encoding AKT3 (1q44; $p = 0.033$) and PARP1 (1q42.12; $p = 0.033$) which encodes the chromatin-associated enzyme, poly (ADP-ribosyl) polymerase, involved in a range of cellular processes such as DNA damage repair and cell proliferation. No regions of focal genomic loss associated with posterior fossa ependymomas were identified.

4.3.1.8 Focal regions of genomic imbalance associated with supratentorial ependymomas

Focal regions of copy number gain associated with supratentorial ependymomas (compared to posterior fossa tumours) were also identified in the primary intracranial cohort and ranked according to statistical significance (Table 4.7). Again, several genes encompassed within significant regions were detected, including TULP4 (6q25.3; $p = 0.006$) which is implicated in protein ubiquitination and the Type IV collagen encoding COL4A1 (13q34; $p = 0.02$). The only significant region of genomic loss associated with supratentorial tumours was within 11q12.2 ($p = 0.025$), encompassing FADS1 encoding one of the fatty acid desaturase family members.

4.3.1.9 Focal regions of genomic imbalance associated with posterior fossa ependymomas from patients aged below three years

Within the primary posterior fossa cohort, focal regions of copy number gain associated with tumours from patients below three years of age (compared to those from children aged above three years) were identified and ranked according to statistical significance (Table 4.8). Genes detected within these regions included TELO2/CLK2 (16p13.3; $p = 0.014$) and HDAC10 (22q13.33; $p = 0.037$). TELO2 encodes a putative cell cycle S-phase checkpoint protein, while HDAC10 encodes a member of the histone deacetylase family. Genes encoding transmembrane proteins such as TMEM18 (2p25.3; $p = 0.037$) were also identified. No regions of focal genomic loss associated with the younger age group were identified.

4.3.1.10 Focal regions of genomic imbalance associated with posterior fossa ependymomas from patients aged above three years

Focal regions of copy number gain associated with tumours from patients above three years of age (compared to those from children aged below three years) were also identified in the posterior fossa cohort and ranked according to statistical significance (Table 4.9). The most significant regions were located on chromosomes 1q, 9 and 18. These regions encompassed genes such as the dynamin family member DNM3 (1q24.3;

p = 0.027) and the putative oncogene TXN (9q31.3; p = 0.014) which encodes an oxidoreductase enzyme overexpressed in several malignancies (Grogan, Fenoglio-Prieser et al. 2000; Kakolyris, Giatromanolaki et al. 2001; Raffel, Bhattacharyya et al. 2003). No regions of focal genomic loss associated with the older age group were identified.

4.3.1.11 Focal regions of genomic imbalance associated with anaplastic histology in intracranial ependymomas

Focal regions of copy number gain associated with a WHO grade III tumour classification (compared to a WHO grade II classification) were identified for both primary supratentorial and posterior fossa tumours and ranked according to statistical significance (Tables 4.10 and 4.11). The supratentorial cohort comprised six grade II and five grade III tumours, while the posterior fossa cohort comprised 13 grade II and 10 grade III tumours.

Anaplastic supratentorial ependymomas were associated with focal gains across chromosomes 7q, 11q, 13q and 19. Numerous genes were encompassed within these regions, such as COL4A1 (13q34; p = 0.015). Other genes included the fibroblast growth factor family member FGF3 (11q13.3; p = 0.015), the cyclin dependent kinase encoding CDK8 (13q12.13; p = 0.015) and the VEGF (vascular endothelial growth factor) receptor gene FLT1 (13q12.2; p = 0.015). Few focal regions of gain were associated with high grade posterior fossa ependymomas. Most were located on chromosome 12, specifically within 12q24.31, encompassing genes such as KNTC1/ROD (p = 0.024) which encodes a mitosis checkpoint protein. In contrast, sites of genomic loss were not associated with grade III ependymomas from either CNS location.

Table 4.6: Focal regions of genomic gain associated with primary paediatric posterior fossa ependymomas.

Locus	Start (bp)	End (bp)	Gene symbol	Additional genes in this region	PF	ST	p-value
1q23.3	163763105	163772128	RXRG (d) – LRRC52 (u)		10/24	0/11	0.015
1q32.1	201414285	201432734	CHI3L1 (u) – CHIT1 (u)	CHI3L1	10/24	0/11	0.015
1q42.13	225822390	225908366	ZNF678 (i)		10/24	0/11	0.015
1q42.13	225925871	226167742	ENST00000323562 (d) – WNT9A (u)	C1orf142, MPN2	10/24	0/11	0.015
1q42.2	230984253	231057198	ENST00000366656 (i)		10/24	0/11	0.015
1q42.2	232617486	232651815	TARBP1 (i)		10/24	0/11	0.015
1q42.3	232704412	232792979	TARBP1 (u)		10/24	0/11	0.015
1q21.1	146183314	146292286	ENST00000386140 (d) - ENST00000365538 (d)	ENST00000365538	9/24	0/11	0.033
1q23.3	159925404	159951193	FCGR2C (d) – FCRLA (d)	FCRLA	9/24	0/11	0.033
1q23.3	159971944	160297525	FCRLB (d) – OLFML2B (d)	ATF6, OLFML2B	9/24	0/11	0.033
1q24.1	165086908	165099823	POGK (i) – TADA1L (i)		9/24	0/11	0.033
1q24.3	170104490	170339258	DNM3 (i)		9/24	0/11	0.033
1q25.2	175395635	175413784	ASTN1 (i) – FAM58B (i)		9/24	0/11	0.033
1q25.3	180687071	180723046	RGSL2 (i)		9/24	0/11	0.033
1q25.3	181989553	182289682	RGL1 (i) – C1orf19 (i)	GLT25D2	9/24	0/11	0.033
1q32.1	199371212	199529953	TMEM9 (3'UTR) – PKP1 (i)	ENST00000263947, ENST00000367325 DKFZp434B1231	9/24	0/11	0.033
1q32.1	201339300	201412789	ADORA1(i) – MYBPH (u)		9/24	0/11	0.033
1q32.1	201454716	201475471	CHIT1 (i-d)		9/24	0/11	0.033
1q32.1	201955555	202096361	ATP2B4 (i) – ZC3H11A (d)	ZC3H11A, LAX1	9/24	0/11	0.033
1q41	212643105	212714251	PTPN14 (i)		9/24	0/11	0.033
1q42.12	224512151	224652139	LIN9 (i) – PARP1 (i)		9/24	0/11	0.033
1q42.12	224703718	224738529	ENST00000363938 (d) – ENST00000317080 (u)		9/24	0/11	0.033
1q42.12	224749874	224777493	C1orf95 (u)		9/24	0/11	0.033
1q42.13	225330114	225488622	CDC42BPA (i)		9/24	0/11	0.033
1q42.2	230505719	230525587	ENST00000366656 (i)		9/24	0/11	0.033
1q42.2	232486852	232542353	SLC35F3 (i) – GNPAT (u)		9/24	0/11	0.033
1q42.3	232826070	232893942	IRF2BP2 (u)		9/24	0/11	0.033
1q43	237772841	237847789	CHRM3 (u)		9/24	0/11	0.033

Locus	Start(bp)	End(bp)	Gene symbol	Additional genes in this region	PF	ST	p-value
1q44	241777030	242064416	AKT3 (i)		9/24	0/11	0.033
1q44	243159455	243272300	HRNPU (d) – EFCAB2 (i)		9/24	0/11	0.033
1q44	243448255	243490562	ENST00000329504 (u)		9/24	0/11	0.033
1q44	243515509	243720615	ENST00000366519 (d) – ENST00000366518 (u)	ENST00000329504	9/24	0/11	0.033
1q44	245790293	246010859	C1orf150 (i) – OR1C1 (d)	OR6F1, OR1C1	9/24	0/11	0.033
1q44	246038341	246071237	OR5AT1(d) – OR11L1 (CDS)		9/24	0/11	0.033
1q44	246098389	246346107	TRIM58 (i) – OR2L13 (d)	OR2AK2, OR2L13	9/24	0/11	0.033
1q44	246531667	246645245	OR2T12 (d) – OR2T2 (u)	OR2T6	9/24	0/11	0.033
1q44	74268812	74352226	ENST00000362858 (d)		9/24	0/11	0.033
20q13.33	60094968	60121514	TAF4 (d) – LSM14B (u)		9/24	0/11	0.033

Note: The gene symbols representing the start and end of each specified genomic region are shown, together with additional encompassed genes identified from the Affymetrix® annotation file (Netaffx file build 07.12.07). Gained regions associated with posterior fossa tumours are ranked in order of statistical significance, as deemed by a two-tailed Fisher's exact test (SPSS 16) performed against supratentorial tumours. The precise numbers of samples within the posterior fossa and supratentorial cohorts demonstrating genomic gain for each region are included. PF = posterior fossa, ST = supratentorial, bp = base pair, i = intronic, u = upstream of annotated gene, d = downstream of annotated gene, 3'UTR = 3' untranslated region. The gene highlighted in red had its copy number validated by real time qPCR.

Table 4.7: Focal regions of genomic gain associated with primary paediatric supratentorial ependymomas.

Locus	Start (bp)	End (bp)	Gene symbol	Additional genes in this region	ST	PF	p-value
1p35.1	33741021	33753817	CSMD2 (u-3'UTR)		5/11	0/24	0.001
1p32.3	52966384	52983591	ZYG11B (i)		6/11	2/24	0.006
1p33	47592596	47647450	CMPK (i) – FOXE3 (u)		4/11	0/24	0.006
2q13	113700504	113745190	PAX8 (i-d)	ENST00000333145	4/11	0/24	0.006
6q21	107137365	107147668	RTN4IP1 (i)		4/11	0/24	0.006
6q25.3	158788606	158801752	TULP4 (i)		4/11	0/24	0.006
6q25.3	159639229	159650534	FLJ27255 (d)		4/11	0/24	0.006
10q26.3	131615303	131697615	EBF3 (i-d)		6/11	3/24	0.015
19q13.31	49678063	49691015	ZNF180 (i)		6/11	3/24	0.015
6p25.1	4666045	4718085	CDYL (i)		5/11	2/24	0.02
8p12	38229372	38345434	DDHD2 (i) – WHSC1L1 (i)		5/11	2/24	0.02
11p11.2	44632464	44649769	TP53I11 (d) - TSPAN18 (u)		5/11	2/24	0.02
13q34	109634045	109748036	COL4A1 (i)		5/11	2/24	0.02

Note: The gene symbols representing the start and end of each specified genomic region are shown, together with additional encompassed genes identified from the Affymetrix® annotation file (Netaffx file build 07.12.07). Gained regions associated with supratentorial tumours are ranked in order of statistical significance, as deemed by a two-tailed Fisher's exact test (SPSS 16) performed against posterior fossa tumours. The precise numbers of samples within the supratentorial and posterior fossa cohorts demonstrating genomic gain for each region are included. ST = supratentorial, PF = posterior fossa, bp = base pair, i = intronic, u = upstream of annotated gene, d = downstream of annotated gene, 3'UTR = 3' untranslated region.

Table 4.8: Focal regions of genomic gain associated with primary paediatric posterior fossa ependymomas from patients aged below three years.

Locus	Start (bp)	End (bp)	Gene symbol	Additional genes in this region	U3	O3	p-value
2q37.3	240534261	240740155	FLJ45964 (d) – GPC1 (u)	NDUFA10, MYEOV2, OTOS	6/12	0/12	0.014
6p23	13895819	13944145	CCDC90A (u) – RNF182 (u)	CCDC90A	6/12	0/12	0.014
6q27	168073683	168101356	MLLT4 (i)		6/12	0/12	0.014
16p13.3	1499839	1566706	TELO2 (CDS) – IFT140 (i)		6/12	0/12	0.014
2p25.3	129574	168356	SH3YL1(d)		5/12	0/12	0.037
2p25.3	435185	604210	ENST00000389719 (i)		5/12	0/12	0.037
2p25.3	658618	926687	TMEM18 (3'UTR) – SNGT2 (u)		5/12	0/12	0.037
2p25.3	959866	1484084	SNTG2 (i) – TPO (i)		5/12	0/12	0.037
2q21.1	131525906	131671717	ENST00000354183 (i) – PLEKHB2 (i)		5/12	0/12	0.037
2q37.3	241674555	241787007	ENST00000310397 (i) – TMEM16G (i)	MTERFD2, PASK, PPPR17	5/12	0/12	0.037
3q21.2	127614144	127665610	CCDC37 (i) – ZXDC (i)		5/12	0/12	0.037
3q21.3	127698840	127772052	UROC1 (i) – TR2IT1 (u)		5/12	0/12	0.037
3q21.3	127812600	127922704	ENST00000360201 (CDS) – CHCHD6 (i)		5/12	0/12	0.037
6p25.3	402748	549537	IRF4 (d) – EXOC2 (i)		5/12	0/12	0.037
6p25.2	4045539	4074092	C6orf201 (i)		5/12	0/12	0.037
6p25.1	4093181	4118652	PECI (d)		5/12	0/12	0.037
6q27	167062637	167279633	ENST00000366867 (i) – RNASET2 (i)		5/12	0/12	0.037
6q27	167927207	167967136	C6orf123 (u-d)		5/12	0/12	0.037
6q27	169654969	169714869	THBS2(d) – WDR27 (i)		5/12	0/12	0.037
10p15.3	1406472	1660999	ADARB2 (i)		5/12	0/12	0.037
10p15.2	3063036	3138116	PFKP (u-i)		5/12	0/12	0.037
16q24.2	86333714	86349893	KLHDC4 (i)		5/12	0/12	0.037
17q21.2	36488761	36499917	ENST00000332991 (u-d)		5/12	0/12	0.037
17q21.2	36542729	36586246	KRTAP4-2 (d) – ENST00000377726 (u)		5/12	0/12	0.037
22q12.3	30905149	30918352	RP1-127L4.6 (d) – ENST00000248983 (i)		5/12	0/12	0.037
22q13.33	48481817	48564816	C22orf34 (d) – BRD1 (i)		5/12	0/12	0.037
22q13.33	48993966	49205915	RP3-402G11.5 (i) – SAPS2 (i)	TUBGCP6, HDAC10	5/12	0/12	0.037

Note: The gene symbols representing the start and end of each specified genomic region are shown, together with additional encompassed genes identified from the Affymetrix® annotation file (Netaffx file build 07.12.07). Gained regions associated with posterior fossa tumours from children aged less than three years are ranked in order of statistical significance, as deemed by a two-tailed Fisher's exact test (SPSS 16) performed against posterior fossa tumours from children aged above three years. The precise numbers of samples within the posterior fossa age cohorts demonstrating genomic gain for each region are included. U3 = posterior fossa tumours from children aged under three years, O3 = posterior fossa tumours from children aged over three years, bp = base pair, i = intronic, u = upstream of annotated gene, d = downstream of annotated gene.

Table 4.9: Focal regions of genomic gain associated with primary paediatric posterior fossa ependymomas from patients aged above three years.

