Recurrent Paediatric Ependymoma: A Multicentre Analysis of Clinical Features and Tumour Biology in the Molecular Era

Timothy Albert Ritzmann, BMBS, MRCPCH

Thesis submitted to the University of Nottingham
for the degree of doctor of philosophy

February 2018

Abstract

Introduction: Ependymoma is the second most common malignant brain tumour of childhood. 50% of children with primary disease recur; three-quarters of these do not achieve long term survival. In the 'molecular era' of cancer research, diagnosis combines advanced molecular profiling with histopathological assessment. Whilst primary ependymomas can be classified based on epigenetic and transcriptomic features, there is little information on molecular signatures at recurrence. However, some small studies have implicated cancer immunity. Trials of novel therapies at recurrence have been disappointing. This study undertook molecular profiling of recurrent ependymoma, combined with contemporary clinical data, to better understand recurrence biology and potential therapy options. **Methods:** Clinical outcomes for 188 children with recurrent ependymoma were analysed. Cases with primary and matched recurrent samples were included in DNA methylation (n=56), RNA sequencing (n=52) and immunohistochemical (IHC) (n=56) analyses. RNA sequencing from FFPE tissue was validated to expand the cohort. **Results:** Recurrence was the strongest predictor of long term survival. Treatment approach at primary diagnosis was not associated with survival, but radiotherapy at first recurrence was associated with better short-term outcomes. Children with the commonest DNA methylation based diagnoses, EPN_PFA and EPN_RELA, had equally poor outcomes. RNA sequencing from FFPE tissue was effective, therefore tumours sequenced from FFPE and FF tissue were included in paired gene expression analyses. Transcriptomic and DNA methylation analyses identified three similar subgroups in FFPE and FF cohorts (PF1, PF2 and ST). At first recurrence, PF1 was associated with downregulated immune and inflammatory ontologies, which may indicate tumour immune escape. PF2 and ST subgroups demonstrated upregulation of ontologies associated with adaptive immunity. Despite this, there was little evidence of change in either immunogenicity or T-cell effector activity at first recurrence. IHC analysis identified a fall in inflammatory cells in posterior fossa tumours at recurrence and indicated that ependymoma is an immune excluded tumour. **Conclusions:** This study highlights both the abysmal prognosis for this disease, and the need for a better understanding of tumour biology to improve outcomes. This study has contributed novel data on changes at recurrence across molecular subgroups, and identified the immune excluded nature of ependymoma, which may be important in guiding therapy. The validation of RNA-seq from FFPE in childhood brain tumours has facilitated access to a large set of previously uninvestigated samples.

Acknowledgements

I must first extend my gratitude to Dr Judith Grant and Professor Helen Budge who facilitated three years away from my clinical paediatric training and provided a great deal of support and encouragement along the way. My two supervisors, Professor Richard Grundy and Dr Hazel Rogers provided constant support and advice, and helped me to identify ways forward in times of uncertainty. Professor Grundy also sourced the funding for this project and identified me as someone who would be able to undertake it – for this opportunity I am extremely grateful. Dr Rogers was particularly helpful in the structure and layout of the final text and made many useful comments on the experimental approach throughout.

During my time at the CBTRC I was lucky enough to work with many exceptional people. Dr Jennifer Ward, Dr Rebecca Chapman and Dr Maria Estevez-Cebrero taught me various techniques and provided constant supplies of sensible advice. Dr Lisa Storer was fundamental in helping me to understand the databases and how to go about collecting more data. Dr Anbarasu Lourdusamy provided useful advice on the bioinformatic analyses. Special thanks go to Dr Simon Paine who taught me a great deal about ependymoma pathology. Thanks also go to the numerous people who collaborated on this project, particularly Dr Nicholas Foreman, Andrew Donson and their team in Denver, Colorado; and Dr Thomas Jacques and Dr Alex Virasami at the UCL Institute of Child Health. Kristian Pajtler's advice from Heidelberg on DNA methylation profiling was also invaluable. The number of clinicians and researchers who provided patient data and tissue samples, from across the UK and beyond, are too numerous to mention here but I am grateful to them all.

This project would not have been possible without the support of its funders; the University of Nottingham, Fighting Ependymoma and The James Tudor Foundation. These organisations perform fantastic work and their support has not gone unappreciated.

Finally, love and thanks go to my family. Mum and Dad who have been nothing less than 100% supportive. Virginia – my beautiful wife – who has been a fountain of tolerance, love and support throughout this process and has provided extremely valuable tangible support by proofreading my thesis. Joshua – our son – who turned our world upside down and inside out when he arrived nearly one year ago, but has enriched it immeasurably ($p<0.001$).

Table of Contents

Table of Tables

Table of Figures

Abbreviations

BAM	Binary Sequence Alignment/Map
BJLE	BJ hTERT + SV40 Large T+ (Cell line)
BLAST	Basic Local Alignment Search Tool
BP	Base Pairs
CBTRC	Children's Brain Tumour Research Centre
CCLG	Children's Cancer and Leukaemia Group
CD	Cluster of Differentiation
CGH	Comparative Genomic Hybridisation
CI	Confidence Interval
CNS	Central Nervous System
CPH	Cox Proportional Hazards
CSC	Cancer Stem Cell
CT	Chemotherapy
CTA	Cancer-Testis Antigen
CYT	Cytolytic Activity
DAB	3,3'-diaminobenzidine
DAVID	Database for Annotation, Visualisation and Integrated Discovery
DIG1	Desmoplastic Infantile Glioma
DKFZ	The German Cancer Research Centre
DNA	Deoxyribonucleic Acid
DPX	Distyrene, Plasticiser, Xylene
ECM	Extracellular Matrix
EFS	Event Free Survival
ENCODE	Encyclopaedia of DNA Elements
EPN_MPE	Myxopapillary Ependymoma
EPN_PFA	Posterior Fossa A Ependymoma
EPN_PFB	Posterior Fossa B Ependymoma
EPN_RELA	RELA Fusion Related Ependymoma
EPN_YAP	YAP Fusion Related Ependymoma
FDR	False Discovery Rate
FF	Fresh Frozen
FFPE	Formalin Fixed, Paraffin Embedded
GBM	Glioblastoma Multiformae
GC	Guanine-Cytosine
GO	Gene Ontology
GOrilla	Gene Ontology Enrichment Analysis and Visualisation Tool
GSEA	Gene Set Enrichment Analysis
GTF	Gene Transfer Format
GTR	Gross Total Resection
H&E	Haematoxylin and Eosin
HGNET_BCOR	CNS High Grade Neuroepithelial Tumour with BCOR Alteration
HGNET_MN1	CNS High Grade Neuroepithelial Tumour with MN1 Alteration
HLA	Human Leucocyte Antigen
HPC	High Performance Cluster

HPF	High Powered Field
IDAT	Raw Intensity Data File
IFN	Interferon
IHC	Immunohistochemistry
LFC	Log Fold Change
MAPK	Mitogen-activated Protein Kinase
MDS	Multidimensional Scaling
MHC	Major Histocompatability Complex
mRNA	Messenger Ribonucleic Acid
NCBI	National Center for Biotechnology Information
NES	Normalised Enrichment Score
NF-kB	Nuclear Factor Kappa-Light-Chain-Enhancer of Activated B Cells
NGS	Next Generation Sequencing
NRT	No Reverse Transcriptase
OS	Overall Survival
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PE	Paired End
PF	Posterior Fossa
PNET	Primitive Neuroectodermal Tumour
PRC2	Polycomb Repressive Complex 2
QC	Quality Control
REVIGO	Reduce and Visualise Gene Ontology
RIN	Ribonucleic Acid Integrity Number
RLD	R-log Transformation (implemented by DeSeq2)
RNA	Ribonucleic Acid
RNA-seq	Ribonucleic Acid Sequencing
rRNA	Ribosomal Ribonucleic Acid
RT	Radiotherapy
SE	Single End
SIOP	International Society of Paediatric Oncology
SNP	Single Nucleotide Polymorphism
SP	Spine
ST	Supratentorial
STR	Subtotal Resection
Th1	T Helper 1
Th2	T Helper 2
TPM	Transcripts per Million
TRAP	Telomeric Repeat Amplification Protocol
TSG	Tumour Suppressor Gene
TTP	Time to Progression
WHO	World Health Organisation

1 Introduction

1.1 Childhood cancer

1.1.1 Epidemiology

Cancer is a leading cause of death in young people in the developed world. In the UK, it is the leading cause of death in 5-14 year olds, and second only to injuries and poisonings in under four year olds. In 15-19 year olds it is superseded by death related to risks and behaviours (Royal College of Paediatrics and Child Health and National Children's Bureau, 2014).

In the UK, around 1600 children are diagnosed with, and 250 die from, cancer annually. By age 14, 1 in 500 children develop cancer (Cancer Research UK, 2015a). The overall incidence for paediatric cancer in Europe is 130.9 per million (0-14 years) and 157 per million (0-19 years) (Steliarova-Foucher et al., 2004). Other reported incidences vary from 122.1 per million (England and Wales) to 160.6 per million (American Hispanics in Los Angeles) (Kaatsch, 2010). Boys develop malignancy more frequently than girls (Kaatsch, 2010).

The epidemiology of childhood cancer has been well studied through large national and international registries including SEER (Surveillance, Epidemiology and End Results programme), ACCIS (Automated Childhood Cancer Information System) and EUROCARE (European Cancer Registry based study on survival and care of cancer patients). Since the early 1970s the incidence of childhood cancer of all types has increased (Howlader N et. al., 2013; Steliarova-Foucher et al., 2004). Whilst this may be accounted for, in part, by improved diagnostic and reporting methods, the lack of uniformity across all cancer types suggests a genuine increase (Kaatsch et al., 2006).

In Europe, five-year survival has improved from 50% in the 1970s (Steliarova-Foucher et al., 2004) to 77.9% between 1999 and 2007 (Gatta et al., 2014). Whilst survival is improving, implications for long-term health are dire and attention must be paid to ameliorating late effects. In children who have a life expectancy of 70-80 years, simply reporting the risk of death to five years is inadequate to assess the true impact of the disease.

A disparity in survival has been noticed between the developed and developing world. Even comparing outcomes within Europe, survival is better in the West.

Factors suggested for this difference include variable diagnostic techniques and the availability of specialist centres and medications (Gatta et al., 2014).

1.1.2 Classification

Whilst adult malignancies are classified by site, cancer in childhood is less anatomically defined. Many adult malignancies are of epithelial tissue whilst in children this does not hold true (Birch and Marsden, 1987). Childhood cancer is a disease of development arising from early progenitor cells, with few or no apparent genetic mutations (Scotting et. al., 2005). As a consequence, these neoplasms may not be restricted to a specific location. With this in mind, a classification of childhood cancer, first described in 1987, was based on morphology rather than anatomy. More recently, authors have suggested that morphological classifications should be superseded by molecular classifications (Pajtler et al., 2015; Schwalbe et al., 2017; Sturm et al., 2016).

The International Classification of Childhood Cancer (ICCC-3) is based on the morphological codes outlined in the International Classification of Diseases for Oncology (ICD-O). The ICCC-3 was intended for international adoption by the large cancer registries (Steliarova-Foucher et. al., 2005). 12 main diagnostic groups were identified (Table 1-1).

Major Diagnostic Groups of Childhood Cancer	Incidence (/million)		Relative %	
	0-14yrs	0-19yrs	0-14yrs	0-19yrs
(1) Leukaemias, myeloproliferative and myelodysplastic diseases	42.4	37.3	32.3	25.8
(2) Lymphomas and reticuloendothelial neoplasms	14.3	21.0	10.9	14.5
(3) CNS and miscellaneous intracranial and intraspinal neoplasms	28.1	29.5	21.5	20.4
(4) Neuroblastoma and other peripheral nervous cell tumours	9.8	7.1	7.5	4.9
(5) Retinoblastoma	3.8	3.3	2.9	2.3
(6) Renal tumours	8.5	7.1	6.5	4.9
(7) Hepatic tumours	1.4	1.4	1.1	1.0
(8) Malignant bone tumours	5.4	7.6	4.1	5.3
(9) Soft tissue and other extra osseous sarcomas	8.3	9.4	6.3	6.5
(10) Germ cell tumours, trophoblastic tumours, and neoplasms of gonads	4.0	8.0	3.1	5.5
(11) Other malignant epithelial neoplasms and malignant melanomas	3.4	9.9	2.6	6.8
(12) Other malignant neoplasms	1.2	2.9	0.9	2.0
Total:	130.9	144.6	100	100

Table 1-1: The classification of childhood cancer into 12 subgroups. Incidence and percentage contributions of each type according to age. CNS malignancies are the second most common type overall. Adapted from Howlader N et.al., 2013; Kaatsch, 2010.

1.2 Paediatric brain tumours

Tumours of the central nervous system (CNS) are the second most common neoplasm in children, accounting for 26% of the total (Cancer Research UK, 2015a), and the most common type of solid tumour (Kaatsch, 2010). They grow in a rapidly developing anatomical compartment, associated with eloquent structures. This has implications for all stages in care including diagnosis, investigation, treatment and follow up.

1.2.1 Classification

CNS tumours may be classified according to:

(1) Cell origin and relationship to the brain;

(2) Location within the CNS;

(3) Histological grade;

(4) The World Health Organisation (WHO) classification, including molecular features for specific tumours.

Tumours arising from glial, neural or choroid plexus cells are intrinsic to the brain parenchyma, whilst tumours arising from Rathke's pouch and germ cells are extrinsic to the brain parenchyma. Tumours can then be further subdivided by cell of origin (Table 1-2) (Kieran et al., 2015).

Relationship to brain	Cell Type	Tumour Type
Extrinsic	Rathke's Pouch Epithelium	Craniopharyngioma
	Germ cell	Germinoma
Intrinsic	Glial	Astrocytoma
		Ependymoma
		Oligodendroglioma
	Neural	Medulloblastoma
		PNET*
	Choroid Plexus	Choroid Plexus Carcinoma

Table 1-2: Cellular origin and relationship to brain of paediatric brain tumours. *Primitive Neuroectodermal tumours (PNETs) are increasingly recognised as a group of distinct entities (Sturm et al., 2016) and have recently been removed from the WHO classification.