Locus	Start (bp)	End (bp)	Gene symbol	Additional genes in this region	U3	O3	p-value
9p24.1	6889149	6953799	JMJD2C (i)		0/12	6/12	0.014
9q21.11	70922928	70936691	TJP2 (u)		0/12	6/12	0.014
9q21.11	70938747	70985333	FXN (d) – TJP2 (i)		0/12	6/12	0.014
9q31.3	111892156	111904290	AKAP2 (i)		0/12	6/12	0.014
9q31.3	112050484	112063686	TXN (i) – TXNDC8 (u)		0/12	6/12	0.014
9q32	115654279	115684020	ZNF613 (u)		0/12	6/12	0.014
9q32	115973270	116134866	COL27A1 (i) – ORM2 (i)	ORM1	0/12	6/12	0.014
9q33.3	125900331	125913604	NEK6 (u)		0/12	6/12	0.014
18q21.1	43913365	43948279	ZBTB7C (d) - ENST00000384217 (d)	ENST00000384217	0/12	6/12	0.014
1q23.3	161319642	161340376	RGS4 (d) – RGS5 (u)		1/12	7/12	0.027
1q24.2	167176576	167249654	ATP1B1 (u)		1/12	7/12	0.027
1q24.3	170360753	170394207	DNM3 (i)		1/12	7/12	0.027
1q41	219488068	219510247	HLX (d) – C1orf140 (u)		1/12	7/12	0.027
1q43	237490347	237768272	CHRM3 (u)		1/12	7/12	0.027
9p13.1	38512055	38761831	IGFBPL1 (d) – CNTNAP3 (u)	ANKRD18A, ENST00000377689 ENST00000357927	1/12	7/12	0.027

Note: The gene symbols representing the start and end of each specified genomic region are shown, together with additional encompassed genes identified from the Affymetrix® annotation file (Netaffx file build 07.12.07). Gained regions associated with posterior fossa tumours from children aged above three years are ranked in order of statistical significance, as deemed by a two-tailed Fisher's exact test (SPSS 16) performed against posterior fossa tumours from children aged below three years. The precise numbers of samples within the posterior fossa age cohorts demonstrating genomic gain for each region are included. U3 = posterior fossa tumours from children aged under three years, O3 = posterior fossa tumours from children aged over three years, bp = base pair, i = intronic, u = upstream of annotated gene, d = downstream of annotated gene. The gene highlighted in red had its copy number validated by real time qPCR.

Table 4.10: Focal regions of genomic gain associated with primary paediatric supratentorial ependymomas of WHO grade III.

Cytoband	Start (bp)	End (bp)	Gene symbol	Additional genes in this region	Gd3	Gd2	p-value
7q34	138643716	138814158	ENST00000384116 (d) – KLRG2 (i)	C7orf55, LUC7L2	4/5	0/6	0.015
7q36.3	155958047	156088256	C7orf13 (d)		4/5	0/6	0.015
7q36.3	156454985	156476488	NOM1 (CDS-d)		4/5	0/6	0.015
11q13.2	68933687	68990408	FLJ44258 (d) – CCND1 (u)		4/5	0/6	0.015
11q13.3	69198344	69363887	ORAOV1 (i) – FGF3 (d)	FGF3	4/5	0/6	0.015
11q13.4	72082606	72206621	CENTD2 (i) – ATG16L2 (i)		4/5	0/6	0.015
11q13.4	118774194	118861892	USP2 (u) – THY1 (d)	THY1	4/5	0/6	0.015
11q23.3	118896089	119078599	PVRL1 (u-i)	NM_001001681	4/5	0/6	0.015
13q12.11	19666144	19748497	GJB6 (u) – CRYL1 (d)	GJB6	4/5	0/6	0.015
13q12.12	23996413	24011645	PARP4 (u)		4/5	0/6	0.015
13q12.13	25651251	25734780	RNF6 (i) – CDK8 (i)		4/5	0/6	0.015
13q12.2	27299578	27312265	GSX1 (d) – PDX1 (u)		4/5	0/6	0.015
13q12.2	27443268	27482744	CDX2 (u) – FLT3 (i)		4/5	0/6	0.015
13q12.2	27724983	27806090	PAN3 (i) – FLT1 (i)		4/5	0/6	0.015
13q14.11	40530244	41053103	ELF1 (d) – C13orf15 (d)	KBTBD6, WBP4, MTRF1, NARG1L, C13orf15	4/5	0/6	0.015
13q13.2	97622038	97645101	FARP1 (i)		4/5	0/6	0.015
13q32.3	99051217	99063992	TM9SF2 (d) – CLYBL (i)		4/5	0/6	0.015
13q34	109605936	109629867	COL4A1 (i)		4/5	0/6	0.015
19p13.3	6147373	6218757	ACSBG2 (d) – MLLT1 (i)		4/5	0/6	0.015
19p12	23092981	23156405	ENST00000327867 (u)		4/5	0/6	0.015
19q13.43	63631939	63731511	TRIM28 (i)	ZNF132	4/5	0/6	0.015

Note: The gene symbols representing the start and end of each specified genomic region are shown, together with additional encompassed genes identified from the Affymetrix® annotation file (Netaffx file build 07.12.07). Gained regions associated with WHO grade III supratentorial tumours are ranked in order of statistical significance, as deemed by a two-tailed Fisher's exact test (SPSS 16) performed against WHO grade II tumours. The precise numbers of samples within the supratentorial grade cohorts demonstrating genomic gain for each region are included. Gd3 = WHO grade III supratentorial tumours, Gd2 = WHO grade II supratentorial tumours, bp = base pair, i = intronic, e = exonic, u = upstream of annotated gene, d = downstream of annotated gene.

Table 4.11: Focal regions of genomic gain associated with primary paediatric posterior fossa ependymomas of WHO grade III.

Cytoband	Start (bp)	End (bp)	Gene symbol	Additional genes in this region	Gd3	Gd2	p-value
11p15.1	16984904	17034804	PLEKHA7 (i) – RPS13 (u)		4/10	0/13	0.024
12q24.31	121469421	121557232	CLIP1 (i) – RSRC2 (i)	ZCCHC8	4/10	0/13	0.024
12q24.31	121628547	121729801	KNTC1 (i) – GPR109A (u)		4/10	0/13	0.024
12q24.31	124022488	124105945	DHX37 (i) – AACS (u)	BRI3BP	4/10	0/13	0.024
19p13.2	9447719	9462391	ZNF560 (i)		4/10	0/13	0.024

Note: The gene symbols representing the start and end of each specified genomic region are shown, together with additional encompassed genes identified from the Affymetrix® annotation file (Netaffx file build 07.12.07). Gained regions associated with WHO grade III posterior fossa tumours are ranked in order of statistical significance, as deemed by a two-tailed Fisher's exact test (SPSS 16) performed against WHO grade II tumours. The precise numbers of samples within the posterior fossa tumour grade cohorts demonstrating genomic gain for each region are included. Gd3 = WHO grade III posterior fossa tumours, Gd2 = WHO grade II posterior fossa tumours, bp = base pair, i = intronic, u = upstream of annotated gene, d = downstream of annotated gene.

170

4.3.2 Genomic gain encompassing selected genes on chromosome 1q in intracranial ependymoma and association with patient survival

Survival analysis was performed across the primary SNP array cohort to establish whether ependymomas demonstrating focal gain of regions harbouring selected genes within chromosome 1q were associated with an adverse patient outcome. The genes NSL1, NAV1, CHI3L1/YKL40, PRUNE and BNIPL were chosen as a result of preceding findings from the SNP array analysis. NSL1 (1q32.3) was within one of the most frequently gained focal regions in the primary intracranial cohort. NAV1 (1q32.1) was located in the most common focal region of increased copy number in the intracranial recurrent cohort. Gain of the locus encompassing CHI3L1/YKL40 (1q32.1) was associated exclusively with posterior fossa ependymomas and was also a common feature of intracranial recurrent tumours. The 1q21.2 – 21.3 region was previously identified to be the most common imbalanced cytoband in a set of ependymomas associated with an adverse patient outcome ('group two') and was the most frequently gained chromosome 1q locus across the primary cohort (Chapter 3, section 3.3.5). High resolution analysis of the 1q21.2 – 21.3 region was thus performed across the primary ependymoma cohort to detect the most commonly shared site of focal gain (at least five consecutive SNPs spanning a region greater than 10kb) which was associated with the most unfavourable outcome for the group from accompanying survival data. This process identified a small region on 1q21.3, which encompassed two genes according to the accompanying annotation data file. These were the nucleotide phosphodiesterase encoding PRUNE and BNIPL, a gene involved in regulating apoptosis. This locus was thereby termed 'PRUNE/BNIPL', although subsequent Ensembl database analysis also identified two other genes encompassed within the region, the Rho-GTPase modulator CDC42SE1 and AF1q, a gene of unknown function that has been reported to promote the proliferation and invasion of breast cancer cells in vitro (Chang, Li et al. 2008).

Univariate analysis (Table 4.12) revealed NAV1 gain was associated with a significantly worse event-free patient survival (estimated mean EFS: 2.3 ± 0.5 years versus 6.6 ± 1.3 years, p = 0.013) (Figure 4.5). This finding was replicated on multivariate analysis (hazards ratio 3.18 (95 % CI $1.065 - 9.52$), p = 0.038) (Table 4.13). No association with NAV1 gain and overall survival was observed.
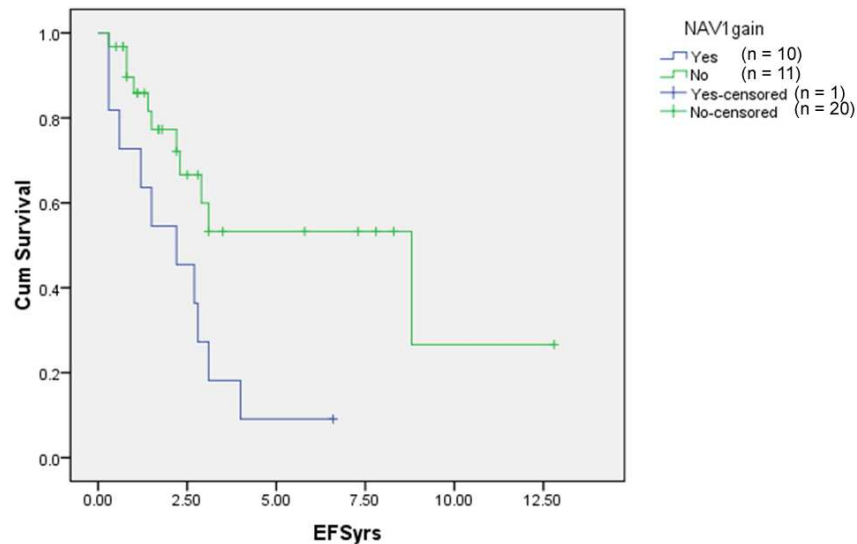
Figure 4.5: Kaplan-Meier EFS curve, comparing 11 ependymoma patients demonstrating SNP array gain of NAV1 (blue lines) against the rest of the primary SNP array cohort (green lines). Patients with NAV1 gain had a significantly worse EFS (percentage from each group attaining five year EFS = 9.1 $\pm$ 8.7 % (NAV1 group) versus 53.3 $\pm$ 11.6 % (remaining primary cohort), p = 0.013).

Univariate analysis (Table 4.12) also revealed gain of the PRUNE/BNIPL region was associated with a significantly worse event-free and overall patient survival (estimated mean EFS: 3.1 $\pm$ 0.8 years versus 7.4 $\pm$ 1.3 years, p = 0.045; estimated mean OS: 5.9 $\pm$ 1 years versus 16.6 $\pm$ 2 years, p = 0.023) (Figure 4.6).
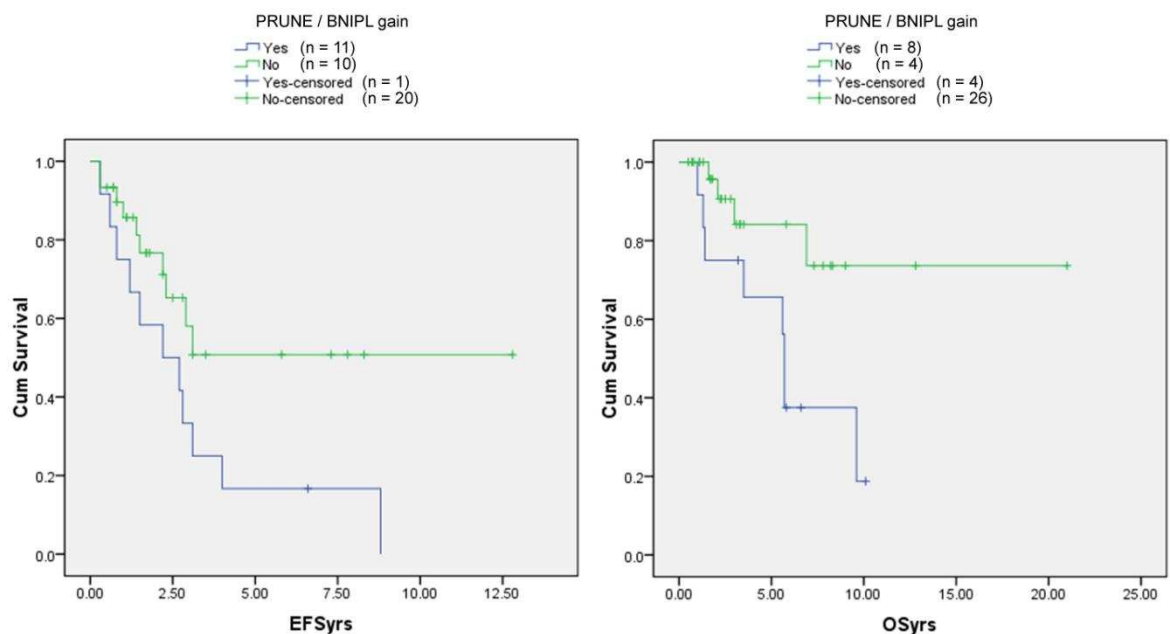
Figure 4.6: Kaplan-Meier EFS and OS curves, comparing 12 ependymoma patients demonstrating SNP array gain of the PRUNE/BNIPL region (blue lines) against the rest of the primary SNP array cohort (green lines). Patients with gain of this focal region had a significantly worse EFS (percentage from each group attaining five year EFS = 16.7 $\pm$ 10.8 % (PRUNE/BNIPL gain group) versus 50.8 $\pm$ 12.1 % (remaining primary cohort), p = 0.045) and OS (percentage from each group attaining five year OS = 65.6 $\pm$ 14 % (PRUNE/BNIPL gain group) versus 84.1 $\pm$ 8.6 % (remaining primary cohort), p = 0.023).

172

Table 4.12: Univariate survival data for gain of selected genes on chromosome 1q identified from the 500K SNP array analysis.

Gene identified from SNP analysis	Number of cases with gain (%)	Percentage of cases attaining five year EFS (cases with gain versus without gain)	P value	Percentage of cases attaining five year OS (cases with gain versus without gain)	p-value
NSL1	18 (43 %)	30.3 % $\pm$ 11.2 % vs 46.4 % $\pm$ 14.6 %	0.365	76 % $\pm$ 10.4 % vs 77.7 % $\pm$ 11.8 %	0.479
CHI3L1	11 (26 %)	10.6 % $\pm$ 10 % vs 51.5 % $\pm$ 11.2 %	0.09	71.6 % $\pm$ 14 % vs 79.7 % $\pm$ 9.4 %	0.293
NAV1	11 (26 %)	9.1 % $\pm$ 8.7 % vs 53.3 % $\pm$ 11.6 %	0.013	71.6 % $\pm$ 14 % vs 81 % $\pm$ 8.8 %	0.220
PRUNE & BNIPL	12 (29 %)	16.7 % $\pm$ 10.8 % vs 50.8 % $\pm$ 12.1 %	0.045	65.6 % $\pm$ 14 % vs 84.1 % $\pm$ 8.6 %	0.023

The primary SNP array cohort of 42 ependymomas was analysed. The number of tumours demonstrating gain of each selected gene within the cohort is shown both numerically and as a percentage of the primary cohort. Statistically significant differences in survival between cases with and without gain of the selected gene are highlighted in yellow. Trends towards significance are highlighted in pale yellow.

Table 4.13: Multivariate event-free survival analysis incorporating SNP array gain of NAV1.

Cox regression multivariate analysis (n=42)	Event-Free Survival		
Factor	Hazards ratio	95 % CI	p-value
Histology (WHO grade II vs III)	1.069	0.432 – 2.648	0.885
Tumour location (ST vs PF)	0.609	0.195 – 1.901	0.393
Resection status (complete vs incomplete)	2.261	0.871 – 5.87	0.094
Patient age (below 3 years vs above 3 years)	1.634	0.429 – 6.232	0.472
NAV1 (gain vs no gain)	3.183	1.065 – 9.520	0.038

The primary SNP array cohort of 42 ependymomas was analysed. Radiotherapy and chemotherapy were not included as variables for event-free survival due to missing clinical data regarding commencement dates (Chapter 3, section 3.2.3). 95 % CI = 95 % confidence interval, ST = supratentorial, PF = posterior fossa. Statistically significant results are highlighted in yellow. Trends towards significance are highlighted in pale yellow.

Table 4.14: Multivariate overall survival analysis incorporating SNP array gain of the PRUNE/BNIPL region.

Cox regression multivariate analysis (n=42)	Overall Survival		
Factor	Hazards ratio	95 % CI	p-value
Histology (WHO grade II vs III)	1.074	0.291 – 3.961	0.914
Tumour location (ST vs PF)	0.540	0.066 – 4.413	0.565
Resection status (complete vs incomplete)	1.176	0.274 – 6.242	0.819
Patient age (below 3 years vs above 3 years)	1.499	0.209 – 10.773	0.687
Radiotherapy (received vs not received)	0.308	0.064 – 1.472	0.14
Chemotherapy (received vs not received)	1.752	0.226 – 13.567	0.591
PRUNE/BNIPL region (gain vs no gain)	4.295	1.071 – 17.227	0.04

The primary SNP array cohort of 42 ependymomas was analysed. ST = supratentorial, PF = posterior fossa, 95 % CI = 95 % confidence interval. Significant results are highlighted in yellow.

While the adverse effect on event-free survival reached a trend towards statistical significance on multivariate analysis (hazards ratio 2.425 (95 % CI 0.89 – 6.59) p = 0.083), gain of the PRUNE/BNIPL region remained significantly associated with a worse overall patient survival (hazards ratio 4.295 (95 % CI 1.071 – 17.227, p = 0.04) (Table 4.14).

No statistically significant association with survival was identified for gain of the regions encompassing NSL1 or CHI3L1/YKL40 on univariate (Table 4.12) or multivariate analysis.

4.3.3 Candidate regions of focal aUPD in paediatric ependymoma.

SNP probes demonstrating aUPD were present, albeit not extensively, in all 44 paediatric ependymomas analysed against patient-matched constitutional DNA (Figure 4.7). In the majority of tumours (33/44, 75 %), aUPD was evident in fewer than 5 % of the total number of SNP probes examined. Indeed, in six cases (samples 6P, 27P, 30P, 31P, 32P and 45P), less than 1 % of the probes revealed aUPD.

Focal, gene-specific regions of aUPD were detected in 12/38 (32 %) primary and 4/6 (67 %) intracranial first recurrent tumours (Appendix 10H). Of the twelve primary ependymomas that demonstrated focal aUPD, five were located in the posterior fossa (samples 9P, 21P, 25P, 28P, 40P), four were spinal tumours (24P, 27P, 38P, 42P), two were supratentorial (20P, 23P) and for one case the location was unknown (3P). Of the four recurrent ependymomas that demonstrated focal aUPD, three were supratentorial tumours (18R1, 20R1 and 43R1) and one was located in the posterior fossa (16R1). All of the regions of focal aUPD detected in the recurrent ependymomas were not present in the corresponding patient's primary tumour.

Focal regions of aUPD were not frequently shared across the primary ependymoma cohort, the most common regions being observed in only two tumours (2/38, 5 %). Nevertheless, various loci were identified across chromosomes 1p, 2, 6 and 13q (Table 4.15). These focal regions encompassed genes reported to be associated with DNA replication (RPA2 (1p35.3); RBMS1/MSSP2 (2q24.2)), cell motility and invasiveness

(ITGB6 (2q24.2), tumour angiogenesis (BAI3 (6q12); ELF1 (13q14.11)) and neurite growth and guidance (DPYSL5/CRMP5 (2p23.3); TIAM2/STEF (6q25.2)) (Takai, Nishita et al. 1994; Shiratsuchi, Nishimori et al. 1997; Fukada, Watakabe et al. 2000; Matsuo, Hoshino et al. 2002; Weisshart, Pestryakov et al. 2004; Huang, Brown et al. 2006; Ramsay, Keppler et al. 2007). Other genes identified included DNAJC6 (1p31.3), which encodes a HSP40 heat shock protein family member and ERBB4 (2q34), a gene encoding a member of the RTK1 (Receptor Tyrosine Kinase 1) protein family also which also comprises ERBB2, ERBB3 and EGFR.

Table 4.15: The most common focal regions of aUPD in 38 primary ependymomas.

Locus	Shared region (physical position bp)	Gene symbol	Number of samples
1p35.3	28093071 - 28106912	RPA2	2/38 (5 %)
1p31.3	65578142 - 65616570	DNAJC6	2/38 (5 %)
1p22.3	85207201-85228120	MCOLN2	2/38 (5 %)
2p24.3	15335505 - 15368097	NAG	2/38 (5 %)
2p23.3	26934210 - 26986326	DPYSL5	2/38 (5 %)
2p13.2	73593933 -73647562	ALMS1	2/38 (5 %)
2q13	111344951 - 111357734	ACOXL	2/38 (5 %)
2q14.1	115957656 -115994455	DPP10	2/38 (5 %)
2q21.2	132932549 - 132944012	GPR39	2/38 (5 %)
2q21.3	135328892 - 135373941	ACMSD	2/38 (5 %)
2q22.1	137685713 - 137700731	ENST00000272643	2/38 (5 %)
2q24.2	159792719 - 159815318	TANC1, WD50B1	2/38 (5 %)
2q24.2	160692135 - 160722430	ITGB6	2/38 (5 %)
2q24.2	160950541 - 161037890	RBMS1	2/38 (5 %)
2q31.3	182597228 - 182623659	ENST00000280295	2/38 (5 %)
2q32.1	183013344 - 183024436	PDE1A	2/38 (5 %)
2q33.3	207693709 - 207715065	KLF7	2/38 (5 %)
2q34	212395736 - 212409273	ERBB4	2/38 (5 %)
2q36.3	225978417 - 225998949	ENST00000272907	2/38 (5 %)
6p25.2	3364148 - 3379241	ENST00000380298	2/38 (5 %)
6p22.3	16754457 - 16767074	ATXN1	2/38 (5 %)
6q12	69597671-69611902	BAI3	2/38 (5 %)
6q14.1	83709716-83724186	C6orf157	2/38 (5 %)
6q25.2	155277331 - 155287746	TIAM2	2/38 (5 %)
13q13.1	32221486 - 32246568	PDS5B	2/38 (5 %)
13q14.11	40436259 - 40476109	ELF1	2/38 (5 %)

Note: The physical genomic position of each shared region of aUPD is shown, together with the corresponding gene symbol from the annotation file (Netaffx file build 07.12.07) and the frequency of occurrence across the cohort. bp = base pairs.

Only one site, within chromosome 9p22.1, was found to be the most common focal region of aUPD across the intracranial first recurrent cohort. This was observed in two tumours (2/6 samples, 33 %) (Table 4.16). The region encompassed a gene of unknown function, C9orf138.
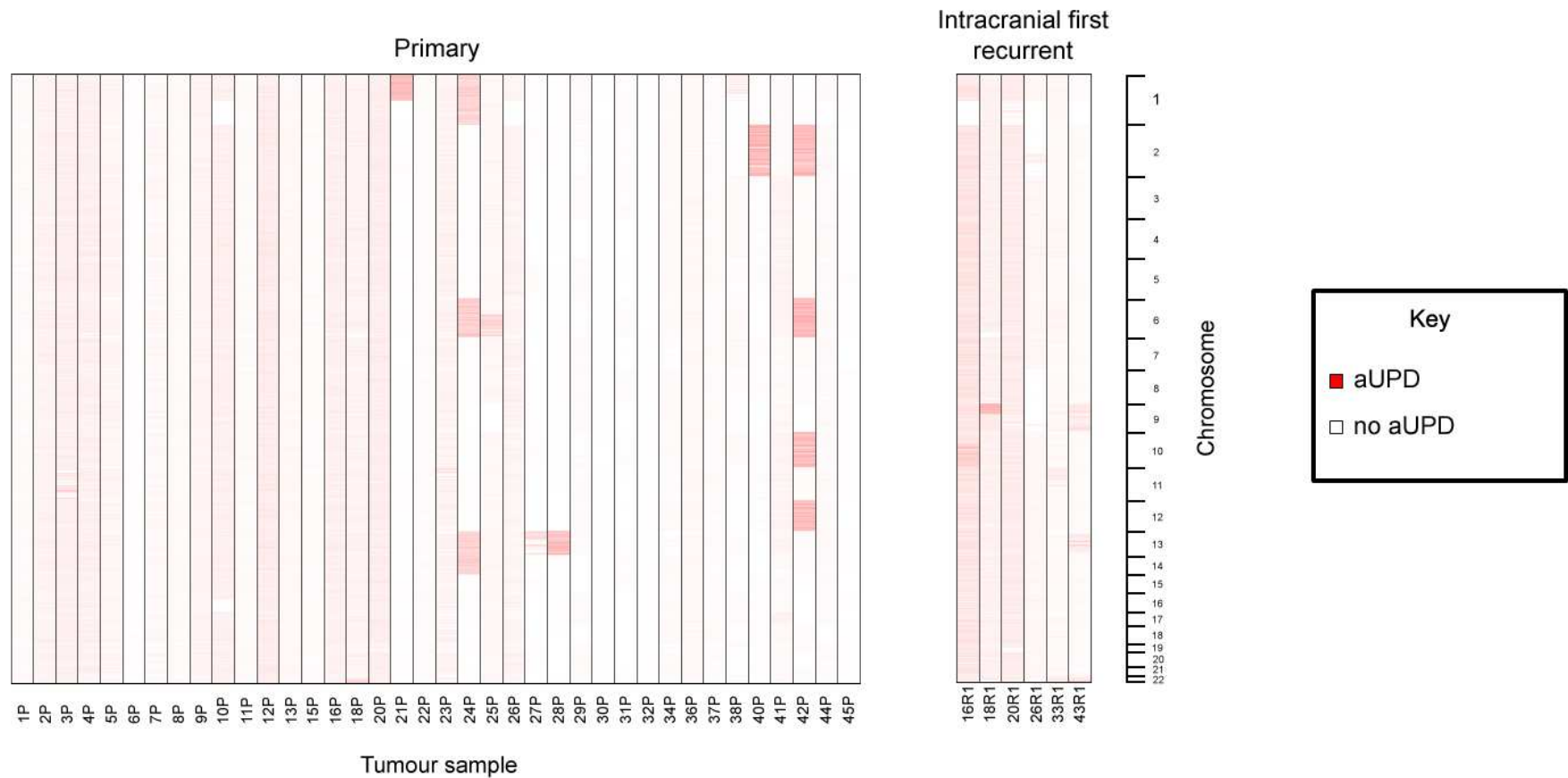
Figure 4.7: Spotfire Decision Site® heatmap identifying regions of aUPD across the genome for 38 primary and six intracranial first recurrent paediatric ependymomas, as detected using the Affymetrix® 500K SNP array. Regions of aUPD are coloured red, while regions not demonstrating aUPD are coloured white. P = primary, R1 = 1st recurrence.

176

Table 4.16: The most common focal region of aUPD in six intracranial first recurrent ependymomas.

Locus	Shared region (physical position bp)	Gene symbol	Number of samples
9p22.1	18984588 - 19003483	C9orf138	2/6 (33 %)

Note: The physical genomic position of each shared region of aUPD is shown, together with the corresponding gene symbol from the Affymetrix® annotation file (Netaffx file build 07.12.07) and the frequency of occurrence across the cohort. bp = base pairs.

4.3.4 Genes differentially methylated between clinical subgroups in paediatric ependymoma

Of all the clinical subgroups analysed (section 4.2.2.1), significant differential gene methylation was only identified between ependymomas from different CNS locations (Tables 4.17 – 4.19).

4.3.4.1 Differentially methylated genes characterising spinal and intracranial ependymomas

Compared to spinal ependymomas, intracranial tumours demonstrated hypermethylation of genes such as the protein kinase encoding RIPK3 (14q11.2) and potential tumour suppressor RASSF1A (3p21.3), both of which have been reported to promote apoptosis (Yu, Huang et al. 1999; Oh, Lee et al. 2006) (Table 4.17). Spinal ependymomas revealed relative hypomethylation of the transcription factor EYA4 (6q23), an upregulated signature gene for spinal ependymomas (Taylor, Poppleton et al. 2005).

In contrast, intracranial ependymomas demonstrated comparative hypomethylation of genes such as the RAS oncogene superfamily member RAN (12q24.3) responsible for regulating DNA synthesis. Other genes included the transcription factor encoding SPDEF (6p21.3), the transforming growth factor gene TGFB3 (14q24) and BCR (22q11.23), a gene of uncertain function that can form the BCR-ABL fusion gene found in chronic myeloid leukaemia. The hypomethylation profile of the latter three genes in intracranial ependymomas was attributed to the contribution from posterior fossa tumours (Table 4.19 and section 4.3.4.2).

The methylation status of the HOX family members, including HOXA5 (7p15.2), were not significantly different between spinal and intracranial tumours.

Table 4.17: Genes with a differential methylation profile between paediatric spinal and intracranial ependymomas.

SP hypomethylated genes		SP β value	Intracranial β value	β difference	p-value
CRIP1	(probe P874R)	0.146	0.749	0.603	5.18×10^{-4}
KLK11	(probe P103R)	0.487	0.897	0.410	5.72×10^{-4}
GP1BB	(probe E23F)	0.486	0.838	0.352	9.38×10^{-4}
IRAK3	(probe P185F)	0.315	0.715	0.400	2.06×10^{-3}
RIPK3	(probe P24F)	0.441	0.865	0.424	2.43×10^{-3}
MEST	(probe P4F)	0.406	0.767	0.361	6.19×10^{-3}
RASSF1	(probe E116F)	0.439	0.783	0.344	7.48×10^{-3}
EYA4	(probe P794F)	0.276	0.699	0.423	7.92×10^{-3}
SP hypermethylated genes		**SP β value**	**Intracranial β value**	**β difference**	**p-value**
MC2R	(probe P1025F)	0.873	0.243	0.630	3.3×10^{-4}
JAK3	(probe P1075R)	0.817	0.331	0.486	3.4×10^{-4}
TGFB3	(probe E58R)	0.863	0.326	0.537	5.18×10^{-4}
BCR	(probe P346F)	0.835	0.272	0.563	5.72×10^{-4}
SPDEF	(probe P6R)	0.655	0.285	0.370	6.19×10^{-4}
HRASLS	(probe P353R)	0.821	0.428	0.393	6.53×10^{-4}
MBD2	(probe P233F)	0.766	0.303	0.463	6.82×10^{-4}
PYCARD	(probe P393F)	0.617	0.192	0.425	9.77×10^{-4}
IL1RN	(probe P93R)	0.683	0.319	0.364	1.46×10^{-3}
AOC3	(probe P890R)	0.759	0.397	0.362	2.43×10^{-3}
TRPM5	(probe P721F)	0.789	0.401	0.388	3.1×10^{-3}
CCKAR	(probe E79F)	0.709	0.329	0.380	3.37×10^{-3}
RAN	(probe P581R)	0.798	0.438	0.360	4.22×10^{-3}

Genes with a methylation profile differentiating paediatric spinal ependymomas from intracranial tumours are shown. A methylation score (β value) of one represents complete methylation of a gene's CpG target site, while a score of zero represents no methylation of a CpG target site. The p-values shown are post Benjamini-Hochberg False Discovery Rate correction. The gene highlighted in red is an upregulated signature gene of spinal ependymomas established from gene expression array work (Taylor, Poppleton et al. 2005). SP = spinal.

4.3.4.2 Differentially methylated genes characterising supratentorial and posterior fossa ependymomas

Several genes revealed distinct methylation profiles that were associated with either supratentorial or posterior fossa ependymomas (Tables 4.18 and 4.19). Moreover, certain genes demonstrated significantly different methylation states between tumours from the two intracranial locations. For instance, the Wnt pathway member WNT10B (12q13) and tyrosine kinase encoding gene HCK (20q11) demonstrated hypomethylation amongst supratentorial ependymomas, yet were both relatively hypermethylated across the posterior fossa cohort.