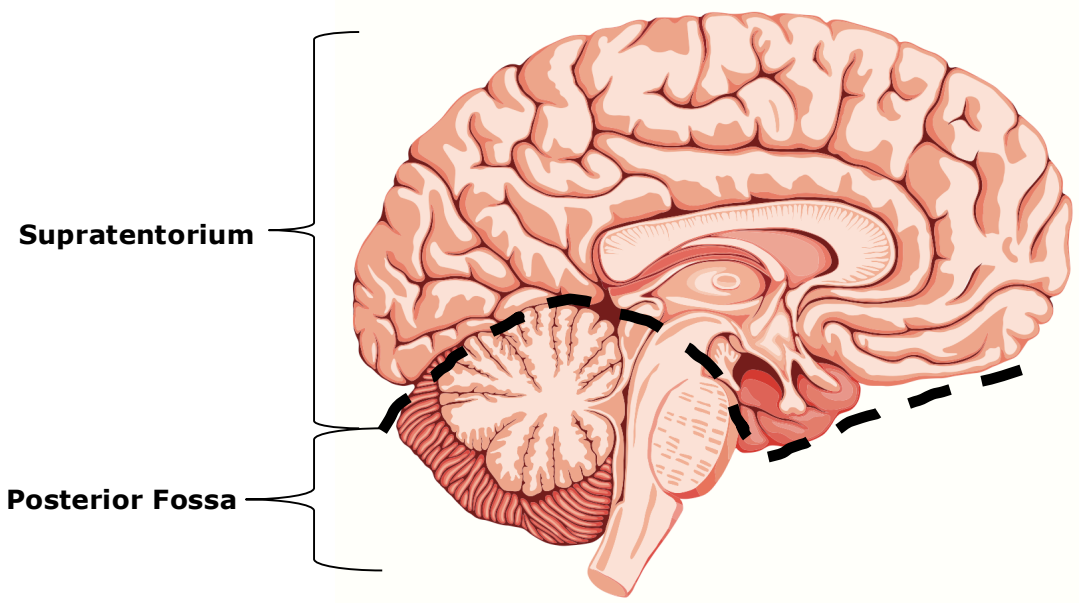


Figure 1-1: Location of supratentorial (ST) and posterior fossa (PF, infratentorial) regions. The supratentorium contains the cerebrum and the posterior fossa contains the cerebellum and brain stem. The dashed black line indicates the boundary between the two areas. Image modified from stock.adobe.com/#93141325.

Tumours defined by cell of origin may occur in more than one anatomical location. For brain intrinsic tumours, location can be divided between the posterior fossa and supratentorium. The posterior fossa encompasses the region

below the tentorium cerebelli and above the foramen magnum, containing the cerebellum and brainstem. The supratentorium includes the remainder of the cranial cavity above the tentorium cerebelli and contains the cerebrum. Paediatric tumours may also occur in the spine (Figure 1-1).

Histological grade differs across tumour types, but in general:

- Grade 1: Low proliferative potential and possible surgical cure;
- Grade 2: Infiltrative with low level proliferative activity and often recur;
- Grade 3: Evidence of malignancy (may include nuclear atypia and increased mitotic activity);
- Grade 4: Malignant, cytologically active, necrotic tumours which rapidly progress and result in death (Louis et al., 2007).

The WHO classification has previously been based upon histological appearance, taking into account tumour grade (Louis et al., 2007). The 2016 revision of the classification included, for the first time, the addition of molecular stratification. New molecularly defined entities include IDH-mutant glioblastomas, H3-K27M mutated diffuse midline gliomas, SHH- and WNT-activated medulloblastomas and *RELA* fusion-positive ependymomas (Louis et al., 2016) (Table 1-3). The inclusion of molecular features as part of this classification is likely to expand over time as more evidence of the biological differences between tumour types emerges.

WHO grades of select CNS tumours	
Diffuse astrocytic and oligodendroglial tumours	
Diffuse astrocytoma, IDH-mutant	II
Anaplastic astrocytoma, IDH-mutant	III
Glioblastoma, IDH-wildtype	IV
Glioblastoma, IDH-mutant	IV
Diffuse midline glioma, H3 K27M–mutant	IV
Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	II
Anaplastic oligodendroglioma, IDH-mutant and 1p/19q-codeleted	III
Other astrocytic tumours	
Pilocytic astrocytoma	I
Subependymal giant cell astrocytoma	I
Pleomorphic xanthoastrocytoma	II
Anaplastic pleomorphic xanthoastrocytoma	III
Ependymal tumours	
Subependymoma	I
Myxopapillary ependymoma	I
Ependymoma	II
Ependymoma, *RELA* fusion–positive	II or III
Anaplastic ependymoma	III
Other gliomas	
Angiocentric glioma	I
Chordoid glioma of third ventricle	II
Choroid plexus tumours	
Choroid plexus papilloma	I
Atypical choroid plexus papilloma	II
Choroid plexus carcinoma	III
Neuronal and mixed neuronal-glial tumours	
Dysembryoplastic neuroepithelial tumour	I
Gangliocytoma	I
Ganglioglioma	I
Anaplastic ganglioglioma	III
Dysplastic gangliocytoma of cerebellum (Lhermitte–Duclos)	I
Desmoplastic infantile astrocytoma and ganglioglioma	I
Papillary glioneuronal tumour	I
Rosette-forming glioneuronal tumour	I
Central neurocytoma	II
Extraventricular neurocytoma	II
Cerebellar liponeurocytoma	II
Tumours of the pineal region	
Pineocytoma	I
Pineal parenchymal tumour of intermediate differentiation	II or III
Pineoblastoma	IV
Papillary tumour of the pineal region	II or III
Embryonal tumours	
Medulloblastoma (all subtypes)	IV
Embryonal tumour with multilayered rosettes, C19MC-altered	IV
Medulloepithelioma	IV
CNS embryonal tumour, NOS	IV
Atypical teratoid/rhabdoid tumour	IV
CNS embryonal tumour with rhabdoid features	IV
Tumours of the cranial and paraspinal nerves	
Schwannoma	I
Neurofibroma	I
Perineurioma	I
Malignant peripheral nerve sheath tumour (MPNST)	II, III or IV
Meningiomas	
Meningioma	I
Atypical meningioma	II
Anaplastic (malignant) meningioma	III
Mesenchymal, non-meningothelial tumours	
Solitary fibrous tumour / haemangiopericytoma	I, II or III
Haemangioblastoma	I
Tumours of the sellar region	
Craniopharyngioma	I
Granular cell tumour	I
Pituicytoma	I
Spindle cell oncocytoma	I

Table 1-3: Selected features of the WHO classification of the tumours of the central nervous system and possible grades for each tumour. Reproduced from Louis et. al. 2016.

1.3 Paediatric ependymoma

Ependymoma is the second most common malignant brain tumour of childhood (Peris-Bonet et al., 2006). Morphologically, it mimics the ependymal lining of the central nervous system (Lehman, 2008). It can arise anywhere in the neuraxis and presents in children and adults, with a median age of diagnosis of around five years (Perilongo et al., 1997). Childhood tumours arise most commonly in the supratentorial (ST) or posterior fossa (PF) regions of the brain, and adult tumours in the spine. Diagnosis is based upon clinical presentation, imaging and histopathology. Symptoms of a developing tumour depend upon location of the lesion, but commonly include headaches, vomiting and ataxia.

With the advent of molecular biology, new subtypes of ependymoma are being proposed, but the implications of these for clinical outcome and treatment strategy are not yet clear (Hoffman et al., 2014a; Pajtler et al., 2015; Wani et al., 2012; Witt et al., 2011). In the future, DNA methylation profiling may have a role (Pajtler et al., 2015) and is being introduced into clinical trials.

1.3.1 Epidemiology

Ependymoma constitutes approximately 10% of all paediatric brain tumours. Incidences have been reported as 4 per million in 0-14 year olds and 3.7 per million in 0-19 year olds, between 2007 and 2011, in the USA (Howlader N at. al. 2013). In Europe between 1978 and 1997, incidences of 3.4 per million in 0-14 year olds were reported (Peris-Bonet et al., 2006). It is most common in the first three years of life (Peris-Bonet et al., 2006; Rickert and Paulus, 2001), and it has been suggested that these younger children have the worst outcomes (Agaoglu et al., 2005; Jaing et al., 2004; Messahel et al., 2009; Pollack et al., 1995).

Reported overall survival (OS) and event free survival (EFS) rates vary in published clinical trials (Evans et al., 1996; Grill et al., 2001; Grundy et al., 2007; Robertson et al., 1998; Venkatramani et al., 2013). Some studies indicate very poor outcomes with two year OS and EFS of 40% and 23% respectively (Duffner et al., 1998), whilst others found better outcomes with five-year OS and EFS of 71% and 57% (Garvin et al., 2012). These studies have been conducted over long time periods and now represent relatively old case series; it is therefore possible that outcomes have improved over time. This hypothesis is supported by survival data from more recent series published after 2010. A large study of 146 patients found a five-year OS and EFS of 82.6% and 68.9% respectively (Godfraind et al., 2012). Another indicated three year OS of 100% and 73% for

ST and PF tumours respectively (Venkatramani et al., 2013). A third study found five-year OS and EFS of 81.1% and 65.4% in a series of 160 patients with mixed tumour locations (Massimino et al., 2016).

A key factor in ependymoma survival is recurrence. The recurrence rate approaches 50% at five years and only 24-27% of children survive five years after the first recurrence (Messahel et al., 2009).

The only other consistently reported clinical factor associated with improved EFS is extent of surgical resection (Agaoglu et al., 2005; Grill et al., 2001; Jaing et al., 2004; Massimino et al., 2004; Perilongo et al., 1997; Robertson et al., 1998), however this is not a universal finding (Grundy et al., 2007). Achieving a gross total resection (GTR) is difficult due to the delicate nature of the surrounding structures and has been reported to only be attainable in 50-60% of cases (Bouffet et al., 1998; Perilongo et al., 1997). However, some centres report rates of up to 81.7% (Merchant et al., 2009). Resection rates can be improved by second-look surgery (Vinchon et al., 2005).

One study found that duration of symptoms before diagnosis was an independent risk factor for poor outcome in both univariate and multivariate analyses (Pollack et al., 1995). This finding supports the recent drives for earlier diagnosis of paediatric brain tumours by HeadSmart, a UK campaign that has reduced the time from first symptoms to diagnosis (Shanmugavadivel et al., 2015; Wilne et al., 2010).

1.3.2 Current approaches to treatment

The main accepted treatment strategies in ependymoma management include surgery, radiotherapy and chemotherapy.

Surgery is used in all patients, with an aim to achieve GTR (Jaing et al., 2004; Paulino et al., 2002). Adjuvant radiotherapy is often given to older children at diagnosis, and radiotherapy is recommended for all at recurrence (Grundy et al., 2007; Messahel et al., 2009; Perilongo et al., 1997). Despite evidence of its efficacy (Koshy et al., 2011; Merchant et al., 2009; Snider et al., 2017), children under three generally do not receive radiotherapy in view of concerns about late effects; more recently it has been suggested that this age cut off should be lowered to 12-18 months (Rudà et al., 2017). There is an increasing role for proton beam radiotherapy (Indelicato et al., 2017). Instead of radiotherapy, the youngest children often receive chemotherapy (Grill et al., 2001; Grundy et al.,

2007; Rudà et al., 2017). Some argue that the use of chemotherapy in younger children has been a relatively successful approach, with studies demonstrating that radiotherapy may be delayed in around 40% of infants with non-metastatic disease (Duffner et al., 1993; Geyer et al., 2005; Grill et al., 2001; Grundy et al., 2007; Strother et al., 2014). Others have suggested that chemotherapy is only beneficial in supratentorial disease (Venkatramani et al., 2013), or not at all (Evans et al., 1996).

Long-term effects of cancer therapy must be balanced against the risk of disease progression. A recognised dilemma in many paediatric cancers, including acute lymphoblastic leukaemia and optic pathway glioma (Taylor et al., 2008; Vora et al., 2013). Radiotherapy has significant neurocognitive impact and is associated with declines in memory, processing speed, attention, visual, motor and intellectual and executive functioning (Shortman et al., 2014; Spiegler et al., 2004).

A cohort of 26 patients from 12 Canadian centres provided evidence that, in children aged over two years with low proliferative markers and in whom GTR had been achieved, radiotherapy might be avoided completely (Ailon et al., 2014). This supports other research describing a group of PF ependymomas (EPN_PFB) with a markedly improved prognosis (over 90% EFS and OS), that may be amenable to more conservative treatment (Pajtler et al., 2015; Ramaswamy et al., 2016; Witt et al., 2011). This finding needs to be verified in an adequately powered, prospective trial before introduction into the clinical environment.

The development of evidence based treatment strategies requires large prospective cohorts and prospectively designed, randomised, trials rather than small cohorts from single centres (Bouffet et al., 1998). Some progress has been made in this respect through multi-centre co-operation and a summary table of prospective clinical trials of different combinations of conventional therapies in ependymoma is presented (Table 1-4). Despite this progress, of the 11 studies considered, only one included in excess of 100 patients (Massimino et al., 2016). A further difficulty in interpreting these studies is that whilst all used the conventionally accepted approaches of surgery, with or without chemotherapy and radiotherapy, the clinical cohorts and experimental protocols varied, generating a heterogeneous dataset.

Study	Patients	OS	EFS	Intervention	Main Outcome
(Evans et al., 1996)	42	39% (10yr)	36% (10yr)	Surgery and RT +/- CT.	CT not effective in improving outcome.
(Duffner et al., 1998)	48	40-50% (2yr)	23% (2yr)	2 year delayed RT in children <2 years, 1 year delayed in children >2 years. CT until RT delivered.	Youngest children had worse outcomes. Attributed to longer RT delay.
(Robertson et al., 1998)	32	64% (5yr)	50% (5yr)	Surgery and RT plus one of two possible CT regimens.	Outcome not affected by either CT regimen.
(Grill et al., 2001)	73	59% (4yr)	22% (4yr)	CT to avoid RT in children <5 years.	RT avoided or delayed in substantial proportion of <5 year olds.
(Geyer et al., 2005)	74	59% (5yr)	32% (5yr)	Children <3 years randomised to one of two CT arms. RT for progression.	No difference in outcomes between the two CT arms.
(Grundy et al., 2007)	89	63% (5yr)	42% (5yr)	CT to avoid RT in children <3 years.	RT avoided or delayed in substantial proportion of under 3 year olds.
(Massimino et al., 2011)	41	37% (5yr)	26% (5yr)	Upfront CT to avoid RT in young children.	High failure rates demonstrated. Neurocognitive outcomes not improved for those who did not receive RT.
(Garvin et al., 2012)	84	71% (5yr)	57% (5yr)	CT pre-RT in children with STR.	Those with near total resection had equivalent outcomes to GTR. Those with STR had suboptimal outcomes.
(Venkatramani et al., 2013)	19	73% (3yr)	27% (3yr)	Intensive induction and consolidation CT prior to RT.	Potential to improve outcomes in children with ST disease. Appeared ineffective in PF disease.
(Strother et al., 2014)	82	29% (10yr)	18% (10yr)	DI versus standard CT in children <3 years. RT for progression.	Treatment with DI resulted in better EFS but not OS.
(Massimino et al., 2016)	160	65% (5yr)	81% (5yr)	3 arms stratified by extent of resection and grade: (1) STR: CT + RT (2) GTR+grade II: RT (3) GTR+grade III: CT+RT	Worst outcomes seen in children with STR followed by GTR and grade III. Best outcomes in GTR and grade II disease.

Table 1-4: Summary of multicentre prospective clinical trials using conventional therapies in paediatric ependymoma. CT: Chemotherapy. RT: Radiotherapy. DI: Dose intensification.

1.3.3 Clinical and molecular features associated with outcome

Numerous clinical and molecular features have been associated with clinical outcome in ependymoma. The level of evidence for each feature is variable.

1.3.3.1 Tumour grade

The World Health Organisation (WHO) classification describes four histological types of ependymoma (Louis et al., 2007):

- Myxopapillary ependymoma (WHO grade I);
- Subependymoma (WHO grade I);
- Ependymoma (WHO grade II);
- Anaplastic ependymoma (WHO grade III).

Morphologically, ependymomas vary from highly differentiated tumours with densely cellular areas and perivascular rosettes, to anaplastic types with areas of necrosis, chaotic vascularity and calcification (Kilday et al., 2009; Louis et al., 2007). The level of heterogeneity seen in assessment of tumour histology is summarised in the grading scheme proposed by Ellison (Table 1-5).

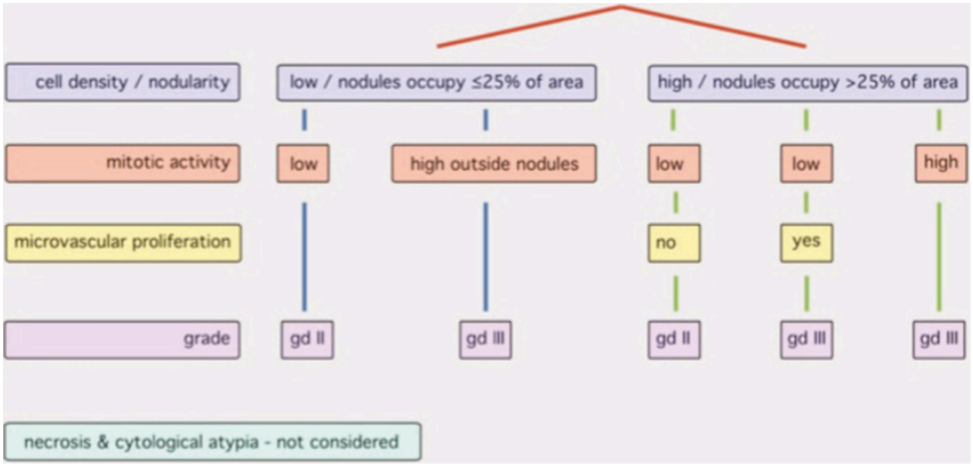


Table 1-5: New criteria for histological assessment based on the tumour samples collected from patients involved in 4 European ependymoma trials. Reproduced from Ellison et. al. 2011.

The results of multiple studies of tumour grade as a marker of prognostic significance are mixed. Some studies provide no evidence that tumour grade at presentation correlates with either OS or EFS (Agaoglu et al., 2005; Bouffet et al., 1998; Ellison et al., 2011; Li et al., 2015; Paulino et al., 2002; Ridley et al., 2008; Robertson et al., 1998; Snider et al., 2017; Tabori et al., 2008). This may be due to small study size. However others, including a recent large prospective study, identified significant differences in outcome for children with different

tumour grades (Massimino et al., 2016; Tihan et al., 2008). A further explanation for the conflicting evidence may be that accurate, reproducible grading is challenging (Ellison et al., 2011).

1.3.3.2 Resection status

It has been consistently reported that GTR is the most important and reproducible predictor of long-term survival (Agaoglu et al., 2005; Aizer et al., 2013; Amirian et al., 2012; Cage et al., 2013; Godfraind et al., 2012; Grill et al., 2001; Jaing et al., 2004; Mendrzyk et al., 2006; Perilongo et al., 1997; Pollack et al., 1995; Robertson et al., 1998; Snider et al., 2017; Tabori et al., 2008). However, achieving this can be challenging (Bouffet et al., 1998; Perilongo et al., 1997). In some cases STR may need to be followed up with second look surgery to obtain GTR (Foreman et al., 1997; Godfraind et al., 2012; Vinchon et al., 2005). In one example the GTR rate was improved to 80% using this approach (Godfraind et al., 2012).

1.3.3.3 Location

Paediatric ependymomas make up 6-12% of all childhood intracranial tumours and 30% of spinal tumours. 90% of cases are intracranial and around 70% of these arise in the PF (Jaing et al., 2004; Vinchon et al., 2005). Spinal tumours have a better prognosis (McGuire et al., 2009). There is conflicting evidence about the implications of intracranial tumour location on outcome. Some studies report no difference between PF and ST tumours (Agaoglu et al., 2005; Jaing et al., 2004; Paulino et al., 2002; Perilongo et al., 1997; Pollack et al., 1995; Robertson et al., 1998; Snider et al., 2017). Some report that ST tumours have worse survival outcomes (Cage et al., 2013; Mansur et al., 2005; Nambirajan et al., 2014) and some report that PF tumours have worse outcomes (Grill et al., 2001; Ridley et al., 2008; Venkatramani et al., 2013). This data is difficult to interpret in view of the heterogeneity of the study designs and lack of knowledge about the molecular subgroup of the tumours included in the studies.

Figarella-Branger and colleagues reported that PF tumours located laterally had worse outcomes than those located centrally (Figarella-Branger et al., 2000). This might account for some of the discrepancies in results for outcomes of PF versus ST tumours. Consequently various studies have looked at further sub-classification of tumours in one anatomical location; in particular the posterior fossa (Hoffman et al., 2014a; Wani et al., 2012; Witt et al., 2011). DNA methylation classification, which will be introduced later, supports the concept of midline and lateral PF tumours being different entities (Pajtler et al., 2015).

1.3.3.4 Age at diagnosis

Young children have the poorest outcomes (Agaoglu et al., 2005; Amirian et al., 2012; Bouffet et al., 1998; Figarella-Branger et al., 2000; Jaing et al., 2004; McGuire et al., 2009; Perilongo et al., 1997; Snider et al., 2017; Tabori et al., 2008). It has taken time for age to emerge as a clear prognostic factor because studies have used different age bands and included adult patients in their case series (Bouffet et al., 1998). Additionally, younger children are more difficult to clinically assess and there is appropriate reluctance to use radiotherapy on the developing brain (Agaoglu et al., 2005; Gilbertson et al., 2002; Jaing et al., 2004; Massimino et al., 2004). Despite this, it would be erroneous to conclude that the differing therapeutic approaches due to age are the only cause for poorer outcomes, as it is now emerging that younger children may have more biologically aggressive tumours, irrespective of treatment (Carter et al., 2002; Hirose et al., 2001; Pajtler et al., 2015; Ramaswamy et al., 2016; Wani et al., 2012; Witt et al., 2011).

1.3.3.5 Genomic profiles

Ependymoma has been described as 'genetically bland' and 'enigmatic' (Kilday et al., 2009; Mack et al., 2013; Robertson et al., 1998). Whilst there are multiple examples of DNA copy number change (Bouffet et al., 1998; Dyer et al., 2002; Hirose et al., 2001; Kilday et al., 2012; Witt et al., 2011b) there are few genetic mutations (Mack et al., 2013). Reasons cited for this include:

(1) heterogeneous nature of the tumour resulting in missing important and focal genetic changes;
(2) presence of epigenetic, but not genetic, drivers;
(3) insufficient depth of analysis (Mack & Taylor, 2009).

Despite this, a number of genomic abnormalities have been linked with ependymoma and these present an area for investigation.

Associations have been described between chromosome 1q gain, disease progression and higher tumour grade. This copy number change has gained a great deal of prominence regarding its association with tumour prognosis and has been repeatedly associated with outcome in multiple studies (Carter et al., 2002; Dyer et al., 2002; Godfraind et al., 2012; Hirose et al., 2001; Kilday et al., 2012; Mendrzyk et al., 2006; Rousseau et al., 2010).

Loss of chromosome 22 has been associated with sporadic ependymoma and is present in 17-75% of tumours (Carter et al., 2002; Jeuken et al., 2002; Reardon

et al., 1999; Zheng et al., 2000). It was initially thought that the *NF2* gene was the candidate gene associated with this chromosomal loss, however it emerged that this was only the case in spinal ependymoma (Mack & Taylor, 2009).

Ependymomas in different CNS locations appeared to have different chromosomal constitutions in some studies. Spinal tumours were associated with gains of chromosomes 7, 9, 11, 18 and 20 and loss of 1, 2 and 10 whilst intracranial ependymomas showed gain of chromosome 1q and loss of 22, 3, 9p and 13q (Mack & Taylor, 2009; Rousseau et al., 2010).

In 2002 Dyer and colleagues analysed 42 primary and 11 recurrent ependymomas by array-CGH. They clustered tumours into three groups based on levels of genomic imbalance:

(1) Numerical: 13 or more imbalances (12%);
(2) Structural: 1-6 imbalances and high ratio of partial to whole chromosome imbalances (45%);
(3) Balanced (43%).

When these groups were analysed in relation to their clinical behaviour, children under three years had mainly balanced profiles whilst older children fell into numerical or structural groups. Multivariate analysis demonstrated that children with tumours in the structural group had the worst outcomes when controlled for tumour location, histology, extent of resection, adjuvant therapy and age at diagnosis. The study also found that tumours exhibiting 1q gain had a trend towards poorer five-year OS (15% versus 50%). In analysing their small cohort of recurrent tumours, 10 out of 11 of the specimens showed a structural profile. Gain of 1q was identified in eight cases. The authors concluded that tumours in the structural group are clinically more aggressive. A larger study of 292 cases confirmed similar genomic subgroups (Korshunov et al., 2010).

One study reported that the presence of chromosomal imbalance was more likely to be associated with recurrent disease, in particular gains of chromosome 9q33 and 9q34 (Puget et al., 2009).

Researchers in the USA demonstrated that copy number aberrations (CNAs) were affecting oncogenes and tumour suppressor genes (TSGs) resulting in upregulated and downregulated gene expression. CNAs were characterised for a set of 204 ependymomas from which the authors identified 84 and 39 candidate

oncogenes and TSGs. They transduced the genes in radial-glial cells which were then injected in pools into immunocompromised mice. PCR analysis from the subsequent mouse tumours determined which oncogenes or TSGs had driven tumour formation. These genes were again transduced, one per pool, into cells of origin and injected into 209 different mice. 53% developed tumours histologically consistent with ependymoma, validating ten tumour suppressor genes and eight oncogenes (Johnson et al., 2010; Mohankumar et al., 2015). The genes that emerged from this analysis are outlined in Table 1-6 and Table 1-7.

Tumour Suppressor Gene	**Function**
ALDH3A1	Aldehyde dehydrogenase activity
ACTR1A	Regulates vesicle trafficking
SNX6	Regulates vesicle trafficking
ULK2	Regulates vesicle trafficking
PCMT1	Protein repair/degradation
DNA2	DNA modification and repair
SUFU	Negative regulator in Hedgehog signalling pathway
STAG1	DNA modification and repair
TET1	DNA modification and repair
ST13	Protein binding

Table 1-6: Tumour suppressor genes from Mohankumar et. al. 2015. Function from www.genecards.org.

Oncogene	**Function**
ZNF688	Links chromatin relaxation state to DNA repair
BCL7C	Member of the SWI/SNF chromatin regulatory complex
RAB3A	Late-stage vesicle trafficking and exocytosis in neuronal cells
PRDX2	Regulates oxidation induced apoptosis
RTBDN	Belongs to folate receptor family
AKT2	Mediator of cell survival
TMEM129	Mediates HLA class I protein degradation
MRPS17	Mitochondrial ribosomal protein

Table 1-7: Oncogenes from Mohankumar et. al. 2015. Function from www.genecards.org.

1.3.3.6 Gene expression profiles

Through the use of gene expression arrays, numerous authors have correlated transcriptional profiles of ependymoma with clinicopathological features, including age, grade, gender and tumour location. Researchers have also attempted to look for associations between transcriptional profiles and clinical outcomes. Specific gene expression profiles have been linked most frequently to location within the

CNS with large numbers of differentially expressed genes between spinal, posterior fossa and supratentorial locations. However, associations have also been suggested with tumour grade (Korshunov et al., 2003; Palm et al., 2009), gender (Pajtler et al., 2015) and patient age (Korshunov et al., 2003; Pajtler et al., 2015; Wani et al., 2012; Witt et al., 2011) (Table 1-8).

Initial studies undertook supervised approaches to these investigations, performing differential expression analysis on groups of tumours with particular clinical characteristics (Korshunov et al., 2003; Lukashova-v.Zangen I. et al., 2007; Modena et al., 2006; Taylor et al., 2005). One of the drawbacks of this approach was the tendency to be able to find expression profiles based on predefined ideas about the disease, risking biased conclusions. More recently, studies have undertaken unsupervised clustering approaches. This has resulted in the delineation of numerous subgroups which were annotated clinically to identify numerous associations with specific locations, epidemiological features and clinical outcomes. Some of the proposed subgroups were complex and not validated in the literature (Johnson et al., 2010).

A key development was the description of two PF based gene expression subgroups, initially named A and B. Group A was found to be associated with a particularly aggressive phenotype with poor outcomes, a balanced genome (apart from isolated 1q gain) and younger age. Group B was associated with better outcomes, a more unbalanced genome, midline location, and an older age profile (Wani et al., 2012; Witt et al., 2011). Biomarkers were identified which were representative of these two gene expression groups: *Laminin-alpha-2* (*LAMA2)* selected to represent group A, and *Neural Epidermal Growth Factor Like 2* (*NELL2)* selected to represent group B. Each group was also associated with specific gene ontology terms, including wound healing and the inflammatory response for group A, and terms related to cilia and motility for group B. Interestingly, given that A and B gene expression groups were found to have specific associations with clinical features, it is unsurprising that authors prior to 2011 had also found clinical associations with gene expression. However, the definition of A and B subgroups allowed the knowledge of the associations to be linked with well-defined molecular phenotypes.

Study	Patients (Location)	Approach	Age	Location	Grade	Relapse/ Outcome	Gender	Molecular Subgroup
(Korshunov et al., 2003)	39 (All)	Supervised	Yes	Yes	Yes	No	No	No
(Taylor et al., 2005)	32 (All)	Supervised	No	Yes	No	No	No	No
(Modena et al., 2006)	24 (IC)	Supervised	No	Yes	No	Yes	No	No
(Lukashova-v.Zangen I. et al., 2007)	47 (All)	Supervised	No	Yes	No	Yes	No	No
(Palm et al., 2009)	34 (All)	Supervised and unsupervised	No	Yes	Yes	No	No	No
(Donson et al., 2009)	19 (IC)	Supervised	No	No	No	Yes	No	No
(Johnson et al., 2010)	83 (All)	Unsupervised	No	Yes	No	No	No	No
(Witt et al., 2011)	177 (PF)	Unsupervised	Yes	N/A	No	Yes	No	PFA, PFB
(Wani et al., 2012)	67 (PF)	Unsupervised	Yes	N/A	No	Yes	No	PFA, PFB
(Hoffman et al., 2014a)	44 (All)	Unsupervised	No	Yes	No	Yes	No	PFA, PFB
(Pajtler et al., 2015)	209 (All)	Unsupervised	Yes	Yes	No	Yes	Yes	Multiple

Table 1-8: Summary of gene expression array studies investigating transcriptomic profiles of ependymoma. Location refers to spinal and intracranial (All), intracranial only (IC) or posterior fossa only (PF). Approach reflects whether the studies tested for differences between predefined clinical factors (supervised) or used hierarchical clustering to generate clusters before comparing the clinical features of each cluster (unsupervised). Columns relating to age, location, grade, relapse/outcome and gender indicate whether the study in question made an association between gene expression profile and each characteristic. The molecular subgroup column indicates whether accepted gene expression subgroups were defined or discussed in the study.

1.3.3.7 DNA methylation profiles

DNA methylation is an epigenetic mark which can alter gene expression and biological function. DNA methylation patterns vary between tissue types, tumour types and age. Given their relative stability, DNA methylation patterns can be used as part of the diagnostic process. This approach is now being used extensively in childhood brain tumours (Pajtler et al., 2015; Schwalbe et al., 2017; Sturm et al., 2016).