Likewise, particular genes were hypermethylated in the supratentorial tumour group yet hypomethylated in posterior fossa ependymomas, such as the growth factor encoding gene LEFTY2/TGFB4 (1q42.1) and PPARG (3p25), a member of the peroxisome proliferator receptor family. Indeed, correlation of the differential methylation data with the expression array findings of Taylor and colleagues (Taylor, Poppleton et al. 2005) identified PPARG, together with the aforementioned SPDEF and BCR, as genes demonstrating both significant hypomethylation and characteristic overexpression in posterior fossa ependymomas. MMP2 (16q13) and CHI3L2 (1p13.3) were also comparatively hypomethylated in the posterior fossa cohort. While MMP2 encodes an extracellular matrix enzyme that degrades type IV collagen, CHI3L2 belongs to the chitinase gene family already associated with infratentorial ependymomas through genomic gain, as established from the SNP array work (section 4.3.1.7).

Table 4.18: Genes with a methylation profile characterising paediatric supratentorial ependymomas.

ST hypomethylated genes		ST β value	Other location β value	β difference	p-value
EMR3	(probe P39R)	0.435	0.911	0.476	1.29×10^{-5}
WNT10B	(probe P993F)	0.238	0.758	0.520	1.61×10^{-4}
HCK	(probe P858F)	0.202	0.702	0.500	4.53×10^{-4}
ST hypermethylated genes		ST β value	Other location β value	β difference	p-value
NAT2	(probe P11F)	0.897	0.539	0.358	1.29×10^{-5}
MAP3K1	(probe E81F)	0.631	0.210	0.421	4.84×10^{-5}
SH3BP2	(probe E18F)	0.528	0.148	0.380	1.93×10^{-4}
RBP1	(probe P426R)	0.633	0.122	0.511	6.8×10^{-4}
LEFTY2	(probe P719F)	0.836	0.433	0.403	1.05×10^{-3}
PPARG	(probe P693F)	0.863	0.383	0.480	1.6×10^{-3}
LCN2	(probe P86R)	0.672	0.300	0.372	1.95×10^{-3}
ASCL2	(probe P609R)	0.712	0.276	0.436	2.73×10^{-3}
EYA4	(probe P794F)	0.925	0.527	0.398	3.72×10^{-3}

Genes with a methylation profile differentiating paediatric supratentorial ependymomas from tumours located in the posterior fossa and spinal canal are shown. A methylation score (β value) of one represents complete methylation of a gene's CpG target site, while a score of zero represents no methylation of a CpG target site. The p-values shown are post Benjamini-Hochberg False Discovery Rate correction. ST = supratentorial.

Table 4.19: Genes with a methylation profile characterising paediatric posterior fossa ependymomas

PF hypomethylated genes		PF β value	Other location β value	β difference	p-value
VAMP8	(probe E7F)	0.260	0.800	0.540	2.5×10^{-13}
EPHX1	(probe P1358R)	0.182	0.655	0.637	2.12×10^{-13}
B3GALT5	(probe E246R)	0.444	0.845	0.401	6.36×10^{-11}
OSM	(probe P34F)	0.400	0.895	0.495	9.72×10^{-11}
KRT5	(probe P308F)	0.483	0.844	0.361	1.37×10^{-10}
PPARG	(probe P693F)	0.259	0.836	0.577	4.79×10^{-10}
SEPT9	(probe P374F)	0.266	0.904	0.638	5.05×10^{-10}
LEFTY2	(probe P719F)	0.328	0.814	0.486	5.35×10^{-10}
CHI3L2	(probe E10F)	0.518	0.904	0.386	1.37×10^{-8}
TGFB3	(probe E58R)	0.219	0.695	0.476	1.37×10^{-8}
MMP2	(probe P303R)	0.070	0.570	0.500	1.26×10^{-7}
WRN	(probe P969F)	0.241	0.593	0.352	4.1×10^{-7}
SPDEF	(probe P6R)	0.192	0.567	0.375	7.24×10^{-7}
BCR	(probe P346F)	0.203	0.599	0.396	1.9×10^{-6}
MAPK1 0	(probe E26F)	0.347	0.825	0.478	4.22×10^{-5}
PF hypermethylated genes		PF β value	Other location β value	β difference	p-value
HCK	(probe P858F)	0.787	0.291	0.496	4.23×10^{-7}
FABP3	(probe E113F)	0.829	0.400	0.429	3.14×10^{-5}
CRIP1	(probe P874R)	0.840	0.376	0.464	3.82×10^{-5}
WNT10B	(probe P993F)	0.792	0.407	0.385	1.44×10^{-3}

Genes with a methylation profile differentiating paediatric posterior fossa ependymomas from tumours located in the supratentorial region and spinal canal are shown. A methylation score (β value) of one represents complete methylation of a gene's CpG target site, while a score of zero represents no methylation of a CpG target site. The p-values shown are post Benjamini-Hochberg False Discovery Rate correction. Genes highlighted in red are upregulated signature genes of posterior fossa ependymomas established from gene expression array work (Taylor, Poppleton et al. 2005). PF = posterior fossa.

4.3.4.3 Differentially methylated genes between a posterior fossa sample subcluster and the remaining primary ependymoma cohort

Unsupervised hierarchical cluster analysis of the primary methylation cohort had previously identified a subgroup of nine ependymomas within a cluster of 34 posterior fossa tumours. These tumours were predominantly from young children and were of an anaplastic histology (Chapter 3, section 3.3.7). While differential gene methylation analysis did not identify disparity between this subgroup and the remaining posterior fossa tumours, differences in the methylation values of certain genes were found when the nine tumours were compared to the rest of the primary methylation cohort (Table 4.20). These included THYl (11q23.3), encoding the haemopoetic stem cell marker CD90 and the Wnt receptor gene FZD9 (7q11.23), which were both relatively hypomethylated across the smaller subgroup and therefore potentially overexpressed.

Table 4.20: Differentially methylated genes between the posterior fossa subgroup of nine tumours and the remainder of the primary ependymoma cohort.

Cluster hypomethylated genes	Cluster β value	Remaining cohort β value	β difference	p-value
HLA-DOA (probe P594F)	0.299	0.655	0.356	1.21×10^{-3}
THYl (probe P20R)	0.102	0.623	0.523	0.024
PRKCDBP (probe E206F)	0.242	0.581	0.339	0.024
FZD9 (probe E458F)	0.141	0.561	0.420	0.05
TNFRSF10D (probe E27F)	0.481	0.842	0.432	0.05

A methylation score (β value) of one represents complete methylation of a gene's CpG target site, while a score of zero represents no methylation of a CpG target site. The p-values shown are post Benjamini-Hochberg False Discovery Rate correction.

The cluster analysis had also identified a subgroup of five supratentorial tumours with a methylation profile distinct to that of the other supratentorial tumours in the cohort (Chapter 3, section 3.3.7). However, high resolution differential analysis failed to identify any significant differences in gene methylation values between this subgroup and either the remaining supratentorial tumours or the entire primary cohort.

4.3.5 Real time qPCR validation of SNP array results

Real time qPCR was performed to validate the copy number alterations found in particular genes of interest from the 500K SNP array analysis. Fifteen candidate genes were chosen (Table 4.21). All had either the potential to contribute to the pathogenesis of cancer based on their ontology, or had been associated with a variety of tumours, including ependymomas, from previous work. The focal regions of genomic imbalance incorporating these genes of interest were also aligned against a published map of normal copy number variation (CNV) across the human genome (Redon, Ishikawa et al. 2006) to detect sites of overlap. This identified only the region encompassing TXN (9q31.3) to be located in an area of known CNV. Nevertheless, this gene was not excluded from further analysis since evidence suggests that CNV regions may facilitate tumourigenesis (Johnson, Wright et al. 2010).

Table 4.21: Fifteen candidate genes selected for real time qPCR validation.

Gene	Locus	Reason for selection from SNP array analysis	Primer efficiency (Tm)
NSL1	1q32.3	Frequent gain in the primary intracranial cohort	91 % (58°C)
DNAJC25	9q31.3	Frequent gain / amplification in the primary intracranial cohort	99 % (58°C)
SEH1L	18p11.21	Frequent gain in the primary intracranial cohort	105 % (58°C)
FILIP1	6q14.1	Frequent loss in the primary intracranial cohort	94 % (57°C)
FRK	6q22.1	Frequent loss in the primary intracranial cohort	93 % (57°C)
HOXA5	7p15.2	Frequent amplification in the primary intracranial cohort	105 % (58°C)
NAV1	1q32.1	Frequent gain in the intracranial recurrent cohort and gain associated with adverse event-free survival	95 % (58°C)
PPARA	22q13.31	Frequent loss in the intracranial recurrent cohort	91 % (58°C)
CDKN2A	9p21.3	Homozygous deletion in the intracranial recurrent ependymoma	100 % (58°C)
CHI3L1	1q32.1	Frequent gain in the intracranial recurrent cohort and gain associated with posterior fossa ependymomas	100 % (58°C)
COL4A1	13q34	Gain associated with supratentorial (high grade) ependymomas	100 % (58°C)
TELO2	16p13.3	Gain associated with posterior fossa ependymomas from children aged below three years	99 % (58°C)
TXN	9q31.3	Gain associated with posterior fossa ependymomas from children aged below three years	104 % (58°C)
BNIPL	1q21.3	Gain associated with adverse overall survival	102 % (58°C)
PRUNE	1q21.3	Gain associated with adverse overall survival	100 % (58°C)
AJAP1	1p36.32	Control gene	93 % (57°C) 100 % (58°C)
ULK4	3p22.1	Control gene	95 % (at 57°C and 58°C)

The fifteen candidate genes selected are shown, along with the two control genes used in the real time PCR experiments, AJAP1 and ULK4. The reason for selecting each candidate gene on the basis of the 500K SNP array results is also given. In addition, the corresponding primer efficiency used for the qPCR of each gene and the optimal annealing temperature (Tm) to attain this percentage is included. Primer efficiency was calculated as described in Chapter 2, section 2.4.2, using the designed primer pairs shown in Chapter 2, Table 2.2.

4.3.5.1 qPCR validation of NSL1 gain

Real time qPCR was performed on nine primary ependymomas for NSL1 copy number validation (Figure 4.8 and Appendix 10I). Copy number gain of NSL1 from the SNP array analysis was verified in 5/6 (83 %) expected cases by qPCR. A normal copy number for NSL1 was verified in 3/3 (100 %) control cases. A positive correlation existed between SNP array and qPCR derived NSL1 copy number values (Spearman's rank correlation coefficient = 0.841; p = 0.005, two-tailed).

4.3.5.2 qPCR validation of DNAJC25 gain

Real time qPCR was performed on 11 primary ependymomas for DNAJC25 copy number validation (Figure 4.9 and Appendix 10I). Copy number gain of DNAJC25 from

the SNP array analysis was verified in 7/8 (88 %) expected cases by qPCR. A normal copy number for DNAJC25 was verified in 2/3 (67 %) control cases. A positive correlation existed between SNP array and qPCR derived DNAJC25 copy number values (Spearman's rank correlation coefficient = 0.735; p = 0.01, two-tailed).

4.3.5.3 qPCR validation of FILIP1 loss

Real time qPCR was performed on seven primary ependymomas for FILIP1 copy number validation (Figure 4.10 and Appendix 10I). Copy number loss of FILIP1 from the SNP array analysis was verified in 4/4 (100 %) expected cases by qPCR. A normal copy number for FILIP1 was verified in 2/3 (67 %) control cases. A positive correlation existed between SNP array and qPCR derived FILIP1 copy number values (Spearman's rank correlation coefficient = 0.874; p = 0.01, two-tailed).

4.3.5.4 qPCR validation of FRK loss

Real time qPCR was performed on seven primary ependymomas for FRK copy number validation (Figure 4.11 and Appendix 10I). Copy number loss of FRK from the SNP array analysis was verified in 4/4 (100 %) expected cases by qPCR. A normal copy number for FRK was verified in 2/3 (67 %) control cases. A positive correlation existed between SNP array and qPCR derived FRK copy number values (Spearman's rank correlation coefficient = 0.874; p = 0.01, two-tailed).

4.3.5.5 qPCR validation of HOXA5 gain/amplification

Real time qPCR was performed on 14 primary ependymomas for HOXA5 copy number validation (Figure 4.12 and Appendix 10I). Amplification of HOXA5 (copy number greater than five) from the SNP array was verified in 3/4 (75 %) expected cases by qPCR. Two other tumours (21P and 34P) revealed lower level gain of HOXA5 from the SNP array analysis (both copy number three). However, both of these tumours exhibited a higher level of genomic gain when assessed by qPCR (copy numbers 6.1 (sample 21P) and 4.5 (sample 34P)). A normal copy number for HOXA5 was verified in 3/8 (38 %) control cases, with the others revealing copy number gain. A positive correlation existed

between SNP array and qPCR derived HOXA5 copy number values (Spearman's rank correlation coefficient = 0.694; p = 0.006, two-tailed).

4.3.5.6 qPCR validation of NAV1 gain

Real time qPCR was performed on 12 ependymomas for NAV1 copy number validation (Figure 4.13 and Appendix 10I). Copy number gain of NAV1 from the SNP array analysis was verified in 4/4 (100 %) expected intracranial first recurrent cases and 3/3 (100 %) expected primary cases by qPCR. A normal copy number for NAV1 was verified in 3/5 (60 %) control cases. A positive correlation existed between SNP array and qPCR derived NAV1 copy number values (Spearman's rank correlation coefficient = 0.748; p = 0.003, two-tailed).

4.3.5.7 qPCR validation of CDKN2A homozygous deletion

Real time qPCR was performed on five ependymomas for CDKN2A copy number validation (Figure 4.14 and Appendix 10I). The SNP array analysis identified one intracranial recurrent ependymoma (sample 18R1) to have homozygous genomic deletion of CDKN2A. This was verified by qPCR. Of the control tumours analysed, copy number gain was verified in one expected case, while a normal copy number for CDKN2A was verified in 3/3 (100 %) additional cases. A positive correlation existed between SNP array and qPCR derived CDKN2A copy number values (Spearman's rank correlation coefficient = 0.918; p = 0.03, two-tailed).

4.3.5.8 qPCR validation of CHI3L1 gain

Real time qPCR was performed on nine primary ependymomas for CHI3L1 copy number validation (Figure 4.15 and Appendix 10I). Copy number gain of CHI3L1 from the SNP array analysis was verified in 5/5 (100 %) expected cases by qPCR. A normal copy number for CHI3L1 was verified in 4/4 (100 %) control cases. A positive correlation existed between SNP array and qPCR derived CHI3L1 copy number values (Spearman's rank correlation coefficient = 0.895; p = 0.001, two-tailed).

4.3.5.9 qPCR validation of TXN gain

Real time qPCR was performed on eight primary ependymomas for TXN copy number validation (Figure 4.16 and Appendix 10I). Copy number gain of TXN from the SNP array analysis was verified in 4/4 (100 %) expected cases by qPCR. A normal copy number for TXN was verified in 2/4 (50 %) control cases. A positive correlation existed between SNP array and qPCR derived TXN copy number values (Spearman's rank correlation coefficient = 0.75; p = 0.035, two-tailed).

4.3.5.10 qPCR validation of BNIPL gain

Real time qPCR was performed on nine primary ependymomas for BNIPL copy number validation (Figure 4.17 and Appendix 10I). Copy number gain of BNIPL from the SNP array analysis was verified in 5/5 (100 %) expected cases by qPCR. A normal copy number for BNIPL was verified in 4/4 (100 %) control cases. A positive correlation existed between SNP array and qPCR derived BNIPL copy number values (Spearman's rank correlation coefficient = 0.921; p = 0.001, two-tailed).