In 2014 Mack and colleagues reported a study of 47 PF ependymomas and found that whilst these tumours were unremarkable when assessed by whole-genome and whole-exome sequencing, they demonstrated heterogeneous DNA methylation profiles. They reported that the EPN_PFA and EPN_PFB groups were replicated when clustered by DNA methylation profile. EPN_PFA ependymomas demonstrated a CpG island methylator or 'CIMP' phenotype (Mack et al., 2014). The authors hypothesised that the reason for the aggressive behaviour of the EPN_PFA tumours could be due to epigenetic silencing of genes promoting cellular differentiation. They found that some of the hypermethylated genes were those silenced by the Polycomb Repressive Complex 2 (*PRC2*) in embryonic stem cells. *EZH2* expression, a component of the PRC complex, has subsequently been associated with poorer five-year OS in childhood ependymoma (Li et al., 2015).

An international collaboration led by researchers in Heidelberg has since identified nine groupings based on DNA methylation status. The DNA methylation profiles of 500 adult and paediatric ependymal tumours were established with Infinium HumanMethylation450 BeadChip Arrays (Illumina). Results were clustered to reveal three subgroups for each of the three tumour locations (Pajtler et al., 2015) (Figure 1-2).

There were no children in the SP-SE, PF-SE and ST-SE groups. Children under the age of 18 years constituted 99% of the EPN_PFA group, 92% of the EPN_YAP group and 76% of the EPN_RELA group but only 19% of the EPN_PFB group. This is important given that different transcriptional profiles have also been attributed to different age groups (Korshunov et al., 2003). The DNA methylation defined groups also correlated perfectly with the transcriptionally defined groups. EPN_PFA and EPN_PFB were reflective of the gene expression patterns of group A and group B described in section 1.3.3.6.

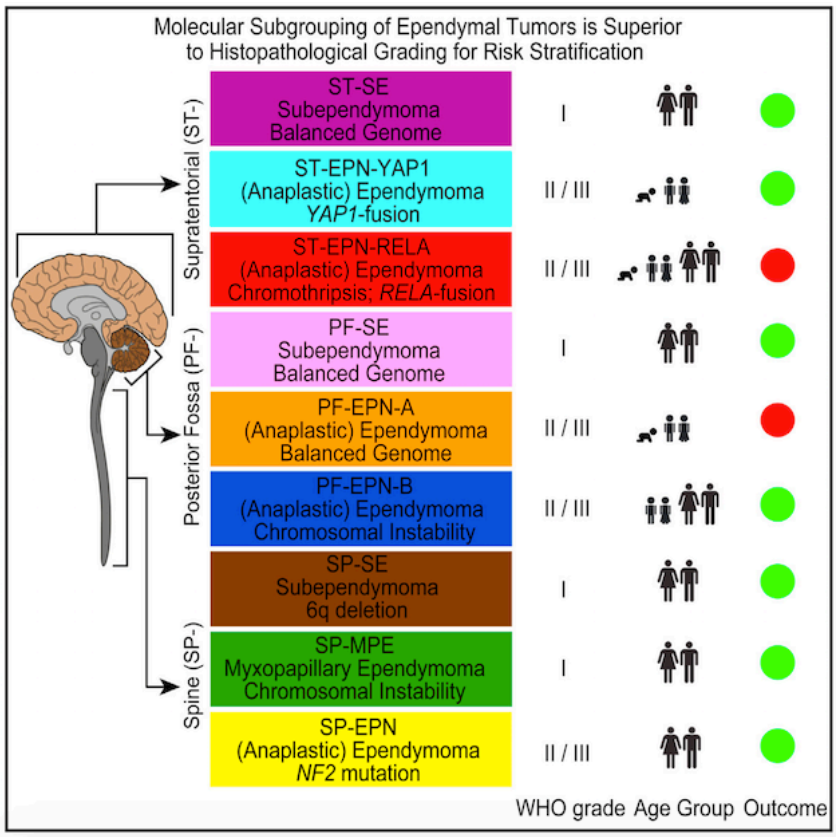


Figure 1-2: The nine proposed subgroups of ependymoma. Note that this classification includes both adults and children. Groups most relevant to paediatric cohorts include ST-EPN-YAP1, ST-EPN-RELA, PF-EPN-A and PF-EPN-B. Classification with poor outcomes are marked with a red dot. Modified with permission from Pajtler et al. 2015.

The DNA methylation defined subgroups were associated with disparate outcomes. Children with EPN_PFA and EPN_RELA tumours had poorer outcomes in contrast to those with EPN_PFB and EPN_YAP tumours. Given that EPN_PFB is more commonly seen in older children and adults, the finding that younger children have poorer outcomes may actually be related to the molecular composition of the tumours experienced by younger children.

Pajtler and colleagues also reported that no tumour changed subgroup at recurrence, however this contradicts gene expression findings (Hoffman et al., 2014a). One explanation for this discrepancy is that DNA methylation profiles are set very early in foetal development and are then relatively permanent. However, other genetic and epigenetic changes could still occur that affect gene expression profiles.

Evidence is emerging that the EPN_PFA tumours can be further subdivided into nine groups, with two major subgroups; PFA-1 and PFA-2. This data has been presented internationally and has been submitted for publication (Pajtler et al., 2017).

1.3.3.8 Fusion genes

Fusion genes form from two genes which were previously spatially separated. They can form through localised genomic rearrangements (deletions, inversions or translocations) or by more widespread rearrangements resulting from chromothripsis; a mechanism first described in the context of chronic lymphocytic leukaemia (Forment et al., 2012). In the process of chromothripsis it is thought that chromosomes are shattered by an extreme insult, followed by incorrect rearrangement resulting in deletions, inversions and translocations. It has been suggested that cancerous cells which develop following chromothripsis are more susceptible to novel targeted cancer therapies as a result of their genomic instability (Forment et al., 2012).

Fusion genes may contribute to ST ependymoma development (Parker et al., 2014; Pietsch et al., 2014). Parker and colleagues performed whole genome and RNA sequencing of 41 and 77 ependymomas respectively from both ST and PF locations, identifying and validating 27 novel fusion genes involving chromosome 11q. These fusions occurred exclusively in ST tumours. Of particular interest were fusions with the *RELA* gene, an important component of NF-kB signalling. A form of the *C11orf95-RelA* fusion was found in 70% of paediatric ST ependymomas, with seven different molecular variants identified.

The *C11orf95-RelA* fusion gene resulted in the upregulation of the NF-kB cellular signalling pathway by the transportation of $RELA^{FUS1}$ protein into the nucleus. Some upregulation is also seen with wild type RELA protein ($RELA^{WT}$) which is probably involved in normal cellular homeostasis. $RELA^{FUS1}$ seemed to induce the NF-kB pathway with more potency and resulted in increased ependymoma formation in mouse models (Parker et al., 2014). Mouse tumours were also induced by another gene fusion, *C11orf95-YAP1*, indicating that multiple fusions are potentially capable of ependymoma oncogenesis.

The work on DNA methylation profiles by Pajtler and colleagues subsequently identified that specific DNA methylation based subgroups were associated with *RELA* and *YAP1* fusions (EPN_RELA and EPN_YAP), associating the presence of different fusion genes with clinical outcomes, epidemiological characteristics and

gene expression patterns. Pajtler and colleagues also reported that whilst RELA fusions were associated with chromothripsis and a disordered genomic profile, YAP1 fusions were associated with balanced profiles but specific copy number aberrations around the YAP1 locus (Pajtler et al., 2015).

1.3.4 A cancer stem cell origin for ependymoma

Cancer stem cells (CSCs) form a small proportion of the overall cancer burden and have the capacity for self-renewal and tumour maintenance (Clarke et al., 2006). Their existence was first indicated in 1937 by a study demonstrating that large numbers of cells needed to be injected into mice to initiate cancer thus inferring that not every cancer cell can propagate malignancy (Furth and Kahn, 1937). It has been theorised that this subpopulation of tumour initiating and propagating cells are relatively refractory to conventional therapies. As a result, their presence may make for tumours that are difficult to eradicate with conventional approaches (Clarke et al., 2006).

CSCs may play a role in cancer recurrence (Esmatabadi et al., 2016). One theory is that, in view of their general resistance to therapy, they are not completely eradicated during initial treatment. Consequently, a proportion lie dormant and recapitulate the tumour when environmental factors permit. Importantly, these cells may lie dormant for many years, resulting in recurrence of the same tumour long after the patient and physician believe it has been eradicated. This suggests that recurrence might develop from asymptomatic residual disease (Aguirre-Ghiso, 2007). It is not clear what maintains CSCs in their dormant state, but one review suggests considering mechanisms antagonising the expansion of the dividing tumour cell population and mechanisms resulting in cessation of tumour growth. The review describes three types of dormancy;

(1) Cellular dormancy: results in dormant cells having entered a G0-G1 arrest, potentially allowing them to evade immune recognition;
(2) Angiogenic dormancy: Maintained by the balance of factors promoting and opposing angiogenesis;
(3) Immunosurveillance: Components of the immune system recognise and suppress proliferating cancer cells (Aguirre-Ghiso, 2007).

Not only does this indicate a role for CSCs in recurrence, but also the importance of an appropriate microenvironment for tumours to develop from a dormant state (Aguirre-Ghiso, 2007; Jandial et al., 2008).

Cancer stem cells (CSCs) can be identified by the expression of cluster of differentiation molecule 133 (CD133) (Li, 2013). Approximately 1% of CD133

positive ependymoma cells display a radial glia-like phenotype, compared to none of the CD133 negative cells. It has been demonstrated that these CD133 positive cells with a radial-glia like phenotype form neurospheres under in vitro conditions, favourable to stem cell growth. Consequently, CD133 expressing radial-glia like neural stem cells (NSCs) have been proposed as a CSC origin for ependymoma. Not only this, but by also using gene expression profiling, Taylor showed that location specific ependymoma gene expression profiles correlated with regionally specified radial-glia-like cells, suggesting that ependymomas in different locations may have distinct origins (Taylor et al. 2005).

It has also been postulated that NSCs arising from different parts of the CNS give rise to biologically distinct tumours. In support of this hypothesis, modified NSCs from different parts of the CNS produced location specific ependymomas in mouse models (Johnson et al., 2010). The studies by Taylor and Johnson marked a shift to considering the developmental biology underpinning ependymoma and help to provide an explanation for a mechanism by which ependymoma may recur.

1.4 Ependymoma recurrence

1.4.1 Epidemiology

Paediatric ependymoma is an aggressive disease in which relapse is common, recurrent, and often fatal. Ependymoma tends to recur at its original site, with metastasis in around 25% of cases (Antony et al., 2014; Jaing et al., 2004; Messahel et al., 2009). In a study of 108 children with recurrence in the United Kingdom five-year OS was 24% for children under three years and 27% for older children (Messahel et al., 2009). Earlier studies have reported even worse outcomes, with five-year OS of 12.3% (Jaing et al., 2004), three-year OS of 0% (Pollack et al., 1995) and two-year OS of 39% (Goldwein et al., 1990). Available studies of recurrence are based on historical patient cohorts. However, given that cancer survival outcomes have improved over the decades (Cancer Research UK, 2015a), it is possible that in more contemporaneous cohorts outcomes are better.

Despite the development of DNA methylation classification, conferring dramatically different prognoses between ependymoma subgroups (Pajtler et al., 2015; Ramaswamy et al., 2016), there has been no detailed study of the clinical features of recurrent disease supported by DNA methylation profiling.

1.4.2 Prognostic factors

Independent prognostic factors impacting OS at recurrence include: extent of repeat surgery; use of radiotherapy; and the presence of metastatic disease (Bouffet et al., 2012; Merchant et al., 2008; Messahel et al., 2009; Tsang et al., 2018; Zacharoulis et al., 2010). Despite this, approaches to the treatment of recurrent disease have not been standardised. Whilst the treatments mentioned have been shown to improve short term prognosis, the long term prognosis of recurrence remains poor for all patients (Messahel et al., 2009; Zacharoulis et al., 2010).

Achieving a GTR can be even more challenging than at first presentation due to tumour infiltration, with resultant high morbidity (Vinchon et. al., 2005). GTR may also be more difficult to achieve in children with multi-site disease.

Children with metastatic disease at recurrence have demonstrated worse outcomes in multivariate analyses in more than one study. In one case median OS was 13 months in the metastatic patients compared to 30 months in those with local disease (Messahel et al., 2009). In another study median OS was not reported, but hazard ratios for patients with metastatic recurrence were significantly increased (Zacharoulis et al., 2010).

Re-irradiation has been suggested to significantly improve survival duration, albeit with the risk of increased neurotoxicity (Bouffet et al., 2012; Lobón et al., 2016; Merchant et al., 2008; Tsang et al., 2018), although this is not a universal finding (Zacharoulis et al., 2010). However, children who received radiotherapy during treatment of their primary tumour had worse OS at recurrence than those who did not (Messahel et al., 2009).

There was no improvement with chemotherapy in recurrent disease in one study, whilst another associated chemotherapy with poorer outcomes after first relapse (Messahel et al., 2009; Zacharoulis et al., 2010).

On the basis of previous research into recurrent ependymoma, European guidelines suggest that therapy at relapse should include repeat surgery and/or radiotherapy, consideration of chemotherapy and consideration of inclusion in a clinical trial (Rudà et al., 2017).

1.4.3 Future therapies

Novel therapies have been investigated in phase I and II clinical trials in patients with recurrence with minimal success. Possible reasons for this include small study numbers, inclusion of multiple types of recurrent brain tumour, and a lack of understanding about the underlying biology of recurrent disease. Novel agents tested include Perifosine (Phase I) (Becher et al., 2017), 5-Fluorouracil (Phase I) (Wright et al., 2015), Erlotinib (Phase II) (Jakacki et al., 2016), Sunitinib (Phase II) (Wetmore et al., 2016), Bevacizumab and Lapatinib (Phase II) (DeWire et al., 2015), Bevacizumab and Irinotecan (Phase II) (Gururangan et al., 2012), interferon (IFN) beta (Phase I-II) (Allen et al., 1991) and Paclitaxel (Phase II) (Hurwitz et al., 2001). Unfortunately, none of these phase II studies showed promise in recurrent ependymoma.

Immunotherapy is not an evidenced based treatment in recurrent ependymoma although it has been a therapy of interest in many other malignancies, including melanoma (Hodi et al., 2010). However, a search of www.clinicaltrials.gov for 'ependymoma' and 'immunotherapy' highlighted eight trials investigating the role of various immunotherapy techniques in brain tumours including ependymoma. In order to assess whether immunotherapeutic interventions stand any chance of being effective, a better understanding of the biological basis of ependymoma recurrence is essential.

1.4.4 Studies of ependymoma biology at recurrence

Minimising the risk of recurrence is a crucial consideration in the management of primary disease. Given that recurrence is so common, a better understanding of its biology is a research priority.

The investigation of recurrent paediatric ependymoma has been hampered by the rarity of the disease and the difficulty in obtaining adequate numbers of matched primary and recurrent pairs. Eight studies investigating tumour biology at recurrence have been published (Table 1-9). These were largely immunohistochemistry based, investigating specific molecular markers (Ridley et al., 2008; Tabori et al., 2008) or comparative genomic hybridisation (CGH), investigating genomic imbalances (Dyer et al., 2002; Puget et al., 2009). Only two studies used array base gene expression profiling. Both were performed before current, DNA methylation defined, molecular subgroups were identified (Mack and Taylor, 2017).

Peyre et al. (2010), investigated 17 matched pairs from all intracranial locations, implicating kinetochore proteins and metallothioneins in tumour behaviour at relapse. Peyre also identified immune related genes in a recurrence signature, with downregulation of chemokines *CXCL5* and *CX3CL1* amongst others. Hoffman et. al. (2014), investigated only PF ependymomas. They suggested that EPN_PFA and EPN_PFB subgroups displayed different patterns of immune response at recurrence. However, subgroups were only defined using gene expression profiling, which recent research has suggested could be misleading (Pajtler et al., 2017). Follow-up suggested that primary tumours with EPN_PFA phenotypes display evidence of immune suppression (Hoffman et al., 2014a).

In addition to studies investigating biology at recurrence, a study by Donson and colleagues in 2009 investigated the differences between primary ependymomas that subsequently relapsed and those that did not. Based on gene expression profiling and immunohistochemistry, they suggested that non-recurrent primaries were associated with enrichment of immune related genes, and that tumours with immune gene enrichment demonstrated a longer time to progression when they did recur. They also correlated increased tumour infiltration with $CD4^+$ T-cells with improved outcomes. Unfortunately this study was based on a small sample size of nine non-recurrent tumours and ten recurrent tumours, making robust conclusions difficult (Donson et al., 2009). There is no evidence that either this work, or the Hoffman paper from 2014, has been validated by other authors or investigated in more detail in the published literature.

In view of the association of relapse with the immune system in two published articles, a search of PubMed for 'ependymoma' and 'immune' across all time periods was undertaken. 46 results were generated, only three of which were specific studies of the immune response in ependymoma, all from the same group (Donson et al., 2009; Griesinger et al., 2017; Hoffman et al., 2014a). The 2009 and 2014 studies have already been described. The 2017 study investigated the upregulation of NF-kB signalling being associated with the epigenetic silencing of the gene *LDOC1* in EPN_PFA tumours.

Following the completion of the literature search, one further study was published relating to the immune environment in ependymoma, again by the same research group (Witt et al., 2018). This study investigated levels of *PD-L1* expression in EPN_RELA disease. The authors suggested that EPN_RELA tumours expressed more PD-L1 and had higher levels of CD4+ and CD8+ infiltration than other

tumour subtypes and concluded that EPN_RELA tumours should be considered for checkpoint inhibitor clinical trial inclusion. Interestingly, the authors did not use DNA methylation classifications to define the subgroups, instead using transcriptomic and protein based approaches. Comparisons between this and other studies where DNA methylation was used for classification should therefore be made with this in mind.

Overall, the lack of published literature on the immune system in ependymoma is surprisingly limited and represents a gap that needs to be addressed.

Study	Number of tumours			Technique	Key findings	Comments
	Primary	Recurrent	Matched Pairs			
(Dyer et al., 2002)	42	11	7	Array CGH	Subtyped ependymoma by chromosomal aberrations: (1) Structural; (2) Numerical;(3) Balanced. At recurrence 91% of tumours were unbalanced.	Few recurrent tumours and few pairs.
(Sowar et. al., 2006)	10	3	1	Gene expression array	Minimum subset required to classify all tumours by recurrence status included 3 genes; PLEK, NF-kB2 and LOC374491.	Small sample size.
(Ridley et al., 2008)	74	23	17	IHC, TRAP	Low nucleolin expression associated with better 5-year EFS. Telomerase reactivation and maintenance appear necessary for tumour progression.	IHC study. TRAP on small cohort.
(Tabori et al., 2008)	83	50	31	IHC, TRAP	hTERT correlated with proliferation markers (MIB-1 and mitotic indices and overall tumour grade). No correlation between telomere length and survival.	IHC study. TRAP on small cohort.
(Puget et al., 2009)	33	26	15	Array CGH	Subtypes similar to Dyer 2002. Greater incidence at recurrence of gain of 9q34, 1q and loss of 6q. In the paired patients 3 profiles remained balanced, one showed loss of chromosome 22, 6 showed new abnormalities and 5 showed fewer imbalances. Gain on 9q33 and 9q34 were associated with recurrence (p=0.003 and 0.009), age over 3 (p=0.019 for 9q34) and PF location (p=0.002 and 0.015)	Small overall sample size and few pairs.

Study	Number of tumours			Technique	Key findings	Comments
	Primary	Recurrent	Matched Pairs			
(Peyre et al., 2010)	17	27	17	Array CGH, gene expression array	Kinetochore proteins and genes involved in neural development (Wnt, CD133 and Notch) upregulated, metallothioneins downregulated. Metallothionein expression had epigenetic control and was restored by histone deacetylase inhibitors. No change in copy number between presentation and recurrence.	Small sample size.
(Hoffman et al., 2014a)	44	14	14	Gene expression array, SNP microarray	Confirmed previously described PF subgroups (Witt et al., 2011) and defined 2 further subgroups (PFA1, PFA2, PFB1, PFB2) based on gene ontologies. 4 tumours changed group at recurrence.	Small sample size.
(Pajtler et al., 2015)	452	48	45	450K Methylation array, gene expression array	Tumours divided into 9 DNA methylation based subgroups. Matched recurrences clustered closely with primary tumours suggesting DNA methylation classification can be applied to recurrent as well as primary disease – tumours did not change group.	Mixed adults and children.

Table 1-9: Summary of studies including matched primary and recurrent paediatric ependymoma cases to date.

PFA: Posterior Fossa Group A, PFB: Posterior Fossa Group B, hTERT: human telomerase reverse transcriptase, IHC: Immunohistochemistry, TRAP: Telomeric Repeat Amplification Protocol, CGH: Comparative Genomic Hybridisation, PF: Posterior Fossa, EFS: Event Free Survival.

1.5 Cancer immunity and recurrence

Given the association of ependymoma, particularly in relation to recurrence, with the immune system, this section considers the role of the immune system in cancer progression in general.

1.5.1 The tumour microenvironment

The National Cancer Institute defines the tumour microenvironment as:

"The normal cells, molecules, and blood vessels that surround and feed a tumour cell" (NCI Dictionary of Cancer Terms).

The tumour microenvironment contains elements which may support or antagonise cancer growth. These include extracellular matrix, immune cells and fibroblasts. The microenvironment is shaped by cancer cells through interactions with the host (Whiteside, 2008). One way in which cancer cells can interact through the microenvironment is via cell signalling, in particular by modulating cytokine expression. Cytokines are small proteins, which are secreted by and act upon immune, tumour and other microenvironment cells. They are able to alter cellular functions such as taxis (movement), cellular proliferation and growth.

Cancers can be considered as 'wounds that do not heal' (Dvorak, 1986). Pro-inflammatory signals, precipitating the wound healing response, do not originate from immune cells as originally predicted, but from cells undergoing apoptotic cell death. Consequently, one postulated microenvironmental mechanism for cancer recurrence is the 'Phoenix rising' pathway, in which apoptotic tumour cell death promotes wound healing and tissue regeneration pathways (Esmatabadi et al., 2016; Li et al., 2010). It is hypothesised that when cancer cells undergo apoptosis as a result of therapy, pro-inflammatory mediators stimulate the surviving cancer stem cells. This may produce a mechanism by which the cancer can recur or progress (Esmatabadi et al., 2016).

1.5.2 The innate and adaptive immune response

The innate immune response is involved in the immediate recognition of pathogens in a non-specific fashion. It includes neutrophils, macrophages and natural killer (NK) cells and, in the brain, microglia. It is based on the recognition of chemical signals produced by pathogens and has no capacity for immunological memory (Reeves and Todd, 2004).

The adaptive immune response is based upon the actions of effector T- and B-cells which are able to develop an immunological memory and therefore recognise specific antigens. Immunological memory is developed through exposure of T- and B- cells to target antigens, which are usually proteins, by antigen presenting cells (Reeves and Todd, 2004). A summary of the key effector cells involved in innate and adaptive immunity is found in Table 1-10.

Cell Type	**Function**	**Markers**	**Immune Response**
Neutrophil	Phagocytic cell. Immediately migrate to affected tissues. Produce cytokines and prostaglandins. Internalise and kill pathogens.	CD16, CD66b	Innate
Macrophage	Phagocytic cell. Accumulate at sites of infection. Secrete products including lysozyme, cytokines and complement proteins. Internalise and kill pathogens.	CD14, CD33	Innate
Natural killer	Kill target cells without antigen-specific activation. Killing functions are enhanced by the presence of interferons.	CD56	Innate
Microglia	Function as the macrophages of the brain.	CD68	Innate
T lymphocyte	Divided into helper T-cells (CD4+) and cytotoxic/effector T-cells (CD8+). Require antigen specific activation. CD4+ cells activate other cells to recognise antigen. CD8+ cells execute the final steps in cell killing by secreting perforins and granzymes onto the surface of target cells.	CD3, CD4, CD8	Adaptive
B lymphocyte	Secrete antibody to opsonise the surface of target cells. Mediate humoral immunity through body fluids.	CD19, CD20	Adaptive

Table 1-10: Summary of key immune cells in innate and adaptive immunity. Modified from (Reeves and Todd, 2004). CD Marker identification modified from (BDBiosciences, 2016). All cells also express the leucocyte common antigen (CD45) which serves as a marker of inflammation.

Both innate and adaptive immune responses have been implicated in cancer. Innate responses are mediated through NK cells and adaptive responses through lymphocytes, such as T- and B-cells (Woo et al., 2015). Immune responses can be pro- or anti-cancer, and mechanisms by which cancer can subvert the immune system and lead to tumour progression have been widely described (Kim and Chen, 2016; Spranger, 2016; Zitvogel et al., 2006).

1.5.3 The immune environment

The immune system has long been postulated to play a role in cancer; physicians were attempting to develop cancer vaccines by inoculation with malignant tissue as early as 1777 (Ichim et al., 2005). There is emerging evidence that loss of immune control of a malignancy may contribute to tumour recurrence.

The immune system is involved in discriminating 'self' from 'non-self' (Blair and Cook, 2008) in the process of immunosurveillance (Zitvogel et al., 2006). Whilst cancer originates from 'self' molecules, subtle changes such as the development of new mutations forming neoantigens or altered DNA methylation patterns, induce the host immune system to target it as 'non-self', enabling tumour destruction (Rooney et al., 2015). Such targets are known as tumour antigens (Table 1-11). However, in some cases the microenvironment induced by the cancer may harness the immune response and either evade, or modify, it to become pro-tumour (Raman et al., 2007; Spranger, 2016). Additionally, immune cell infiltration into the tumour microenvironment may be immune suppressive, aiding in tumour survival (Whiteside, 2008).

Molecule	**Description**
Neoantigens	An antigen arising in a tumour due to a new genetic mutation
Oncoviral proteins	Proteins produced following viral induction of a malignancy.
Glycolipids	Human tumours often express high levels of surface glycoproteins and glycolipids which may be targeted by the immune system.
Cancer-Testis antigens (CTAs)	Preferentially expressed in cells normally sheltered from the immune system, including testes and placenta. Some tumours re-express these markers, inducing an immune response. A database of CTAs was presented by Almeida in 2009.
Hypomethylated DNA	Double stranded DNA is immunogenic and hypomethylated DNA even more so.

Table 1-11: Antigens in human tumours associated with the induction of an anti-tumour immune response. Summary derived from Rooney et al. 2015; Charoentong et al. 2017; Almeida et al. 2009; Serrano et al. 2011.

There is an ongoing balance between the tumour and the immune system which was illustrated by Zitvogel and colleagues in 2006 (Figure 1-3). Pre-malignant lesions are maintained by intact cancer immunosurveillance, in which immune cells detect and destroy malignant cells. As the tumour progresses the presence of an immune response selects for malignant phenotypes which are less

immunogenic in the process of immunoselection (which may also be referred to as immunoediting). This may include downregulation of some of the tumour antigens described in Table 1-11. Eventually tumours undergo immune escape by actively suppressing the immune response, allowing for more rapid growth. This is referred to as immunosubversion (Zitvogel et al., 2006). Tumour immune escape through the process of immunosubversion can take place through a number of mechanisms:

(1) interference with the anti-tumour response;

(2) suppression of effector T- cell function;

(3) downregulation of immune recognition signals;

(4) selection of immunoresistant phenotypes (Whiteside, 2008).

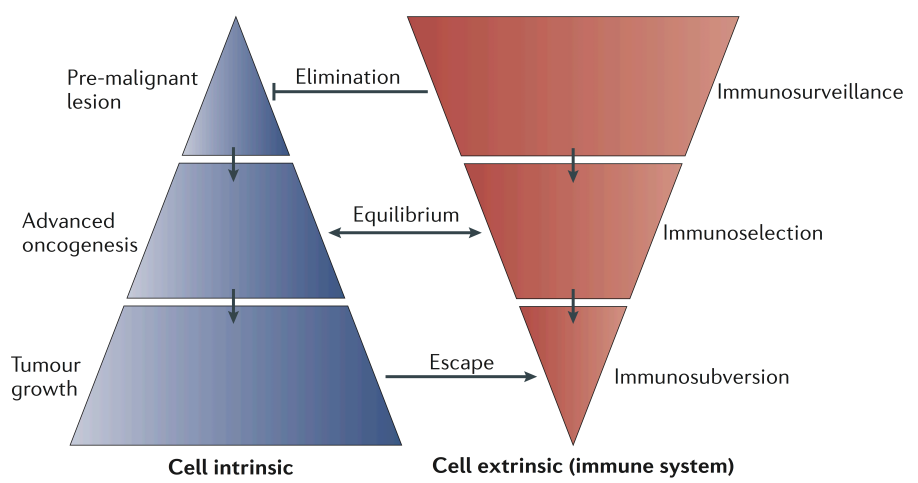


Figure 1-3: The balance between immune surveillance and tumour growth. Intact immunosurveillance results in failure of progression of a pre-malignant to a malignant lesion. Once control is lost, oncogenesis occurs encouraging immunoselection resulting in tumour growth and, as the tumour escapes immune control, immunosubversion. Figure reproduced from (Zitvogel et al., 2006) with permission of the Nature Publishing Group.

The role of the immune system is considered so critical to cancer development and progression that "evading the immune response" has recently been added as one of the new emerging hallmarks to the original six 'Hallmarks of Cancer' (Hanahan and Weinberg, 2011, 2000):

- Sustaining proliferative signalling;
- Evading growth suppression;
- Activating invasion and metastasis;
- Enabling replicative immortality;
- Inducing angiogenesis;
- Resisting cell death.

These hallmarks were designed to provide a framework to help understand the complexity of cancer. Each of the original six hallmarks have links to various components of the immune system (Zitvogel et al., 2006), thus emphasising the critical role of the immune system in cancer development. The other new emerging hallmark is 'deregulating cellular energetics'.

1.5.4 Immunotherapy in cancer

Researchers are now trying to harness the immune response in cancer therapy (Farkona et al., 2016). Attempts thus far have focussed on modifying the adaptive immune response.