4.3.5.11 qPCR validation of PRUNE gain

Real time qPCR was performed on nine primary ependymomas for PRUNE copy number validation (Figure 4.18 and Appendix 10I). Copy number gain of PRUNE from the SNP array analysis was verified in 5/5 (100 %) expected cases by qPCR. A normal copy number for PRUNE was verified in 4/4 (100 %) control cases. A positive correlation existed between SNP array and qPCR derived PRUNE copy number values (Spearman's rank correlation coefficient = 0.935; p = 0.001, two-tailed).

4.3.5.12 qPCR of SEH1L gain, PPARA loss, TELO2 gain and COL4A1 loss

Real time qPCR did not validate the copy number alterations identified in SEH1L, PPARA, TELO2 and COL4A1 from the SNP array study (summarised in Table 4.22 with specific results in Appendix 10I).

Figure 4.8: Real time qPCR validation of NSL1 gene copy number previously ascertained on the 500K SNP array. qPCR was performed at 58°C with ULK4 as the control gene. P = primary.

Figure 4.9: Real time qPCR validation of DNAJC25 gene copy number previously ascertained on the 500K SNP array. qPCR was performed at 58°C with ULK4 as the control gene. P = primary.

186

Figure 4.10: Real time qPCR validation of FILIP1 gene copy number previously ascertained on the 500K SNP array. qPCR was performed at 57°C with AJAP1 as the control gene. P = primary.

Figure 4.11: Real time qPCR validation of FRK gene copy number previously ascertained on the 500K SNP array. qPCR was performed at 57°C with AJAP1 as the control gene. P = primary.

Figure 4.12: Real time qPCR validation of HOXA5 gene copy number previously ascertained on the 500K SNP array. qPCR was performed at 58°C with ULK4 as the control gene. P = primary.
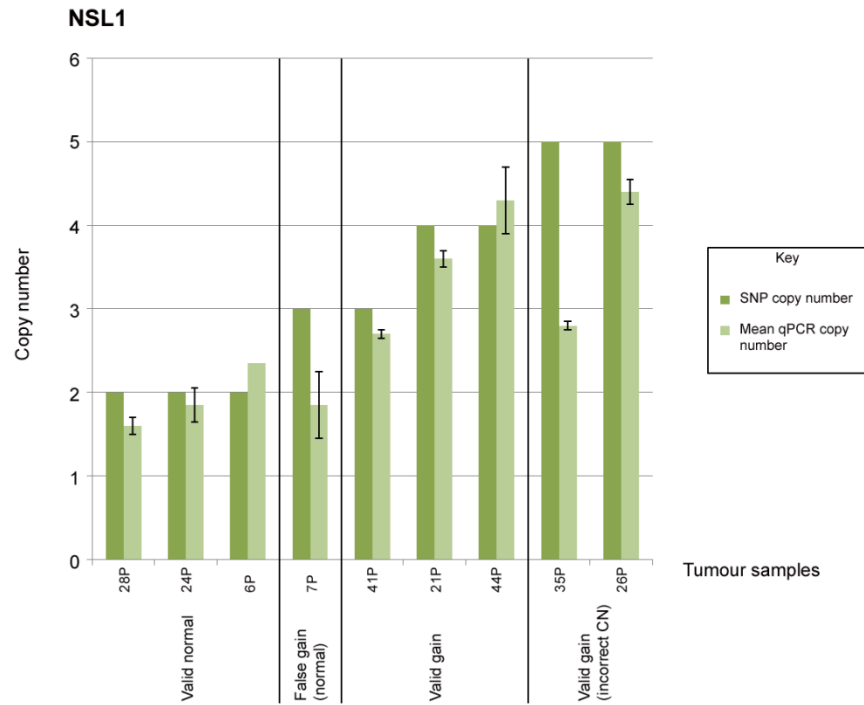

Figure 4.13: Real time qPCR validation of NAV1 gene copy number previously ascertained on the 500K SNP array. qPCR was performed at 58°C with ULK4 as the control gene. P = primary, R1 = 1st recurrence.

Figure 4.14: Real time qPCR validation of CDKN2A gene copy number previously ascertained on the 500K SNP array. qPCR was performed at 58°C with ULK4 as the control gene. P = primary, R1 = recurrence.

Figure 4.15: Real time qPCR validation of CHI3L1 gene copy number previously ascertained on the 500K SNP array. qPCR was performed at 58°C with ULK4 as the control gene. P = primary.

189
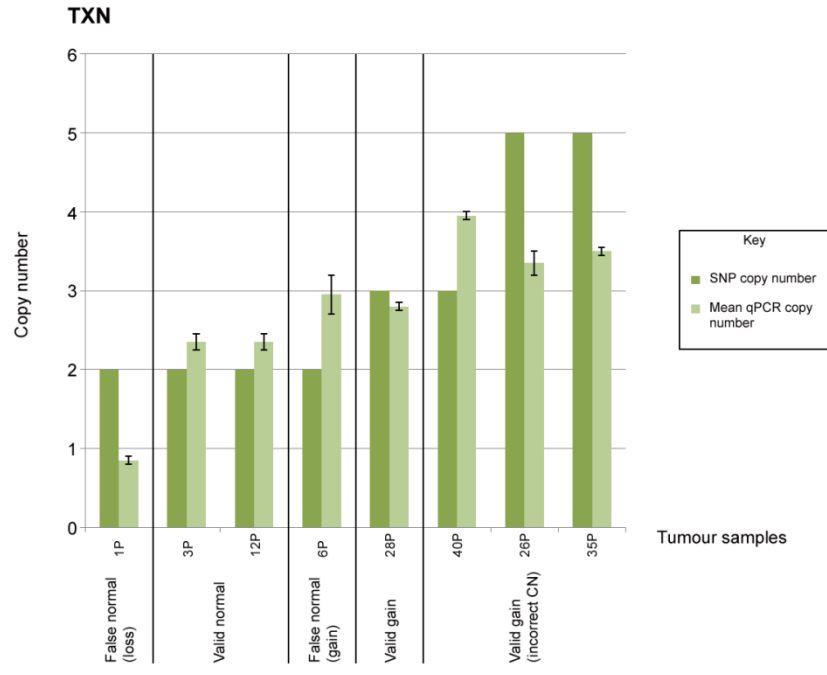
Figure 4.16: Real time qPCR validation of TXN gene copy number previously ascertained on the 500K SNP array. qPCR was performed at 58°C with ULK4 as the control gene. P = primary.
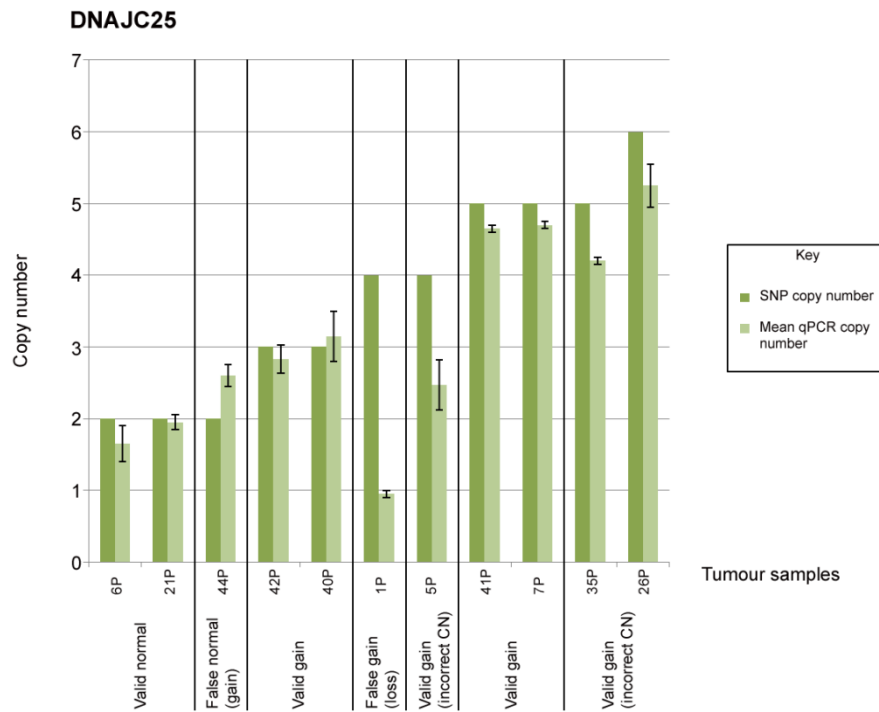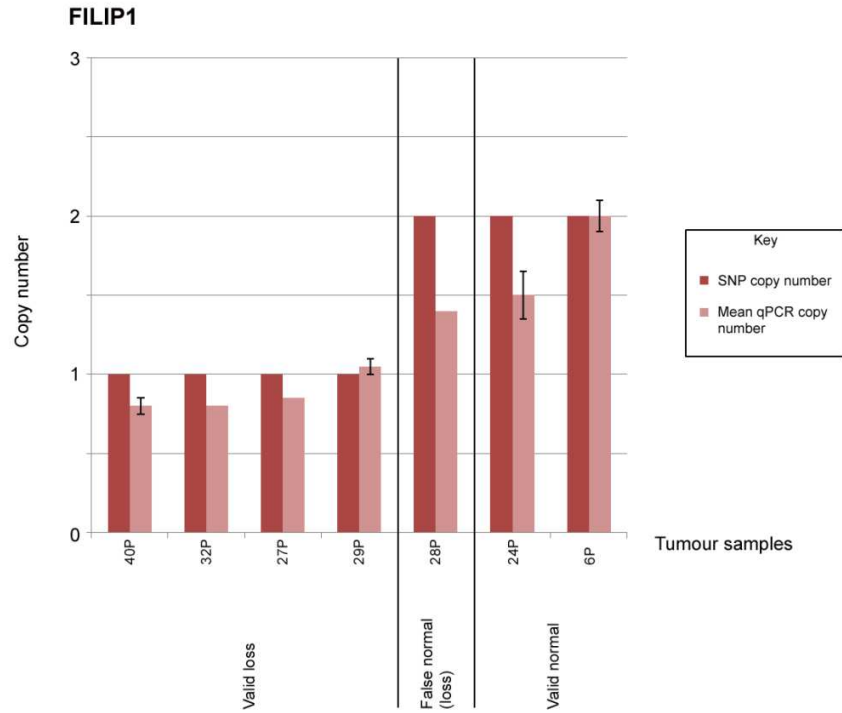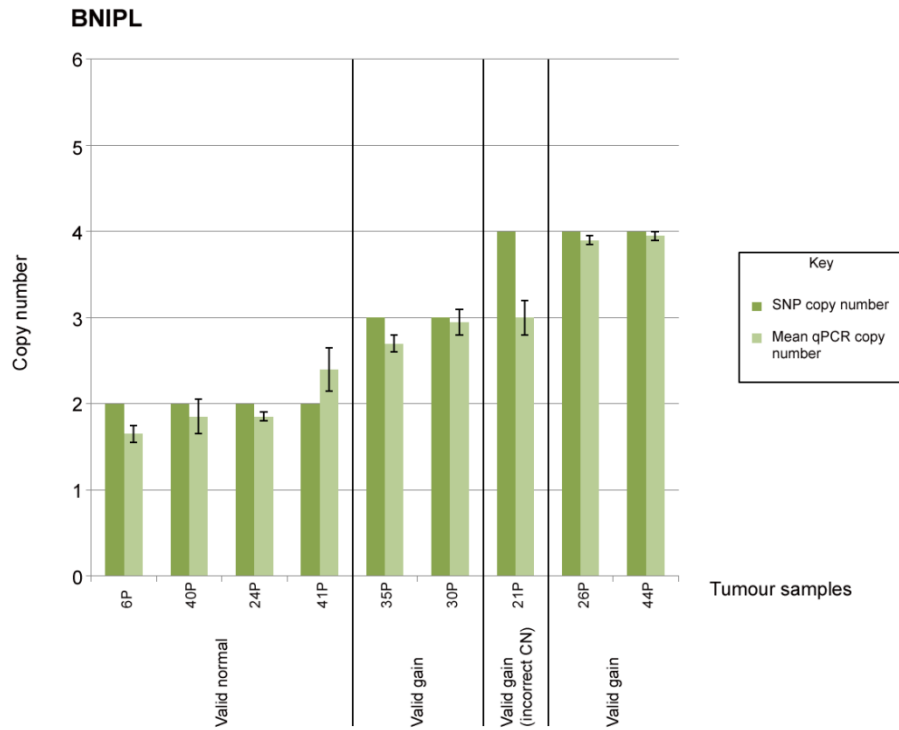