Immune checkpoint blockade has become a well-known approach, arguably due to the dramatic, but often transient, responses sometimes generated. For example, ipilimumab in melanoma (Addeo and Rinaldi, 2013; Hodi et al., 2010). Checkpoint inhibition works by blocking inhibitory immune checkpoints, preventing immune cell inhibition by malignant cells, and theoretically stimulating an increased immune response. Malignant cells are known to express ligands which can interact with these inhibitory checkpoint receptors; suppressing the immune activity of cells they interact with. However, immunotherapy encompasses a multitude of modalities, with checkpoint blockade only beneficial in a relatively small subset of patients (Charoentong et al., 2017).

Examples of other approaches include:

- Monoclonal antibodies: bind specific receptors to induce cell killing, for example, trastuzumab in *HER2* positive breast cancer (Nahta and Esteva, 2003);
- Cytokines: stimulate the host immune response by acting on cellular receptors. Two adjuvant cytokine therapies have been approved in the USA for metastatic melanoma (IL-2 and IFN-α) and renal cell carcinoma (IL-2) (Lee and Margolin, 2011; Sharma et al., 2011);
- Cancer vaccines: induce an immune response against specific tumour antigens and have shown promise in some malignancies, including prostate cancer (Yaddanapudi et al., 2013);
- Chimeric Antigen Receptor T-cell therapy: reprogrammes the patients' own T-cells to recognise and eliminate tumour cells based on the identification of specific tumour antigens (Perica et al., 2015; Yu et al., 2017).

Whilst immunotherapies have shown some promise in solid tumours, progress has been hampered by the development of drug resistance (Sharma et al., 2017).

Within the brain the situation is more challenging. Historically, the brain has been considered an 'immune privileged site' because of the widely held belief that the systemic immune system was prevented from reaching the parenchyma by the blood brain barrier. However, there is emerging evidence that activated T-cells are able to pass through (Ransohoff et al., 2003). A better understanding of how therapeutic interventions might reach the tumour parenchyma is warranted.

In addition to the role of the blood brain barrier, the brain is regulated by its own immune system. Microglia are the resident macrophages whilst astrocytes are able to release pro- and anti-inflammatory mediators which may supress T-cell functioning (Gimsa et al., 2013; Perry and Teeling, 2013). Given this distinct immune environment, cancer therapies that may be effective elsewhere in the body need independent consideration within the brain.

Treatment regimens often include steroids and there is a risk that this could dampen down any induced immune response. Conversely there is also concern that inflammation induced by the immune response may cause significant neurological deficits. Additionally, in paediatric cohorts, there are few approved agents (Sayour and Mitchell, 2017).

1.6 The role and methods of next generation sequencing

RNA sequencing (RNA-seq) can potentially allow a more detailed analysis of gene expression profiles than array based techniques. It is a type of Next Generation Sequencing (NGS); a term used to describe modern, high-throughput, sequencing technologies.

It has been argued that the field of paediatric brain tumours has shown the greatest advance of any scientific field during the 'Next Generation Sequencing era' (Northcott et al., 2015).

1.6.1 The development of next generation sequencing

NGS allows deep sequencing of whole genomes and transcriptomes in a short space of time but at huge computational cost. It has replaced non-automated techniques, leading the cost of sequencing to fall precipitously (Wetterstrand, 2015) (Figure 1-4).

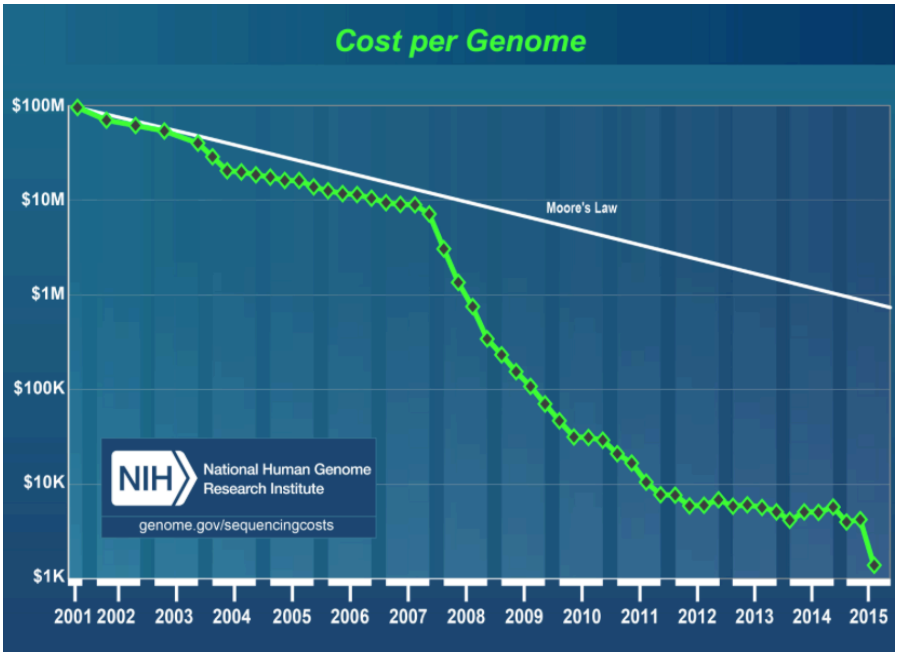


Figure 1-4: Data from the National Human Genome Research Institute demonstrating the rapid fall in whole genome sequencing costs in dollars since the beginning of the millennium. Moore's law refers to the rapid improvement in technology and cost. Reproduced with permission.

A method for sequencing DNA was first described by Frederick Sanger in 1975. This 'Plus and Minus method' was based on the principle of exposing known nucleotides, in the presence of DNA polymerase, to an expanding DNA fragment to determine which base was the next in the sequence. Fragments of different lengths were then separated on acrylamide gels and the position of each base in the fragment was inferred (Sanger and Coulson, 1975).

Subsequently the chain-termination method was proposed (Sanger et al., 1977). DNA fragments were assembled onto a complementary strand of DNA. The reaction occurred in the presence of DNA polymerase I, dTTP and ddTTP plus the other three nucleotides. The ddTTP contained no 3' hydroxyl group meaning that once the initial nucleotide was incorporated into the extending DNA strand, no further additions could take place (chain-termination). The reaction took place

concurrently in three other tubes to allow ddATP, ddCTP and ddGTP to act as terminating nucleotides. The resulting fragments then underwent electrophoresis, allowing quantification by autoradiography (Sanger et al., 1977). Modifications of this method included the use of 'Shotgun' sequencing, in which the DNA is fragmented and then amplified in bacterial vectors before sequencing and re-assembly (Franca et al., 2002; Venter et al., 1996). This approach has high redundancy (the same DNA is sequenced multiple times) with a huge data yield requiring complex computational approaches and was used by Celera in the initial sequencing of the human genome (Franca et al., 2002; Venter et al., 2001).

The study of biology, medicine and the life sciences has increased demand for rapid, reliable and cheap techniques for generating the sequence of a genome, transcriptome or epigenome. Developments such as the use of tiny capillaries rather than polyacrylamide gels for separation of the DNA fragments have allowed the rapid miniaturisation of this technology (Ansorge, 2009). In 1991 an application was made to patent 'Sequencing by Synthesis', an approach in which fluorescently labelled base terminators were annealed to the end of a developing DNA strand (Ansorge, 1991). Once the base annealed it would be detected by a camera before sequencing continued. Rather than being run on gels this would occur on miniature beads (Margulies et al., 2005). This development allowed parallel sequencing of many millions of DNA strands, vastly increasing throughput. Commercial organisations, such as Illumina (formerly Solexa), have developed this technique for extensive use. This approach can be applied to RNA as readily as to DNA, by converting RNA strands into a complementary DNA (cDNA) template prior to sequencing (Ansorge, 2009). This process is known as RNA sequencing.

1.6.2 The method of sequencing by synthesis

Sequencing by synthesis has been claimed to account for 90% of the world's sequencing data (Data calculation on file. Illumina, Inc 2015). A basic outline of the process divides it into four steps and is summarised here from information provided in Illumina documentation (Illumina, 2015):

(1) Library preparation (Figure 1-5A):
    DNA (for genome or exome sequencing) or cDNA (for RNA sequencing) is fragmented to create very short sequences called fragments. Oligonucleotide sequences called adapters are attached to the 3' and 5' ends of each fragment to create a sequencing library;
(2) Cluster amplification (Figure 1-5B):

The adapters are complementary to a lawn of oligonucleotides present on the surface of a glass slide (a flow cell). The sequencing library is applied to the flow cell and the adapter sequences attach the cDNA fragments to its surface. Once the fragments are attached they undergo polymerase chain reaction amplification to create tens of thousands of clusters. The co-ordinates of each cluster on the flow cell is known in order to perform downstream quality control assessments;

(3) Sequencing (Figure 1-5C):

Fluorescently labelled nucleotides are added to the sequencer and bind to the complementary base in the fragment, starting at the first base. Once this base has attached, a fluorescent image is taken of the flow cell and the colour determines the identity of the base. The action of identifying a nucleotide is termed a base call. The reaction repeats with the next base n times. n is set by the user, called the read length, and often ranges from 20-100 bases. This generates a sequence of base calls termed a read. Reads can be single end or paired-end. For single end reads, one read is generated from one end of the fragment, for paired-end reads two reads are generated, one from each end of the fragment. Millions of individual reads are generated for each sample under investigation;

(4) Alignment and data analysis (Figure 1-5D):

Once reads have been generated, they are combined into a fastq file which contains information on the base sequence and quality of each read. Fastq files can then be manipulated bioinformatically, firstly, by being subjected to quality control procedures and then by aligning to the genome and transcriptome. Alignment is the process by which the base sequence in each read is matched to the complementary sequence or sequences in the genome. This then allows for the determination of expressed genes or genetic mutations depending. Details on the nature of file formats used in sequencing analysis are included in Appendix 2.

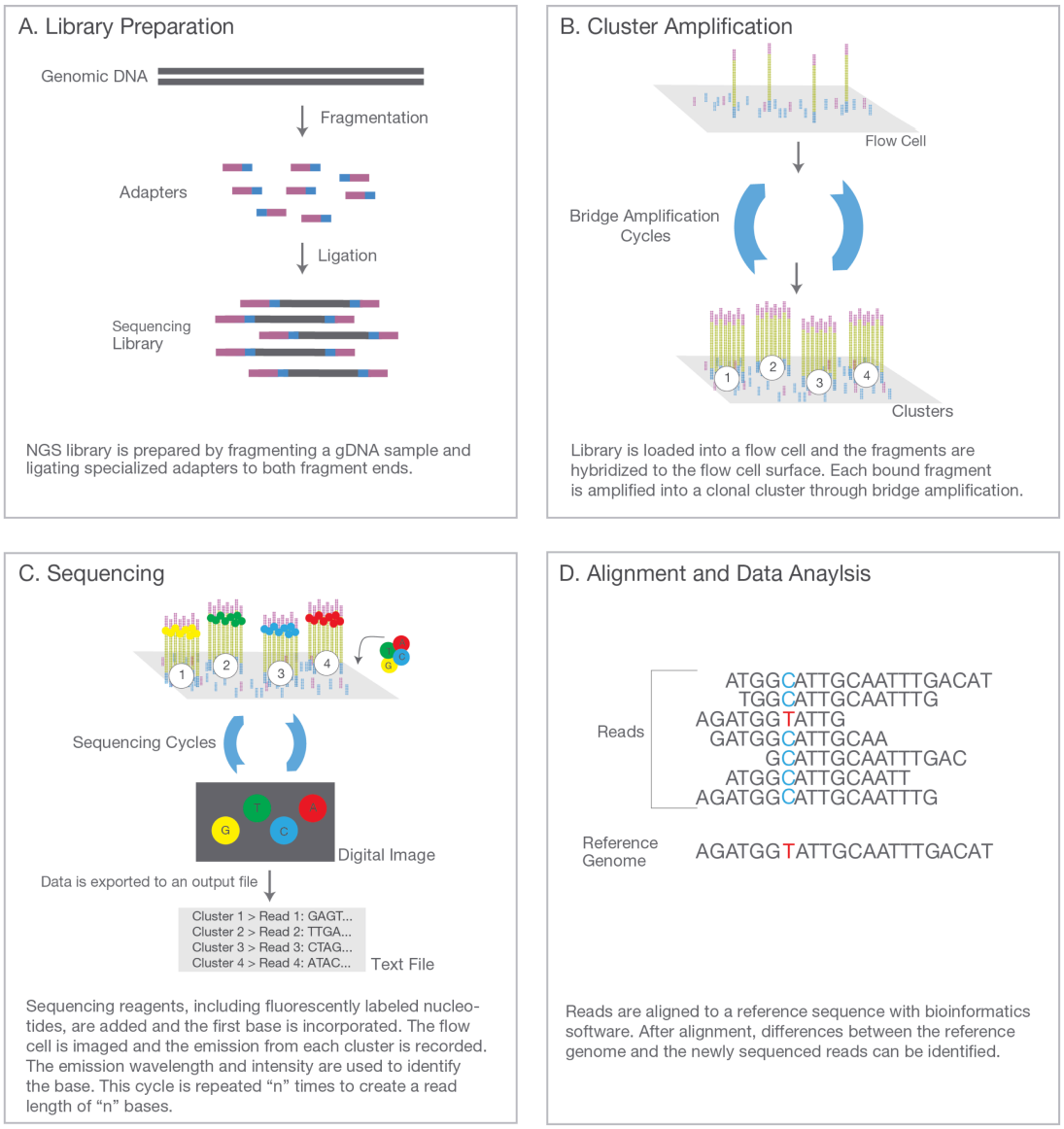


Figure 1-5: The four basic steps of Illumina sequencing by synthesis. Reproduced from (Illumina, 2015). (A) The sequencing library is generated by fragmenting DNA or cDNA before the addition of adapter sequences to the end of each fragment. (B) The fragments in the library are added to the sequencer by attaching the adapter sequence to complementary oligos distributed across a flow cell. Each fragment is then amplified to generate tens of thousands of clusters of fragments. (C) Complementary bases bind to the fragment and as they are fluorescently labelled the identity of the added base can be confirmed by imaging. After each base is confirmed the next base is incorporated and this is repeated for the length of the read (usually 25-100 bases). (D) The data from the reads is incorporated into a fastq file for downstream analysis, including alignment of reads to the genome.

1.6.3 Structure of a cDNA fragment after library preparation

Fragment length refers to the length of the cDNA strand being sequenced and is composed of:

- the adapter sequences at either end of the fragment;
- the paired-end reads;
- the nucleotide bases between the reads.

Insert length refers to the distance between the two adapter sequences.
Inner mate distance refers to the distance between the paired-end reads.
The length of the sequencing adapters and the length of the read are constant, therefore the insert length and inner mate distance are directly affected by the overall length of the fragment (Figure 1-6). A negative inner mate distance indicates that reads overlap. The value of the inner mate distance is required by genome alignment software. Inner mate distance may be calculated by measuring average fragment length on a Bioanalyser or by estimation directly from BAM files.

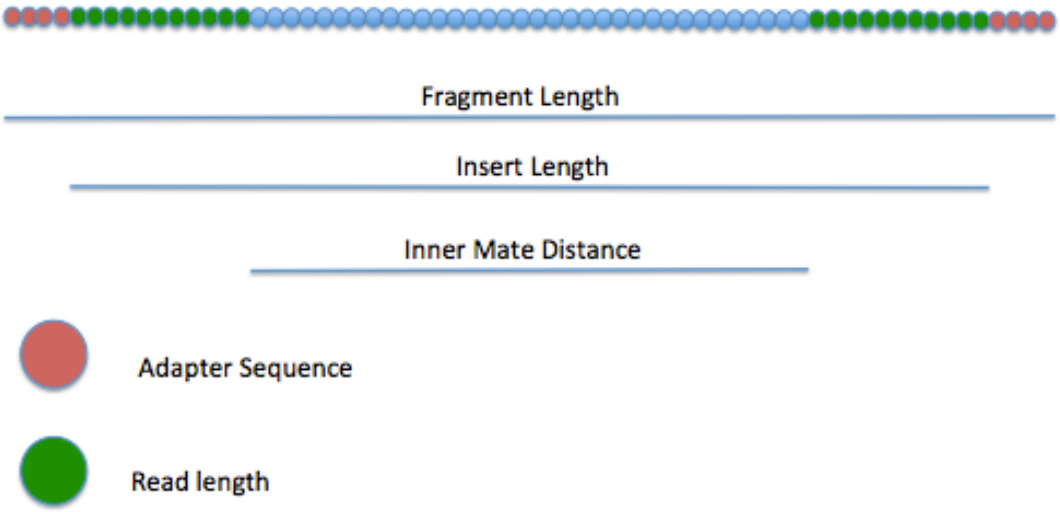


Figure 1-6: A graphical representation of the meanings of fragment length, insert length and inner mate distance. Each circle represents one sequenced base. Red circles represent the adapter attaching the fragment to the sequencing tile. Green circles represent bases actually 'read' by the sequencer. Blue circles represent the bases between the paired-end sequencing reads. The combination of red, green and blue circles represents the entire cDNA fragment.

1.6.4 Transcriptome arrays versus RNA-seq

Array based technologies are used to profile the transcriptomes of various cells and tissues. They can yield a wealth of information but are limited to predefined sequences, requiring a priori knowledge of the genome (Schuster, 2008; Wang et al., 2009). There is also risk of cross-hybridisation of the probe, resulting in high

levels of background noise (Ansorge, 2009; Wang et al., 2009). Arrays are unable to achieve transcriptome comprehensiveness; that is complete coverage of all genes and detection of all RNAs (Mortazavi et al., 2008; Wang et al., 2014, 2009).

In contrast, RNA-seq is able to identify novel transcripts; genes expressed at very high and low levels; small insertions and deletions (Indels); single nucleotide polymorphisms (SNPs); and splice variants (Fang and Cui, 2011; Kratz and Carninci, 2014; Wang et al., 2014, 2009). RNA-seq can produce higher resolution outputs, to the level of individual nucleotides, and requires relatively little input RNA (Wang et al., 2009). Drawbacks to RNA-seq include its relative expense and potential for bias and background noise complicating analysis. It is important to design a rigorous experiment, paying attention to sequencing depth and number of replicates (Fang and Cui, 2011; Kratz and Carninci, 2014).

Errors in RNA sequencing can occur at multiple stages, particularly in library preparation, sequencing and data analysis. When billions of bases are examined, even small inaccuracies can result in significant numbers of incorrect calls, with potential for incorrect genome assembly. The Illumina sequencing platform incorporates information about the reliability of base calls into its fastq output files as the Phred score (Ewing and Green, 1998; Ewing et al., 1998). Phred scores indicate the probability of errors in the sequencing process and are represented using a logarithmic scale (Table 1-12). The Phred score (q) generated by Illumina RNA sequencing protocols ranges between 0 (0% accurate) and 40 (99.99% accurate), and is calculated using the formula:

$$q = -10 \times \log_{10}(p)$$

Where p is the probability of the base call being correct (Ewing and Green, 1998).

Phred scores often deteriorate as a run progresses, therefore adequate data quality control before trimming of poor quality bases is important.

Phred Score	**Probability call is wrong**	**Accuracy of call**
10	1/10	90%
20	1/100	99%
30	1/1000	99.9%
40	1/10000	99.99%
50	1/100000	99.999%

Table 1-12: Illustration of how accuracy of base call changes with Phred score. Modified from http://www.phrap.com/phred/

1.6.5 RNA sequencing of degraded Samples, including FFPE

There is increasing use of RNA-seq for degraded samples. Of particular interest are those samples that have been fixed in formalin and embedded in paraffin (FFPE) shortly after collection. Formalin fixation has been taking place since the late 19th century (Fox et al., 1985), resulting in over one billion FFPE archival specimens worldwide (Blow, 2007). The ability to use these in retrospective biological studies would therefore be of huge potential benefit.

There are several reasons for the degradation of RNA quality in FFPE tissue:

- Formalin forms reversible cross-links between protein (histones) and nucleic acids (Chalkley and Hunter, 1975) and is associated with the addition of mono-methylol groups to nucleotides (Masuda et al., 1999). Both processes can inhibit the polymerase chain reaction during sequencing (Williams et al., 1999);
- Prolonged time between ligation of blood supply and fixation, allows degradation to occur and this process is likely to vary from sample to sample (Blow, 2007);
- Exposure to light and variable temperatures after tissue fixation (von Ahlfen et al., 2007);
- The size of the specimen: formalin penetrates thicker tissue sections more slowly, leaving tissue at the centre unfixed for variable periods (von Ahlfen et al., 2007).

Messenger RNA (mRNA) constitutes 4% of the total RNA, with the majority of the remainder being ribosomal RNA (rRNA). mRNA contains transcripts of expressed genes and is therefore the main molecule of interest for RNA-seq. If libraries made from total RNA are used there will be extensive sequencing of rRNA at the expense of mRNA. Therefore, in order to measure gene expression via RNA sequencing there must be a process of mRNA enrichment.

One technique is to capture mRNA via the polyadenylated 3' end of the molecule, however this region is lost in degraded material. An alternative approach in such a situation is to remove the rRNA via the use of magnetic beads (ribodepletion) prior to cDNA synthesis.

The extent of RNA degradation is indicated by measurement of the RNA Integrity (RIN) score on a Bioanalyzer (Schroeder et al., 2006) and can help to determine the best approach to mRNA enrichment (Figure 1-7).

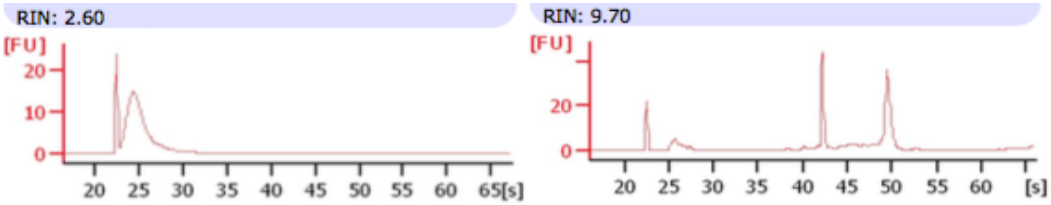


Figure 1-7: Bioanalyser traces for low quality, degraded RNA with a RIN score of 2.60 (left) and high quality RNA with a RIN of 9.70 (right). The high quality RNA exhibits distinct peaks at 40-45s and 50s which represent intact ribosomal RNA. These peaks are not present in the low quality sample indicating that the rRNA is heavily degraded.

One way of overcoming some of the difficulties related to degradation is to increase the depth of sequencing (the overall number of reads) (Robasky et al., 2014). Greater depth also allows better assessment of infrequently expressed transcripts in RNA-seq experiments (Marioni et al., 2008). However, one must be cautious; increased depth can lead to increased false positive base calls and analysis of the results must take this into account (Tarazona et al., 2011). In addition to this a number of studies have found that beyond a certain number of reads, increasing depth adds very little power to expression analysis and increasing the number of biological replicates becomes more important (Ching et al., 2014; Yuwen Liu et al., 2014).

Using kits designed specifically for extracting nucleic acids from degraded material, studies are beginning to show success in RNA-seq of FFPE tissue. These kits digest surrounding materials with proteinase K and attempt to reverse the additions of mono-methylol groups with heat treatment (Masuda et al., 1999). A few studies sequencing matched FFPE and frozen specimens to compare obtained transcriptomic profiles have now been reported (Graw et al., 2015; Hedegaard et al., 2014; Li et al., 2014; Yu Liu et al., 2014; Mittempergher et al., 2011; Morlan et al., 2012). These studies found high correlation coefficients for expressed genes, ranging from 0.7 to greater than 0.9. Most excitingly, the RNA from FFPE tissue from a victim of the 1918 flu pandemic was sequenced and compared with that of a 2009 flu victim. The 1918 and 2009 samples both demonstrated expression of genes related to cellular defence and immune and inflammatory responses. The sequencing of this sample of nearly 100 years of age showed that RNA-seq is applicable to FFPE specimens. However, caution is required in experimental design to maximise the utility of the results (Xiao et al., 2013).

Whilst data from initial studies using FFPE derived RNA is promising, literature searches suggest that there have been no large profiling studies undertaken using FFPE brain material for RNA sequencing. This is important because the composition of RNA in the brain may differ from other tissues, with increased levels of nascent (partially processed) RNA which may make the sequencing process more challenging (Ameur et al., 2011).

1.7 Aims

Using a large clinical cohort, this study aimed to undertake molecular profiling of recurrent ependymoma, combined with contemporary clinical data, to better understand recurrence biology and potential therapies. In particular, improved knowledge of how tumours change from primary to recurrence in the same patient was sought. It was hoped that any new understanding of ependymoma recurrence may provide new insights into future therapies or management strategies.

In planning to achieve the above goals a number of specific aims were developed which included:

(1) Collating and analysing a cohort of recurrent paediatric ependymoma cases to determine:
   a. Patterns of relapse in the entire cohort and location determined subgroups;
   b. Factors impacting upon time to first relapse and OS in a relapsed cohort;
   c. Factors influencing progression and OS after first relapse;
   d. Factors affecting risk of relapse;

(2) Supporting the results of the clinical analysis by generating DNA methylation profiles for samples with tissue availability;

(3) Undertaking RNA sequencing of a cohort of FFPE tumour specimens in order to expand the cohort for primary and recurrence analysis.

(4) Validation of the use of FFPE RNA-Seq against a cohort of fresh frozen specimens in order to make recommendations for future research;

(5) Performing gene expression analysis of matched primary and recurrent pairs to determine changes in expression patterns at relapse and correlate with molecular classifications where DNA methylation data is available;

(6) Establishing, in an independent cohort, whether the immune response is implicated in ependymoma recurrence;

(7) Validating key expression changes using immunohistochemical approaches.

1.8 Thesis structure and navigation

In order to achieve the aims the project was divided into:

Chapters 3 and 4: A clinical analysis of a cohort of 188 children with recurrent ependymoma, supported by DNA methylation profiling;

Chapter 5: Validation of FFPE RNA-Seq on a cohort of 106 FFPE paediatric ependymoma specimens;

Chapter 6: Gene expression RNA-seq analysis of matched primary and recurrent ependymomas from 29 fresh frozen cases and 27 FFPE cases;

Chapter 7: Follow up of gene expression findings and investigation of the level of immune cell markers in recurrent ependymoma.

Methods relevant to the entire project are included in Chapter 2. Methods specific to individual chapters are included in the methods section of that chapter. Chapter locations of the different methods are indicated in Table 1-13.

Chapter	Methods
2	Pathology assessment Nucleic acid extraction Haematoxylin and Eosin staining Basic data analysis protocols
3	Survival analysis
4	Generation of DNA methylation profiles Ependymoma subgroup classification
5	RNA sequencing bioinformatic pipeline Differential expression analysis Hierarchical clustering Gene ontology analysis
6	Quantitative PCR Generation of immunophenotypes
7	Immunohistochemistry

Table 1-13: Summary and location of the methods used in the research.

2 Materials and Methods

2.1 Sample identification

A cohort of primary and recurrent ependymomas was generated by interrogating the Children's Brain Tumour Research Centre's (CBTRC) master database. Tissue for only 20 -30 primary and recurrent pairs were available through this route. Additional requests were sent via the Children's Cancer and Leukaemia Group (CCLG) Biobank to centres across the UK. Collaborations were developed with a centre in Denver, Colorado Great Ormond Street Children's Hospital (GOSH), UK. The final study cohort contained 95 primary and recurrent tumour pairs with a total of 243 individual tumour samples (Appendix 1). Specimens were a mixture of fresh frozen (FF) and FFPE material.

Clinical data was available for a cohort of 208 patients who suffered at least one recurrence. Data included age, gender, number of recurrences, timing of recurrences, treatment protocols, trial inclusion, tumour location, extent of resection and dates of surgery. When unavailable, data was requested from primary treating centres via the Children's Cancer and Leukaemia group. Individual centres were also asked to identify cases that were not accessible through the CBTRC archives in order to increase the size of the cohort.

2.2 Ethics and consent

Ethical approval came from the local research ethics committee. The study was also approved by the CCLG to allow specimens to be obtained from their biobank. Consent had been obtained from all patients or their families where relevant. Where consent was not in place but the samples formed part of existing holdings (obtained before 1st September 2006), data and tissue samples were handled in accordance with the guidelines of the Human Tissue Act. Data was handled with due respect to patient anonymity and held on a secure database within the university.

2.3 Definition of tumour recurrence

Recurrence was defined as the return of a tumour that had been completely resected (GTR) or, for those children who had received a subtotal resection (STR), a clinical or radiological progression of tumour requiring further therapeutic intervention or palliation. Where available, a pathology specimen was

analysed to confirm that the recurrent tumour represented ependymoma rather than another tumour.

2.4 Preparing frozen samples for nucleic acid extraction

Samples were identified and removed from the minus 80ºC freezer and placed onto dry ice in a category two hood. Tissue samples were taken from their tubes one at a time and replaced before the next sample was removed. Each sample was placed on an open petri dish on a shallow bed of dry ice to reduce the risk of nucleic acid degradation. Using a sterile blade (Swann-Morton, UK) and forceps (John Weiss International, UK), approximately 10 mg of tissue was removed and placed into a 1.5 ml vial (Eppendorf, UK). A small amount of tissue was also fixed onto an uncoated glass slide in preparation for pathology assessment by placing it into a Coplin jar of Carnoy solution before staining.

2.5 Haematoxylin and Eosin staining

Haematoxylin and Eosin (H&E) staining was performed for pathology review. The protocols for FFPE sections and frozen smears are outlined below.

FFPE Sections

5 μm wax sections were rehydrated through xylene for 10 minutes, 100% ethanol for 10 minutes and 95% ethanol for 10 minutes before being washed in tap water. The sections were then placed into Gills-3 haematoxylin for five minutes before a further wash with tap water. Next, slides were placed into 1% lithium carbonate, 1% acid alcohol and 1% lithium carbonate again for a few seconds each, with tap water washes between each step. The final stain was performed with 1% eosin for five minutes before a wash with tap water. Sections were dehydrated, for ten seconds each, through 95% ethanol, 100% ethanol and xylene. Slides were mounted using DPX (distyrene, plasticiser, xylene) mountant (Sigma-Aldrich, UK) and left to dry overnight on a heated rack.

Frozen Smears

Slides were removed from Carnoy solution and agitated in haematoxylin for 10 seconds and then placed under running tap water until it cleared. Following this they were agitated in 1% lithium carbonate for 10 seconds and then in 1% eosin for 20 seconds, with tap water washes in between. Following staining the slides were dehydrated in 95% ethanol, 100% ethanol and xylene for 10 seconds each, mounted in DPX and left to dry.