Figure 4.17: Real time qPCR validation of BNIPL gene copy number previously ascertained on the 500K SNP array. qPCR was performed at 58°C with ULK4 as the control gene. P = primary.
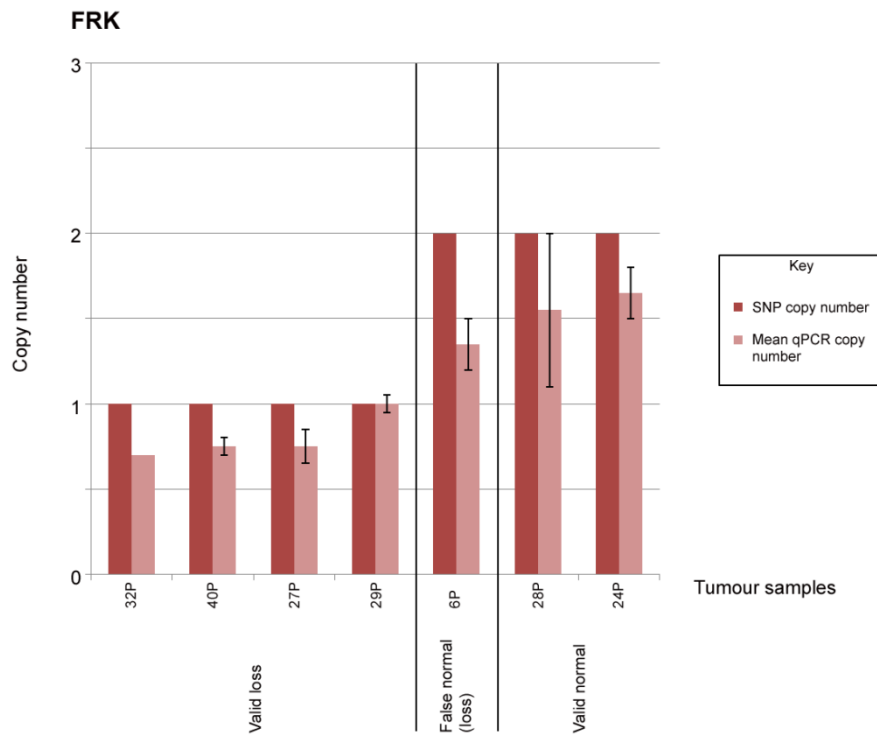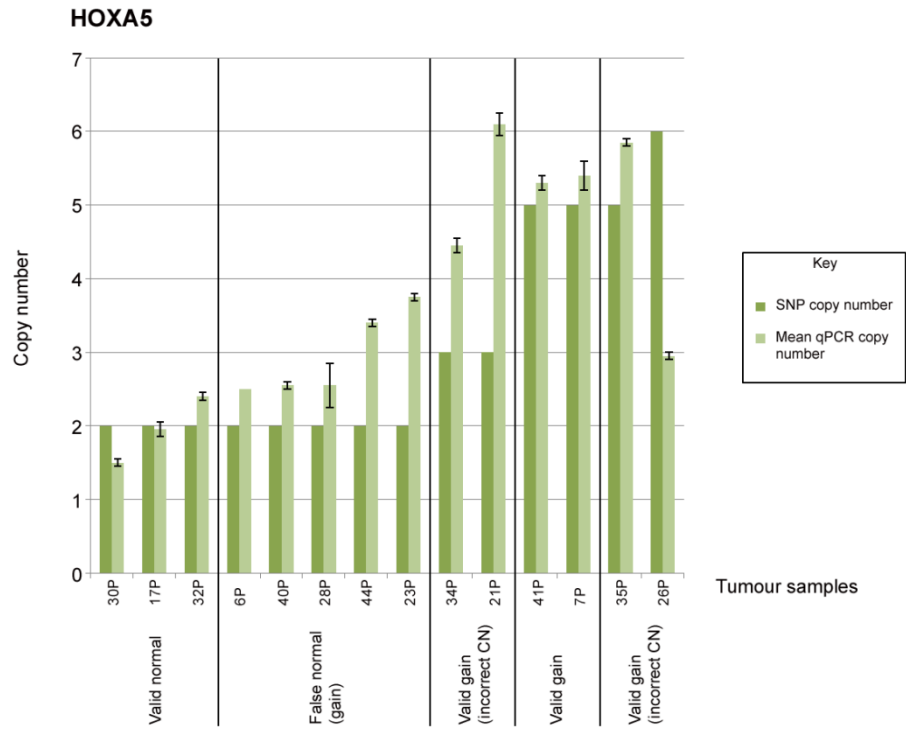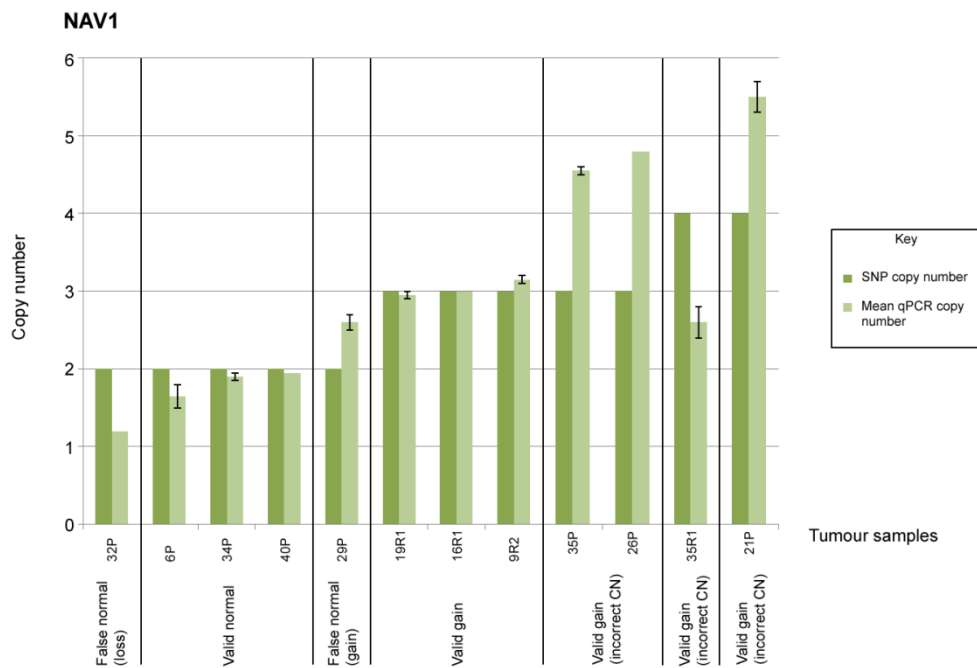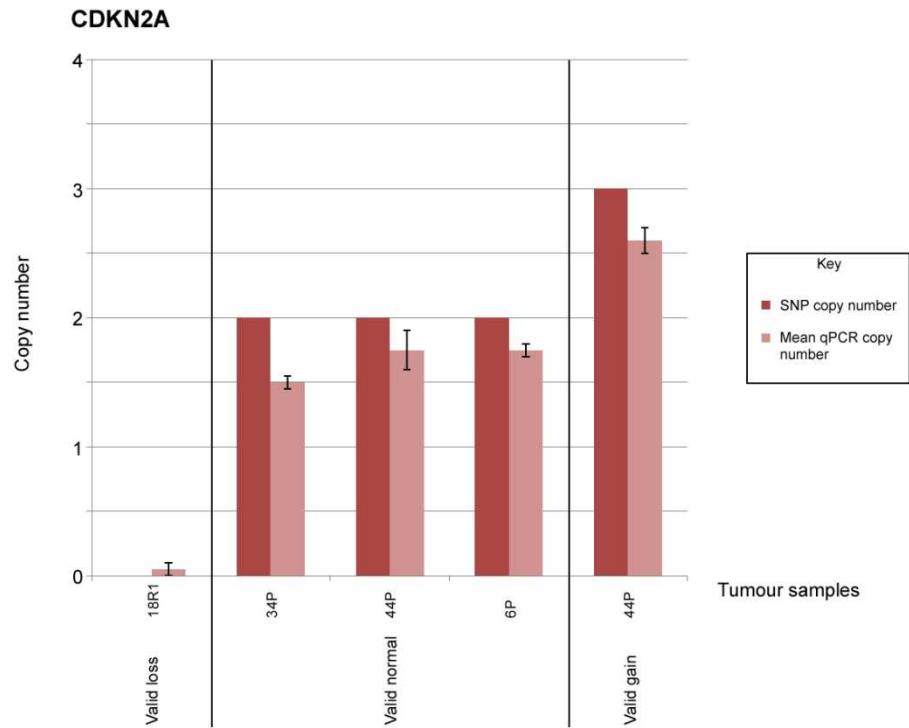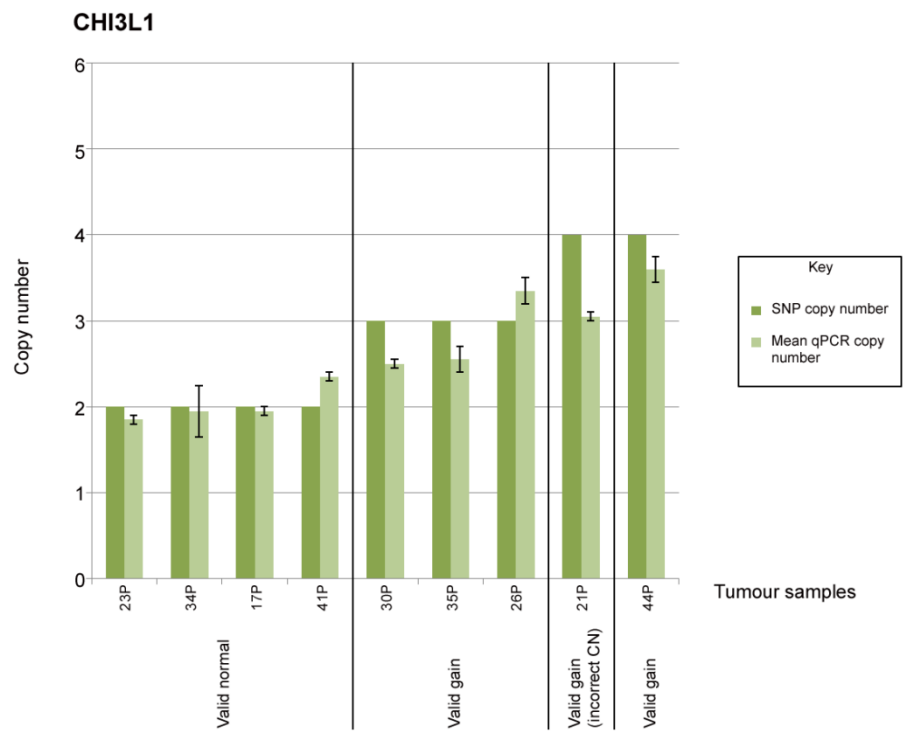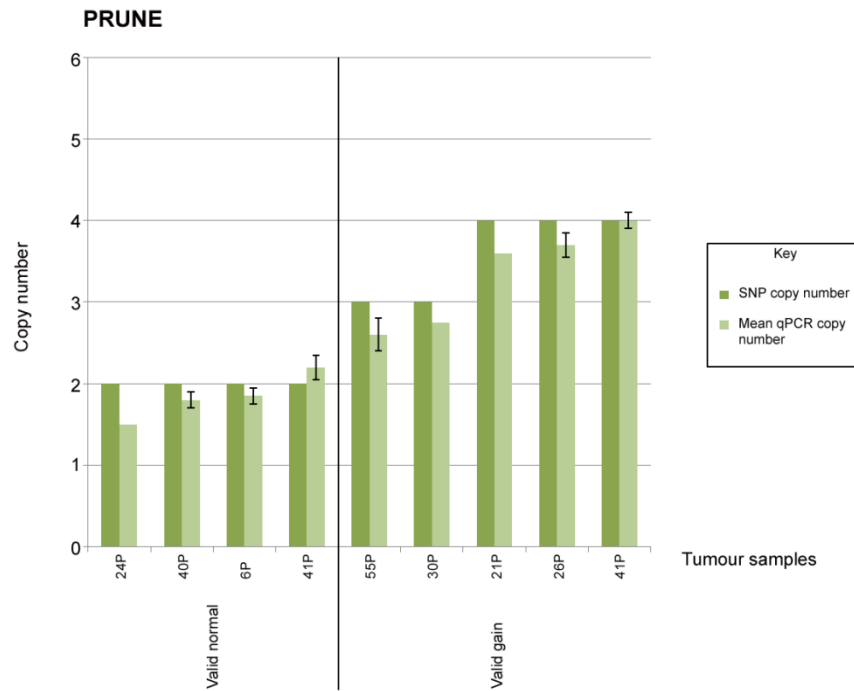
Figure 4.18: Real time qPCR validation of PRUNE gene copy number previously ascertained on the 500K SNP array. qPCR was performed at 58°C with ULK4 as the control gene. P = primary.

Table 4.22: Summary of real time qPCR validation results for SEH1L gain, PPARA

loss, TELO2 gain and COL4A1 loss.

Gene	SNP array derived copy number in ependymomas selected for validation	qPCR derived copy number of the selected ependymomas	Spearman's rank correlation coefficient (two-tailed)
SEH1L	Gain in 6 tumours	Gain = 2 tumours Normal = 4 tumours (Validation = 33 %)	R = 0.409 p = 0.363
	Normal in 1 tumour	Loss = 1 tumour (Validation 0 %)	
PPARA	Loss in 3 tumours	Normal = 1 tumour Gain = 2 tumours (Validation = 0 %)	R = -0.434 p = 0.454
	Normal in 2 tumours	Normal = 2 tumours (Validation = 100 %)	
TELO2	Gain in 3 tumours	Normal = 2 tumours Loss = 1 tumour Validation = 0 %)	Not feasible
COL4A1	Gain in 2 tumours	Normal = 2 tumours (Validation = 0 %)	Not feasible

Since real time qPCR had failed to validate TELO2 gain and COL4A1 loss in ependymomas identified from the 500K SNP array analysis, other tumours in the cohort with apparently normal SNP derived copy numbers for these two genes were not subjected to qPCR in order to avoid sample depletion. As a consequence, the Spearman's rank correlation could not be calculated as indicated in the table. R= Spearman's rank correlation coefficient with two-tailed significance value (p) underneath where calculated. Specific results are shown in Appendix 10I.

4.4 Discussion

4.4.1 Genes potentially implicated in paediatric ependymoma pathogenesis

The high resolution SNP and methylation array work presented in this chapter identified a number of genes potentially implicated in paediatric ependymoma pathogenesis through genomic imbalance, epigenetic alteration or rarely, acquired uniparental disomy. While it is important to consider all genes within a designated region of interest, certain genes identified from this work are now discussed. These include candidates associated with ependymoma pathogenesis from previous studies, genes which exhibited complementary methylation and expression profiles in ependymomas from different CNS locations and, in particular, fifteen SNP array derived candidates selected for copy number validation by qPCR (Table 4.21). The latter group were chosen because of detected aberrations in their copy number which were either frequent across primary and recurrent intracranial cohorts, or associated with particular clinical or prognostic subgroups. All fifteen genes had been implicated in various human cancers from preceding work, or had a cellular function that could potentially contribute to tumour pathogenesis if dysregulated.

4.4.1.1 Candidate genes demonstrating frequent genomic imbalance in primary and recurrent intracranial ependymomas

From the 500K SNP array analysis, NSL1 (1q32.3), SEH1L (18p11.21) and DNAJC25 (9q31.3) were among the most frequently gained genes in the primary intracranial ependymoma cohort (Table 4.1). Both NSL1 and SEH1L promote cell division by encoding subunits of protein complexes responsible for the formation and function of kinetochores, chromosome-associated structures that mediate chromosomal segregation during mitosis (Kline, Cheeseman et al. 2006; Zuccolo, Alves et al. 2007). DNAJC25, one of the most frequently amplified genes in the cohort, encodes a member of the DNAJ/HSP40 group of heat shock proteins that are emerging with roles in various cancers (Mitra, Shevde et al. 2009). The DNAJ/HSP40 family are thought to act as molecular chaperones, important for optimising protein structure and function. They also regulate the activity of other heat shock chaperones, such as HSP70 proteins which

have been implicated in cancer progression (Didelot, Lanneau et al. 2007). Although research interest in a tumourigenic role for DNAJ/HSP40 proteins has recently increased, only 8/41 family members have currently been assessed (Mitra, Shevde et al. 2009). Whereas some appear associated with tumour suppression, overexpression of DNAJA3 has been correlated with colorectal cancer progression (Kurzik-Dumke, Horner et al. 2008), while DNAJA1 reportedly contributes to the in vitro radiotherapy resistance of glioblastomas (Wang, Liao et al. 2006). This ependymoma study is the first to propose a potential oncogenic role for DNAJC25.

The most common regions of focal genomic loss in the primary intracranial cohort were detected primarily on chromosome 6q (Table 4.2). This supported findings from the paediatric ependymoma CGH meta-analysis which identified chromosome 6 deletion as a frequent genomic alteration in primary posterior fossa tumours (Chapter 1, section 1.5.2.2) (Kilday, Rahman et al. 2009) and an array CGH study which associated 6q loss with intracranial ependymomas (Mendrzyk, Korshunov et al. 2006). Putative tumour suppressor genes were encompassed within these focal regions of deletion, including FILIP1 (6q14.1) and FRK/RAK (6q22.1). FILIP1 encodes a filamin interacting protein whose in vitro degradation has been observed in oesophageal carcinoma cells (Shimada, Shiratori et al. 2009). In the CNS, the FILIP1 protein reportedly inhibits neocortical cell migration out of the ventricular zone (Nagano, Yoneda et al. 2002). FRK/RAK encodes a member of the TYR tyrosine kinase family (Lee, Wang et al. 1994). It has been proposed as a tumour suppressor in breast carcinoma (Yim, Peng et al. 2009), potentially through a mechanism promoting the stability and function of PTEN, a recognised tumour suppressor gene in ependymoma (Taylor, Poppleton et al. 2005; Johnson, Wright et al. 2010).

The most common regions of focal copy number gain and loss in the intracranial recurrent cohort were found on chromosomes 1q and 22q respectively (Tables 4.3 and 4.4), a feature observed and discussed further in Chapter 5. Frequent focal regions of gain encompassed genes implicated in a variety of processes, including cell migration and invasion, neurite growth and the inhibition of apoptosis. In addition, genes facilitating mitotic activity by encoding kinetochore associated proteins were also evident. The most frequently gained gene was NAV1 (1q32.1), which encodes a member of the neuron navigator family expressed predominantly in the central nervous system

(Maes, Barcelo et al. 2002). Gain of NAV1 was also evident in the primary ependymoma cohort where it was associated with an adverse event-free patient survival. NAV1 is homologous to the unc-53 gene responsible for axonal guidance in the roundworm Caenorhabditis elegans (Maes, Barcelo et al. 2002). Indeed, recent work has established that mammalian NAV1 localises to the centrosome, associates with the growing ends of microtubules and is potentially involved in neurite outgrowth (van Haren, Draegestein et al. 2009). Other gained genes such as LAMC1 (1q25.3), CENPF (1q41) and CDC42BPA (1q42.13) are reviewed in the fifth chapter.

CBX7 (22q13.1) and PPARA (22q13.31) were encompassed within the regions most frequently demonstrating hemizygous loss. CBX7 encodes a member of the chromobox protein family implicated in gene silencing by facilitating the structural alteration of chromatin (Paro and Hogness 1991). It has already been proposed as a candidate tumour suppressor gene following expression array analysis of paediatric ependymomas (Suarez-Merino, Hubank et al. 2005). The peroxisome proliferator receptor encoded by PPARA activates the transcription of proteins responsible for fatty acid oxidation. It is proposed to exert a negative impact on tumours by inhibiting glycolysis and promoting fatty acid catabolism, thereby disrupting the metabolic pathways thought to be preferred by cancer cells in a state of nutritional deprivation (Grabacka and Reiss 2008). In neoplastic cells exposed to such stress, PPARA is also thought to interact with AMPK (AMP-dependent kinase) to induce cell cycle arrest, promote apoptosis, block Akt oncogenicity, inhibit cell proliferation, reduce cell migration and prevent inflammation. This tumour suppressive activity has been reported in several human cancers, including melanoma, (Grabacka, Plonka et al. 2006), colon cancer (Niho, Takahashi et al. 2003) and medullobalstoma (Urbanska, Pannizzo et al. 2008).

Only one gene region of homozygous deletion was identified across the entire SNP array cohort of ependymomas. This was located in a recurrent supratentorial tumour on chromosome 9p21.3, encompassing the tumour suppressor genes CDKN2A and CDKN2B. Both of these cyclin-dependent kinase inhibitors regulate cell cycle G1 progression, while CDKN2A has also been shown to control neural stem cell proliferation and rapidly expand progenitor cell numbers in developing neural tissue through its depletion (Taylor, Poppleton et al. 2005). Homozygous deletions of both genes are established anomalies in paediatric brain tumours such as CNS PNETs

(Pfister, Remke et al. 2007) and ependymomas, particularly those from a supratentorial location (Taylor, Poppleton et al. 2005; Johnson, Wright et al. 2010) as was the case in this study. Indeed, biallelic CDKN2A loss has been reported as an independent marker of adverse patient survival in ependymoma (Korshunov, Witt et al. 2010). Epigenetic inactivation of both CDKN2A and CDKN2B through promoter methylation has also been reported in ependymomas (Chapter 1, Table 1.7).

4.4.1.2 Candidate genes demonstrating genomic or epigenetic abnormalities associated with ependymomas from particular CNS locations

As the broad regions of genomic imbalance exhibited by spinal ependymomas from the SNP array study encompassed numerous genes, the identification of specific candidates was not feasible and therefore not performed. Nevertheless, the methylation array analysis identified several genes differentially methylated between spinal and intracranial ependymomas (Table 4.17). One such gene, EYA4 (6q23), demonstrated relative hypomethylation in spinal tumours. EYA4 encodes a member of the eyes absent transcription factor family which function through protein phosphatase activity (Tootle, Silver et al. 2003). Overexpression of this gene has been shown to be characteristic of spinal ependymomas when compared to intracranial tumours (Taylor, Poppleton et al. 2005), implying that the expression profile of EYA4 in ependymomas could be influenced by its methylation status. This is supported by integrated epigenetic and expression analyses of oesophageal adenocarcinoma (Zou, Osborn et al. 2005) and non-Hodgkin's lymphoma (Alzein 2005), where aberrant EYA4 promoter methylation has been identified as a common mechanism of altered gene expression and gene silencing.

The methylation array analysis also identified relative hypermethylation of RASSF1A (3p21.3) within the intracranial ependymoma cohort. This tumour suppressor gene encodes a protein with homology to the RAS effector proteins (Dammann, Schagdarsurengin et al. 2003). It has been shown to inhibit cell cycle progression in vitro (Shivakumar, Minna et al. 2002) and appears to be required for death receptor dependent apoptosis (Oh, Lee et al. 2006), while RASSF1A knockout mice have been shown to be susceptible to tumourigenesis (Tommasi, Dammann et al. 2005; Tommasi, Besaratinia et al. 2011). Promoter hypermethylation has been reported as a mechanism

of RASSF1A silencing in several malignancies (reviewed by (Dammann, Schagdarsurengin et al. 2003)). With regards to paediatric CNS tumours, biallelic hypermethylation has been reported as the primary mechanism of RASSF1A inactivation in medulloblastomas (Lusher, Lindsey et al. 2002), while previous reports identifying RASSF1A hypermethylation as a frequent epigenetic aberration in childhood intracranial ependymomas (Hamilton, Lusher et al. 2005; Michalowski, de Fraipont et al. 2006) (Chapter 1, Table 1.7) reinforce the findings presented in this work, warranting further assessment of this gene as a candidate tumour suppressor in paediatric ependymomas from this location.

From the 500K SNP array analysis, HOXA5 (7p15.2) and certain other homeobox (HOX) family members were among the most frequently amplified genes within the primary intracranial ependymoma cohort (Table 4.5A). This gene family are responsible for regulating anteroposterior tissue patterning during embryological development (Dasen, Liu et al. 2003; Kmita and Duboule 2003; Taylor, Poppleton et al. 2005). In contrast to this SNP array finding, gene expression analyses have demonstrated the upregulation of several HOX genes, including HOXA5, in spinal ependymomas as opposed to intracranial counterparts (Korshunov, Neben et al. 2003; Taylor, Poppleton et al. 2005; Modena, Lualdi et al. 2006; Lukashova-v Zangen, Kneitz et al. 2007; Palm, Figarella-Branger et al. 2009). While this suggests the transcriptional repression of HOX genes may occur despite genomic amplification in intracranial ependymomas, the precise putative repressive mechanism of action remains unclear and requires evaluation. Gene promoter methylation appears an unlikely explanation from this work since differential analysis between spinal and intracranial tumours did not establish a difference in the methylation status of any HOX genes, including HOXA5.

Within the paediatric intracranial ependymoma cohort, the SNP array analysis revealed supratentorial tumours were associated with the genomic gain of TULP4 (6q25.3), a gene involved in protein ubiquitination which has previously been proposed as an overexpressed signature gene of supratentorial ependymomas (Taylor, Poppleton et al. 2005), and COL4A1 (13q34) which encodes the type IV alpha collagen chain of cellular basement membranes (Soininen, Haka-Risku et al. 1987) (Table 4.7). Exclusive analysis of the supratentorial ependymoma group also discovered COL4A1 gain was

associated with tumours of an anaplastic histology. Array work has reported the overexpression of COL4A1 in glial derived brain tumours such as paediatric intracranial ependymomas (Suarez-Merino, Hubank et al. 2005) and high grade gliomas (van den Boom, Wolter et al. 2003). Indeed, in vitro studies have demonstrated that normal brain tissue can produce and overexpress type IV collagen in response to confrontation from invading glioma cells (van den Boom, Wolter et al. 2003). Immunohistochemical analysis of gliomas has also revealed type IV collagen expression may be associated with regions of neoplastic vascular proliferation (Bellon, Caulet et al. 1985). These findings may suggest why expression of COL4A1 is increased in higher grade tumours as they often have a propensity for increased infiltration, invasiveness and angiogenic requirements. Nevertheless, explanations for the increased prevalence of SNP-array derived COL4A1 gain in supratentorial ependymomas remain unclear. Further investigation is required to elucidate whether this is a potential consequence of the migration and invasion of supratentorial tumours away from ependymal regions, or simply a reflection of the different biological microenvironments of the CNS compartments.

The methylation array analysis identified 15 genes exhibiting relative hypomethylation in posterior fossa ependymomas compared to supratentorial tumours (Table 4.19). These included the metalloproteinase gene MMP2 (16q13) which encodes a putative marker of microinvasion and prognosis in ependymoma (Snuderl, Chi et al. 2008) (Chapter 1, Table 1.8), and three other genes, PPARG (3p25), SPDEF (6p21.3) and BCR (22q11.23), all designated signature genes for posterior fossa ependymomas because of their characteristic overexpression when compared to ependymomas from other CNS locations (Taylor, Poppleton et al. 2005). As with EYA4 expression in spinal ependymomas, this suggests that the expression profile of these three genes may be influenced by corresponding gene promoter methylation status, a hypothesis requiring validation. PPARG encodes another peroxisome proliferator receptor which controls fatty acid oxidation and glucose homeostasis (Picard and Auwerx 2002; Jones, Barrick et al. 2005). It has been reported to have tumour suppressive roles in colon cancer (Zou, Qiao et al. 2009) and glioblastoma (Cimini, Cristiano et al. 2005), yet oncogenic properties in breast carcinoma (Kourtidis, Srinivasaiah et al. 2009). SPDEF encodes a transcription factor, overexpression of which in vitro has been shown, in conjunction with ERBB2, to promote cell motility and invasiveness in breast cancer (Gunawardane,

Sgroi et al. 2005). While the precise function of BCR is unclear, it can act as a GTPase activating protein of CDC42 (Chuang, Xu et al. 1995), thereby potentially having a role in processes such as neurite growth, migration and cell cycle control. It is also a component of the BCR-ABL fusion gene found in chronic myeloid leukaemia (Grosveld, Verwoerd et al. 1986).

Several genes demonstrating genomic gain on chromosome 1q were associated with primary posterior fossa ependymomas when compared to supratentorial counterparts from the SNP array analysis (Table 4.6). One such candidate was CHI3L1 (previously YKL40, 1q32.1). Gain of this gene was also a relatively common occurrence in the intracranial first recurrent cohort, being identified in 5/9 (55 %) tumours. Of these five recurrent cases, four (80 %) were located in the posterior fossa. CHI3L1 belongs to the chitinase gene family which function to hydrolyse the glycopolymer chitin (Funkhouser and Aronson 2007). Previous gene expression work has demonstrated CHI3L1 upregulation in paediatric ependymomas when compared to normal brain and identified it as the most overexpressed gene at relapse in a patient-matched primary and recurrent tumour pair, again arising from the posterior fossa (Rand, Prebble et al. 2008). The encoded protein is thought to be implicated in cell migration, inflammation and connective tissue remodelling (Volck, Price et al. 1998; Boot, van Achterberg et al. 1999; De Ceuninck, Gaufillier et al. 2001), with overexpression reported as an adverse prognostic marker in glioblastoma (Johansen, Christensen et al. 2003; Pelloski, Mahajan et al. 2005) and malignancies of the breast (Johansen, Christensen et al. 2003), colon (Cintin, Johansen et al. 1999) and ovary (Hogdall, Johansen et al. 2003). However, analysis of CHI3L1 overexpression in paediatric ependymomas has revealed a correlation only with tumour necrosis and not patient outcome (Rand, Prebble et al. 2008). This agrees with the findings from the SNP array analysis which did not find a statistically significant association between CHI3L1 gain and either event-free or overall patient survival (Table 4.12). The SNP and methylation array analyses also identified abnormalities in other chitinase family members that were associated with posterior fossa ependymomas when analysed against tumours from other CNS locations, including the gain of CHIT1 (1q32.1) (Table 4.6) and the relative hypomethylation of CHI3L2 (1p13.3) (Table 4.19). This suggests a role for chitinase gene members in the pathogenesis of childhood posterior fossa ependymomas, warranting further investigation (Chapter 7).

Other genes demonstrating gain which were associated with posterior fossa ependymomas from the SNP array work included PARP1 (1q42.12) and AKT3 (1q44). While AKT3 encodes a protein kinase which regulates cell signalling and has been associated with maintaining the viability of glioma cells in vitro (Mure, Matsuzaki et al. 2010), PARP1 encodes a protein implicated in DNA damage repair and cell proliferation, overexpression of which has been reported in numerous malignancies including colorectal carcinoma and primary glioblastoma multiforme (Wharton, McNelis et al. 2000; Nosho, Yamamoto et al. 2006). In addition, the development of PARP inhibition as a specific therapy for certain malignancies such as BRCA1 and BRCA2 deficient breast cancer is ongoing (Farmer, McCabe et al. 2005; Brown, Power et al. 2008; Rottenberg, Jaspers et al. 2008; Kling 2009).

4.4.1.3 Candidate genes demonstrating genomic and epigenetic abnormalities associated with ependymomas from different paediatric patient age groups

Twenty-four primary posterior fossa ependymomas analysed on the 500K SNP array were categorised equally into two groups according to patient age (below and above three years of age). Comparative analysis of these two subsets enabled the identification of genes demonstrating genomic imbalance which were associated with a particular age group.

Gain of TELO2 (also known as CLK2, 16p13.3) was exclusive to patients aged below three years (Table 4.8). In the roundworm Caenorhabditis elegans, this gene regulates cellular processes including DNA replication, apoptosis and telomere length (Jiang, Benard et al. 2003). As stated in Chapter 1, section 1.5.7, telomeres cap chromosomal termini and erode after each mitotic division in normal somatic cells, whereas telomere maintenance is present in almost all malignant cells (Shay and Bacchetti 1997). As in the roundworm, in vitro overexpression of human TELO2 has also resulted in telomere maintenance and lengthening (Jiang, Benard et al. 2003), suggesting potential oncogenic properties for this gene. Gains incorporating TMEM18G (2p25.3) and HDAC10 (22q13.3) were also associated with the younger age group. While upregulation of TMEM18G may be involved in neural stem cell migration (Jurvansuu,

Zhao et al. 2008), HDAC10 encodes one of the histone deacetylases (HDACs) thought to contribute to tumourigenesis through the epigenetic alteration of chromatin structure and transcriptional repression of target genes (Rahman, Osteso-Ibanez et al. 2010). Indeed, the in vitro inhibition of HDACs in paediatric brain tumour cells has been shown to inhibit proliferation, diminish telomerase activity and promote apoptosis (Rahman, Osteso-Ibanez et al. 2010), while clinical trial evaluation of HDAC inhibitors as anti-cancer therapies are emerging for a variety of malignancies, including brain tumours (Galanis, Jaeckle et al. 2009; Rahman, Osteso-Ibanez et al. 2010).

In contrast, gain of TXN (9q31.3) was exclusive to posterior fossa tumours from children aged over three years (Table 4.9). TXN encodes a member of the thioredoxin family, an oxidoreductase enzyme group thought to protect cancer cells against stress-induced dysregulated oxidation and reduction processes (Powis and Kirkpatrick 2007). High protein expression of TXN has been associated with increased cancer cell proliferation, tumour angiogenesis and decreased apoptosis and has been reported in several malignancies such as lung, gastric and colorectal cancer (Grogan, Fenoglio-Prieser et al. 2000; Kakolyris, Giatromanolaki et al. 2001; Raffel, Bhattacharyya et al. 2003; Powis and Kirkpatrick 2007). Indeed, TXN overexpression in colorectal cancer has been reported as an independent negative marker of patient prognosis (Raffel, Bhattacharyya et al. 2003). The role of this gene in ependymomas from older children therefore merits further assessment at an expression and subsequent functional level if feasible, particularly as drugs inhibiting TXN and the thioredoxin redox system are now in clinical development (Mehta, Rodrigus et al. 2003; Ramanathan, Kirkpatrick et al. 2007).

Differential methylation analysis did not establish genes with a significantly different methylation value between primary ependymomas from young and older children or WHO grade II or III tumours. Nevertheless, unsupervised hierarchical clustering of the primary methylation array cohort did identify a subgroup of nine posterior fossa ependymomas from children aged less than three years which were predominantly of an anaplastic histology (Chapter 3, section 3.3.7). This small set of tumours were characterised by the relative hypomethylation of five genes compared to the rest of the primary cohort, including FZD9 (7q11.23) which encodes a member of the frizzled (FZD) family of receptors involved in activation of the Wnt cell signalling pathway

(Table 4.20). This pathway regulates cell survival, migration, and proliferation and has been implicated in numerous cancers (reviewed by (Taipale and Beachy 2001)). It remains unclear whether the hypomethylation and thereby potential overexpression of FZD9 identified reflects the young patient age or unfavourable tumour histology of the clustering subgroup. Support for anaplasia as the underpinning factor comes from a gene expression analysis of 34 ependymomas which found that WHO grade III tumours significantly overexpressed members of the Wnt pathway, including the FZD genes, when compared to ependymomas of classic histology (Palm, Figarella-Branger et al. 2009).

4.4.1.4 Candidate genes demonstrating genomic abnormalities associated with intracranial ependymomas of anaplastic histology

Anaplastic supratentorial ependymomas were associated with focal regions of genomic gain across chromosomes 7q, 11q, 13q and 19 when compared to location-matched tumours of classic histology. These loci encompassed numerous genes, including FGF3 (11q13.3), CDK8 (13q12.13), FLT1 (13q12.2) and the aforementioned COL4A1 (13q34). FGF3 encodes a fibroblast growth factor, overexpression of which has been has been associated with bladder, breast and non-small cell lung cancer (Naidu, Wahab et al. 2001; Tai, Sham et al. 2006; Zhang, Xie et al. 2006), while CDK8 encodes a cyclin dependent kinase associated with adverse prognosis in colorectal cancer (Firestein, Bass et al. 2008). FLT1, a gene frequently overexpressed in intracranial ependymoma (de Bont, Packer et al. 2008), is thought to positively regulate tumour angiogenesis as it encodes the receptor for vascular endothelial growth factor (VEGF) (Kerber, Reiss et al. 2008). The gain of FLT1 identified from this study thereby emphasised the association of anaplasia with increased tumour vascularity already established from ependymoma expression array work (Palm, Figarella-Branger et al. 2009).

High grade posterior fossa ependymomas were associated with focal regions of gain on chromosome 12 when compared to grade II counterparts, encompassing genes such as KNTC1/ROD (12q24.31). This encodes a protein responsible for mediating cell division (Chan, Jablonski et al. 2000), upregulation of which has been reported in hepatocellular

carcinoma and metastatic prostate cancer (LaTulippe, Satagopan et al. 2002; Mas, Maluf et al. 2009).

4.4.1.5 Candidate genes on chromosome 1q demonstrating genomic imbalance associated with an unfavourable patient prognosis

By integrating 500K SNP array findings for the primary paediatric ependymoma cohort with corresponding patient survival data, the genomic gain of selected genes on chromosome 1q was found to be associated with an unfavourable outcome. These genes were chosen from the numerous potential prognostic candidates on this chromosome arm as a consequence of preceding findings from the SNP analysis, as presented in this chapter. Gain of the aforementioned NAV1 (1q32.1) was associated with a worse event-free survival (Tables 4.12 and 4.13, Figure 4.5). Focal gain within the 1q21.2 − 21.3 region, encompassing the genes BNIPL and PRUNE as identified by the Affymetrix® Netaffx annotation file (build 07.12.07), was also associated with an adverse event-free survival on univariate analysis and moreover, was an indicator of reduced overall patient survival on multivariate analysis (Tables 4.12 and 4.14, Figure 4.6).

BNIPL (1q21.3) encodes two alternative transcripts, BNIPL-1 and BNIPL-2 (Bcl-2/adenovirus E1B 19 kDa interacting protein-2 like 1 or 2) (Shen, Hu et al. 2003; Xie, Qin et al. 2007). Although BNIPL-1 has been reported to promote apoptosis and inhibit cell growth (Shen, Hu et al. 2003), overexpression of BNIPL-2 has been shown to increase hepatocellular cancer cell invasiveness and metastasis in vitro, whilst increase the rate of liver and lung metastases in nude mice in vivo following the hepatic implantation of tumours derived from subcutaneous xenografts (Xie, Qin et al. 2007). PRUNE (1q21.3) encodes a member of the DHH phosphoesterase superfamily (Aravind and Koonin 1998). Through interactions with proteins such as the serine-threonine kinase GSK-3β, the metastasis regulating NM23-H1 and the cell adhesion controlling gelsolin, PRUNE is thought to promote cell motility and invasiveness respectively (Garzia, Roma et al. 2006; Kobayashi, Hino et al. 2006; Marino and Zollo 2007; Garzia, D'Angelo et al. 2008). Clinically, PRUNE overexpression has been reported as a frequent finding in sarcomas (Forus, D'Angelo et al. 2001) and other malignancies such as colorectal, pancreatic, oesophageal, gastric and breast cancer where it has been

202

associated with increased invasiveness, disease progression and adverse patient prognosis (Zollo, Andre et al. 2005; Kobayashi, Hino et al. 2006; Oue, Yoshida et al. 2007; Noguchi, Oue et al. 2009). PRUNE is discussed further in subsequent chapters.

As stated earlier, consideration must be given to all genes within a designated region if interest. While the SNP array annotation file used only acknowledged BNIPL and PRUNE within the 1q21.2 – 21.3 locus, detailed analysis using the genomic database Ensembl revealed two other genes were also encompassed, CDC42SE1 and AF1q. CDC42SE1 encodes a protein that binds to the Rho GTPase CDC42 and modulates its activities, including the regulation of cell cycle progression, cell growth and migration, neurite outgrowth, neuronal development, neural progenitor polarity and fate, malignant transformation and cell invasiveness (reviewed by (Govek, Newey et al. 2005)); (Wilkinson, Paterson et al. 2005; Cappello, Attardo et al. 2006). The precise function of AF1q remains unclear, although it is known to form a fusion gene with MLL in haematological malignancies (Tse, Zhu et al. 1995), while overexpression of its encoded protein has been identified as an adverse prognostic marker in acute myeloid leukaemia (Tse, Meshinchi et al. 2004). In vitro studies have also shown AF1q overexpression promotes proliferation and invasiveness of breast cancer cells (Chang, Li et al. 2008). The recent integrated SNP and expression study of 204 ependymomas by Johnson and colleagues, published at the completion of this study, identified CDC42SE1 and AF1q, yet not BNIPL or PRUNE, as putative ependymoma oncogenes demonstrating copy-number driven overexpression (Johnson, Wright et al. 2010). Consequently, NAV1, BNIPL, PRUNE, CDC42SE1 and AF1q and the proteins they encode warrant further evaluation within the context of establishing new markers of outcome and potential therapeutic targets for paediatric ependymomas. Indeed, in the sixth chapter the prognostic potential of NAV1 and PRUNE protein expression is assessed by immunohistochemistry on paediatric ependymomas from two aged-defined patient cohorts, each being treated uniformly within the confines of a clinical trial (UK CCLG 1992 04 and SIOP 1999 04). This work is now being extended to assess a prognostic role for CDC42SE1, AF1q and BNIPL expression (Chapter 7).

4.4.2 Acquired uniparental disomy as a mechanism in paediatric ependymoma pathogenesis

Whilst requiring validation, this work is the first to indicate aUPD as a potential, albeit infrequent, mechanism of pathogenesis in paediatric ependymoma. Focal regions of aUPD were identified in the primary tumour cohort, implicating genes associated with diverse cellular processes such as motility, invasiveness, angiogenesis, DNA replication and neurite development. However this was uncommon, being observed in only 2/38 (5 %) cases (Table 4.15). As a consequence, further evaluation of aUPD in primary paediatric ependymomas was not performed since it appeared only relevant for an extremely small subset of tumours. Nevertheless, individual recurrent ependymomas demonstrated copy number neutral LOH at specific genomic regions which were not present in the corresponding patient's primary tumour, the most common encompassing C9orf138 within chromosome 9p22.1 (2/6 tumours; 33 %). This suggests a possible role for aUPD in tumour progression which may warrant further investigation.

4.4.3 Accuracy of real time quantitative PCR validation

The copy number alterations found in 11 of the genes identified from the 500K SNP array analysis and reviewed above (NSL1, DNAJC25, FILIP1, FRK, NAV1, CDKN2A, CHI3L1, HOXA5, TXN, BNIPL, and PRUNE) were validated in a subset of ependymomas using qPCR. The genomic imbalances of four other selected genes were unable to be validated (SEH1L, PPARA, COL4A1, TELO2). As a result, the overall correlation of SNP array derived copy number and qPCR derived copy number for the 15 candidate genes analysed was significant (R = 0.664, p = 0.01, two-tailed Spearman's rank test) (Figure 4.19).
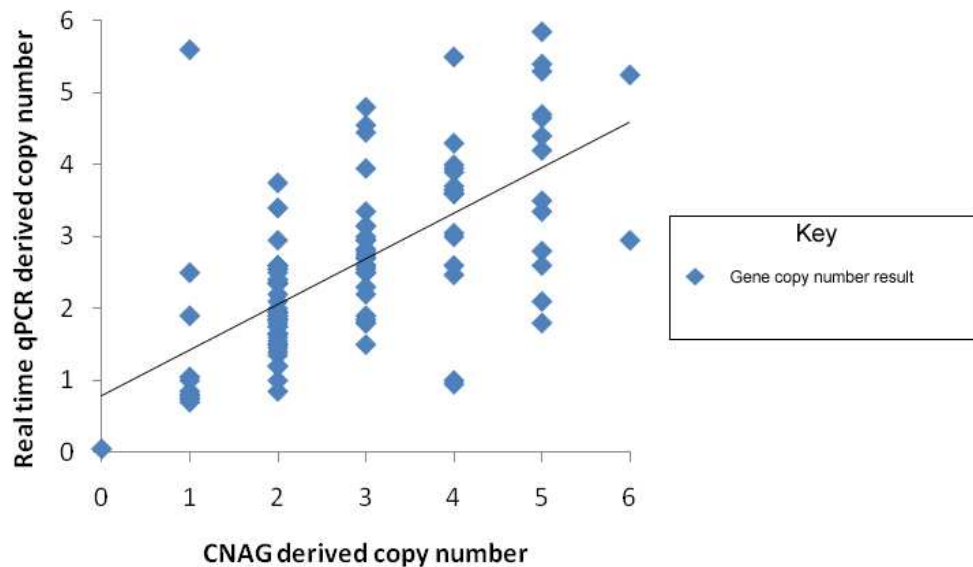
Figure 4.19: Overall correlation of copy number results derived from the 500K SNP array analysis (using the CNAG algorithm) and quantitative PCR for the 15 candidate genes of interest (as defined in Table 4.21). Using Spearman's rank correlation (two tailed), the resulting R value was 0.664, significant at the 0.01 level. The correlation appears relatively strong for genomic loss, copy number balance and low level gain, but weakens at higher copy number values.

In total, 59 independent copy number gains of the 15 selected genes identified from the SNP array were subject to validation by qPCR. The majority were validated (48/59, 81 %). Eleven array-derived gene gains (19 %) were deemed false positives, demonstrating either a diploid copy number or genomic loss on qPCR. Validation was also performed on a total of twelve individual genomic losses of candidates from the SNP array. Of these, nine (75 %) were confirmed by qPCR with a matching copy number value, while three (25 %) were classed as false negatives, demonstrating either a normal or increased qPCR copy number value. In addition to the verification of genomic imbalances, validation of normal copy number was undertaken. Altogether, 50 gene regions with a diploid copy number according to the SNP array were assessed by qPCR, with validation being achieved in most cases (35/50, 70 %). False normals were identified however, with nine cases (18 %) revealing gain on qPCR and six cases (12 %) showing loss. Overall, qPCR appeared a relatively accurate method of confirming genomic loss, diploid copy number states and single copy number gains in genes identified from the 500K SNP array. However this precision diminished when attempting to validate high-level gene gain and amplification (Figure 4.19). FISH would have represented an appropriate alternative validation technique in such circumstances but was not performed due to the time constraints of the project.

The majority of discrepancies between the copy number results from the two methods can be accounted for by the unavoidable use of DNA which had to be re-extracted from certain tumours for the purposes of qPCR validation, since original stocks had been depleted as a result of the array analysis itself, or follow-on collaborative projects that had commenced prior to this project's undertaking (section 4.2.1.4). For example, of the 19 ependymoma samples used in the unsuccessful validation of SEH1L, COL4A1, TELO2 and PPARA copy number changes identified from the SNP array, 16 (84 %) were reliant on tumour samples that had undergone DNA re-extraction. While there is clear evidence that paediatric ependymomas demonstrate histological heterogeneity (Chapter 1, section 1.2.4, Figure 1.1), this finding suggests that genetic diversity also exists between different sections cut from the same tumour and highlights the importance of using a single DNA source for the validation of SNP array projects in such cases.

In contrast to SEH1L, COL4A1, TELO2 and PPARA, the successful copy number validation of eleven genes in this study (NSL1, DNAJC25, FRK, FILIP1, NAV1, CDKN2A, CHI3L1, HOXA5, TXN, BNIPL, and PRUNE) was still achieved, despite incorporating re-extracted ependymoma DNA samples. It can therefore be argued that these latter genes, demonstrating consistently identifiable genomic imbalances within different sections of a heterogeneous tumour, remain more appropriate to evaluate further as candidates of pathogenesis since these aberrations more accurately reflect the biological milieu of the entire tumour. Moreover, such anomalies could potentially be driving clonal expansion in the majority of cells comprising the tumour mass, as opposed to being inconsistent and inconsequential downstream mutations, although this requires verification through functional analysis.

4.5 Summary

Using Affymetrix® 500K SNP arrays and Illumina® GoldenGate® Cancer Panel I methylation arrays, a high resolution genomic and epigenetic analysis of paediatric ependymomas has been performed. While this is the first study to reveal aUPD as a possible yet infrequent mechanism of disease, alterations in the copy number and

methylation state of several genes potentially implicated in ependymoma pathogenesis were identified.

Across primary intracranial ependymomas, the most common regions of genomic gain encompassed genes such as NSL1, SEH1L and DNAJC25, whereas frequent regions of loss harboured the putative tumour suppressor genes FILIP1 and FRK/RAK. Among recurrent tumours, the gain of genes encoding kinetochore associated proteins was again observed, while distinct focal genomic imbalances were identified including the frequent gain of NAV1, loss of candidates such as CBX7 and PPARA, and homozygous deletion of CDKN2A and CDKN2B.

Other alterations were associated with particular clinical and prognostic subgroups. Anomalies specific to ependymoma location were evident, such as EYA4 hypomethylation in spinal tumours and hypermethylation of RASSF1A in intracranial tumours. Within the intracranial cohort, supratentorial ependymomas were associated with the gain of genes such as TULP4 and COL4A1, whilst posterior fossa tumours were characterised by the hypomethylation of MMP2, PPARG, SPDEF and BCR, alterations in members of the chitinase gene family (CHI3L1 and CHIT1 gain, CHI3L2 hypomethylation) and focal gains elsewhere across chromosome 1q, harbouring genes such as PARP1 and AKT3. Genomic imbalances associated with ependymomas from patients below three years of age were observed, including gains encompassing TELO2, TMEM18G and HDAC10, whereas gain incorporating genes such as TXN was a feature of tumours from older children. Focal imbalances also differentiated WHO grade III ependymomas from those exhibiting classic histology. Indeed, anaplastic supratentorial ependymomas were associated with gain of FGF3, CDK8, FLT1 and COL4A1, while gain encompassing KNTC1/ROD characterised high grade tumours from the posterior fossa. Moreover, focal gain encompassing selected candidates on chromosome 1q (NAV1, PRUNE, BNIPL, CDC42SE1 and AF1q) was associated with adverse patient survival.

The genomic imbalances identified in eleven such candidate genes (NSL1, DNAJC25, FRK, FILIP1, NAV1, CDKN2A, CHI3L1, HOXA5, TXN, BNIPL, and PRUNE) were confirmed using real time qPCR. Validation of other findings from both array platforms is discussed in the final chapter. Expression array analysis and immunohistochemistry

are now required to determine whether the copy number alteration of these and all other candidate genes within the focal regions of imbalance presented, is sufficient to drive subsequent gene and protein expression. Indeed, the IHC assessment of NAV1 and PRUNE protein expression in a subset of paediatric ependymomas analysed on the SNP array has been performed, the results of which are presented and discussed in the penultimate chapter. Such analyses would also reaffirm which candidates display an expression profile in paediatric ependymoma that correlates with, and is potentially attributable to, an altered methylation state. Once established, the precise role of genes purported to drive tumourigenesis could then be ascertained functionally, using cell lines derived in-house from childhood ependymomas to manipulate in vitro gene expression. This again is discussed further in the final chapter.