2.6 Pathology review

Cases were reviewed within their originating centres for diagnosis of ependymoma. Those patients who had been included on a clinical trial had undergone central pathology review to confirm diagnosis.

Additionally, H&E stained sections were reviewed by a neuropathologist (Dr Simon Paine) at the CBTRC who confirmed that the sample was consistent with ependymoma and was predominantly tumour. Samples were excluded for lacking sufficient viable tumour or for not being consistent with a diagnosis of ependymoma.

The samples extracted from frozen RNA contributed by collaborators in Denver underwent tissue assessment and nucleic acid extraction in the USA before being shipped to the UK.

2.7 Nucleic acid extraction

To minimise the risk of contamination, gloves were worn at all times and RNA specific pipettes and tips were maintained. The workbench was cleaned prior to commencement of the protocol. RNAse- and DNAse-free tubes (Eppendorf, UK) were purchased to minimise the risk of contamination post extraction.

2.7.1 RNA extraction from frozen tissue

The mirVana™ miRNA isolation kit (ThermoFisher Scientific, UK, AM1560) was used to extract RNA from frozen tissue. Sections were placed into lysis buffer equivalent to 10 times the volume of the tissue in a flat-bottomed vial and homogenised (Stuart, UK). 1/10th the volume of miRNA homogenate additive was added to the tissue lysate and mixed by vortexing, before being left on ice for 10 minutes. The equivalent of one volume of tissue mass of acid-phenol:chloroform mixture was then added to the vial and mixed by vortexing for 30 seconds. The vial was subsequently centrifuged at 10000 x g for five minutes. The resultant aqueous (upper) phase was transferred to a new vial without disturbing the lower phase which was discarded. 1.25 volumes of room temperature absolute ethanol were added. A filter column was inserted into a pre-supplied collection tube and 700 μl lysate/ethanol mixture was pipetted onto it and centrifuged at 10000 x g for 15 seconds. The flow through was discarded. Up to a further 700 μl lysate/ethanol mixture was added again to the filter until depleted. The filter column was kept in the same collection tube for wash steps which involved:

- 700 μl miRNA wash solution 1 followed by 10 second centrifuge;

- 500 μl wash solution 2/3 followed by 10 second centrifuge, repeated once.

Following the wash steps the flow through was discarded and the filter column centrifuged for one minute to remove residual fluid. The filter column was transferred into a fresh collection tube before the addition of 100 μl of preheated (95ºC) nuclease free water. This was spun for 30 seconds to elute the final RNA. Eluted RNA was stored at -80ºC.

2.7.2 RNA extraction from cell lines for qPCR

Cell pellets were harvested from a 70% confluent T75 flask (Eppendorf, UK) and snap frozen. RNA was extracted from cell lines using RNA STAT-60™ Extraction Reagent (Amsbio, UK). 500 μl RNA STAT-60™ was added to each cell pellet. 100 μl chloroform was added in a fume hood before covering and vortexing for 15 seconds. Samples were then incubated at room temperature for two minutes before being centrifuged at 12000 x g for 15 minutes. The aqueous phase was transferred to a new plastic vial and incubated at room temperature for 10 minutes after the addition of 250 μl isopropanol. The sample was centrifuged again at 12000 x g for 15 minutes. The supernatant was discarded and the pellet was washed with 200 μl 75% ethanol before vortexing. The solution was centrifuged at 7500 x g for five minutes, before removal of the supernatant and air drying of the pellet at room temperature for 10 minutes. Finally, the pellet was dissolved in 25 μl RNAse-free water and vortexed before a further incubation at 55ºC for 10 minutes on a hot block (Eppendorf, UK). RNA was stored at -80ºC.

2.7.3 DNA extraction from frozen tissue

The QIamp DNA Mini Kit (Qiagen, Germany) was used to extract DNA from frozen specimens. 10 mg of tissue was disrupted with a homogeniser in 80 μl phosphate buffered saline (PBS). Following this, 100 μl buffer ATL and 20 μl proteinase K was added. The solution was incubated at 56ºC for three hours with occasional vortexing until the tissue had lysed. Subsequently, 200 μl buffer AL was added before a 10 minute incubation at 70ºC on a hot block. 200 μl absolute ethanol was added and vortexed before applying the mixture to a QIAmp Mini Spin Column placed in a 2ml collection tube. The tube was centrifuged at 6000 x g for one minute and the filtrate discarded. The following washes were then performed:

- 500 μl buffer AW1 followed by centrifugation at 6000 x g for one minute;
- 500 μl buffer AW2 followed by centrifugation at 6000 x g for three minutes.

The column was then placed into a new tube and centrifuged at full speed for one minute to dry the membrane. The column was then transferred to a new

collection tube before being incubated for five minutes with 100 μl buffer AE (elution buffer) and spun at 6000 x g for one minute. The elution step was repeated twice to maximise DNA yield. DNA was stored at -20ºC.

2.7.4 RNA and DNA extraction from FFPE tissue

RNA and DNA from FFPE samples were extracted with the AllPrep DNA/RNA FFPE Kit (Qiagen, Germany). Starting material was either eight 10 µm scrolls (where collaborating centres were unable to provide original blocks) or ten 5 µm cores taken from tumour representative areas. The samples were placed into 1.5 ml vials (Eppendorf, UK) until ready for processing. RNA and DNA were extracted from the same piece of tissue, preserving tumour material and theoretically giving better comparability between RNA and DNA studies.

Extraction was carried out according to the recommended three step protocol: (1) *Sample preparation*: Specimens were deparaffinised by adding 1 ml xylene to a vial of crushed cores or scrolls and incubated for 10 minutes on a hot block at 50ºC. The vial was spun at maximum speed on a centrifuge for three to five minutes until a pellet had formed. The xylene was removed using a pipette and 1 ml absolute ethanol was added before further centrifugation to reform a pellet. The ethanol step was repeated once. After deparaffinisation, 150 μl buffer PKD and 10 μl proteinase K were added to the tube. This mixture was incubated for 15 minutes at 56ºC and then placed on ice for three minutes. Following a further 15 minute centrifugation at 20000 x g, the supernatant was transferred to a 2 ml vial whilst the pellet was kept for DNA extraction;

(2) *RNA extraction*: The supernatant from step one was incubated for 15 minutes at 80ºC before the addition of 320 μl buffer RLT and 1120 μl absolute ethanol. 700 μl of this sample was pipetted into an RNeasy MinElute spin column placed in a 2 ml collection tube (supplied with the kit). The spin column was centrifuged for 15 seconds and the step repeated until all of the lysate had been used. Next, 350 μl buffer FRN was added to the column and centrifuged for 15 seconds. 80 μl DNase stock solution, mixed in a 1:7 ratio with buffer RDD, was added to the spin column and incubated at room temperature. After 15 minutes 500 μl buffer FRN was added to the spin column and centrifuged for 15 seconds. The flow through was transferred back into the spin column which was placed in a new collection tube. 500 μl buffer RPE was added to the column and centrifuged. This step was repeated once before placing the column in a new collection tube and centrifuging at full speed for five minutes with an open lid to dry the membrane. Finally, the

spin column was placed in a new collection tube and RNA was eluted in 30 μl RNase-free water by centrifuging for one minute following a one minute incubation. The eluted RNA was stored at -80ºC;

(3) *DNA extraction*: The pellet resulting from step one was resuspended in 180 μl buffer ATL before the addition of 40 μl proteinase K. After vortexing, the sample was incubated for one hour at 56ºC and then for 2 hours at 90ºC. After a short centrifugation, 200 μl buffer AL was added with 200 μl absolute ethanol. This mixture was added to a QIAmp MinElute spin column and centrifuged for one minute. The spin column was placed in a new collection tube before addition of:

(a) 700 μl buffer AW1 with 15 second centrifuge;
(b) 700 μl buffer AW2 with 15 second centrifuge;
(c) 700 μl absolute ethanol with 15 second centrifuge.

On each occasion, the flow through was discarded. The column was then spun for five minutes to dry the membrane before being placed in a new collection tube for DNA elution. 100 μl of buffer ATE was added to the membrane before a one minute incubation and one minute centrifugation to elute the DNA. The eluted DNA was stored at -20ºC.

2.8 Nucleic acid quality control

Following extraction, RNA and DNA purity and quantity was measured by adding 1 μl of nucleic acid to the Nanodrop spectrophotometer (ThermoFisher Scientific, UK). Nucleic acids with 260/280 and 260/230 ratios close to two were considered to be of high purity.

2.9 RNA sequencing

RNA sequencing was performed by Exiqon (Denmark). RNA library preparation was carried out using a ribodepletion technique with Ribo-Zero (Illumina, USA), and libraries were sequenced on an Illumina HiSeq machine with 100 base pair, paired-end sequencing, targeting 50 million reads per sample. The input amount of RNA for the FF and FFPE samples was 100 ng and 600 ng respectively. Sequencing of the FF and FFPE samples was conducted as separate projects, as including RNA of substantially different qualities on the same sequencing run can interfere with the final results (Personal communication, L Klitten). Each RNA-seq sample generated two fastq files (Appendix 2) containing the raw reads, one file with forward reads and one file with reverse reads, representing the paired-end sequencing. Full details of the analysis are included in the methods section of Chapter 5.

2.10 DNA methylation profiling

DNA methylation profiling was performed by UCL Genomics (UK). DNA was hybridised to 450k Illumina methylation arrays following bisulphite conversion. Full details are included in the methods section of Chapter 4.

2.11 Cluster computing and data storage

To conduct the more computationally intense parts of the analysis, access was gained to the High Performance Computing Cluster (HPC) at the University of Nottingham. The cluster was accessed remotely using a login node, but operations were conducted on one (or more) of the compute nodes and submitted as Shell scripts using the 'qsub' command. This approach allowed the bioinformatic pipelines to be run in parallel for different samples on shorter timescales, meaning that each run took several weeks rather than months or longer.

Data was stored securely across a research drive and cloud based facility 'UoN BOX'. Files were also backed up onto encrypted hard drives to protect against data loss.

2.12 Data analysis

Statistical analyses were performed using R (R Core Team, 2014) running in RStudio (version 0.99.489). The code for the underlying analyses is reproduced in Appendix 3. For all statistical tests a p-value of <0.05 was considered significant.

3 A Clinical Analysis of 188 Recurrent Ependymomas

3.1 Introduction

Epidemiological and clinical studies of paediatric ependymoma have tended to be small (less than 100 patients) (Antony et al., 2014; Zacharoulis et al., 2010) or confined to few centres (Antony et al., 2014) or geographical locations (Messahel et al., 2009). A number of studies investigated specific treatment options at recurrence but did not analyse the whole spectrum of relapsed disease (Lobón et al., 2016; Hoffman et al., 2014a; Tsang et al., 2017). Other studies are old and assessment of their relevance to current practice, particularly in view of the development of molecular classification and survival improvements, is required (Goldwein et al., 1990). Consequently, the heterogeneity of the available literature can make general application of the results to recurrent paediatric ependymoma difficult.

50% of paediatric ependymomas relapse and subsequent survival is poor (Messahel et al., 2009; Zacharoulis et al., 2010). It is surprising that there are no recent, large reviews focussing on the features of relapse across multiple centres and geographic locations. This may be in part due to the relative rarity of the disease; a problem that has been highlighted with a plea for more transnational co-operation (Bouffet et al., 1998).

In identifying clinical cases for inclusion in biological studies of matched primary and recurrent ependymoma, a retrospective analysis of clinical data was undertaken. This aimed to assess the features of recurrent ependymoma in an exclusively paediatric cohort, from multiple centres. During data collection, a cohort of patients with non-recurrent ependymoma was established to provide a direct comparison. However, the primary aim was a description of the epidemiology and outcomes of relapsed disease.

The research questions were:

- What is the natural history of relapsed paediatric ependymoma?
- Which, if any, factors confer a higher risk of relapse in paediatric ependymoma?
- What, if any, factors impact upon OS in the relapse cohort?
- What, if any, factors impact upon time to first relapse?

- What is the impact of therapy at relapse?

3.2 Materials and methods

3.2.1 Definition of patient outcomes

Outcomes of interest were overall survival (OS), event free survival (EFS) and time to progression (TTP). OS was defined as the duration from diagnosis of primary ependymoma to death. EFS was defined as the duration from diagnosis of primary ependymoma until progression or death. For OS and EFS, patients who had not progressed or died by the end of the follow up period were right censored at the date of last follow up. TTP was defined as the duration from diagnosis of primary ependymoma to disease recurrence or progression. TTP allowed the assessment of patients who recurred, whilst excluding those who died; providing a quantitatively different result to EFS.

3.2.2 Survival analyses

Univariate analyses

Survival analysis employs techniques for interrogating data where the outcome of interest is the time to occurrence of a pre-specified event. Survival curves and Cox Proportional Hazards (CPH) models were used to undertake this analysis through the R packages 'survival', 'survminer' and 'survMisc' (Dardis, 2016; Kassambara and Kosinski, 2017; Therneau, 2015).

Survival analysis using the 'survival' package was based on the creation of a survival object, which took the format:

```
> SurvObject = Surv(Time, Event)
```

'Time' was a vector containing the time for each patient to reach the outcome and 'Event' was a vector of information about whether the patient experienced the event of interest or was right censored.

A statistical comparison between two or more curves was created using the 'Survdiff' function, which took the format:

```
> Survdiff(SurvObject ~ X, rho=Y)
```

'X' represented the factor by which data was stratified for the statistical comparison, for example a vector of data containing the gender of the patient or

whether they received a certain type of therapy. 'rho' specified the statistical test. Rho=0 used the log-rank test, which required the assumption of proportional hazards to be met. If the curves crossed, the log-rank test was considered insufficiently powered; consequently, all curves were visually inspected. When curves crossed, the supremum (Renyi) set of tests (Fleming et al., 1987) were implemented using the survMisc package.

In the absence of censoring, the Wilcoxon rank-sum test was used to compare the medians of the two groups.

Multivariate analyses
CPH models were used for multivariate analyses by incorporating factors that were statistically significant in univariate analyses.

CPH models were generated in 'survival' using the generic code:

```
> coxph = coxph(Surv(Time, Event) ~ Z)
```

Where 'Z' represented a vector comprised of all of the variables under investigation used to stratify the cohort.

The assumptions of the CPH model are firstly, that censoring is non-informative and secondly, that hazards are proportional over time. The proportionality of hazards was evaluated with a Schoenfeld test using the following code in the 'survival' package:

```
> cox.zph(coxph, transform="rank", global=TRUE)
```

A result suggestive of non-proportional hazards had a global p-value below 0.05.

3.2.3 The non-recurrent cohort

The data for the non-recurrent cohort was collected by a colleague, Dr Hazel Rogers. Patients were included who had a primary diagnosis of ependymoma but had not recurred. Competing risks regression (Fine and Gray, 1999) was used to test for the association with risk of recurrence using STATA (Statcorp, Texas, USA), including death before first recurrence as a competing risk.

3.3 Results

3.3.1 Defining the recurrent and non-recurrent cohorts

438 cases of primary disease were identified from the CBTRC archives. 88 cases were excluded based on having either inadequate clinical data, or a non-WHO grade II or III ependymoma. After excluding cases there was a cohort of 350 children (Figure 3-1A).

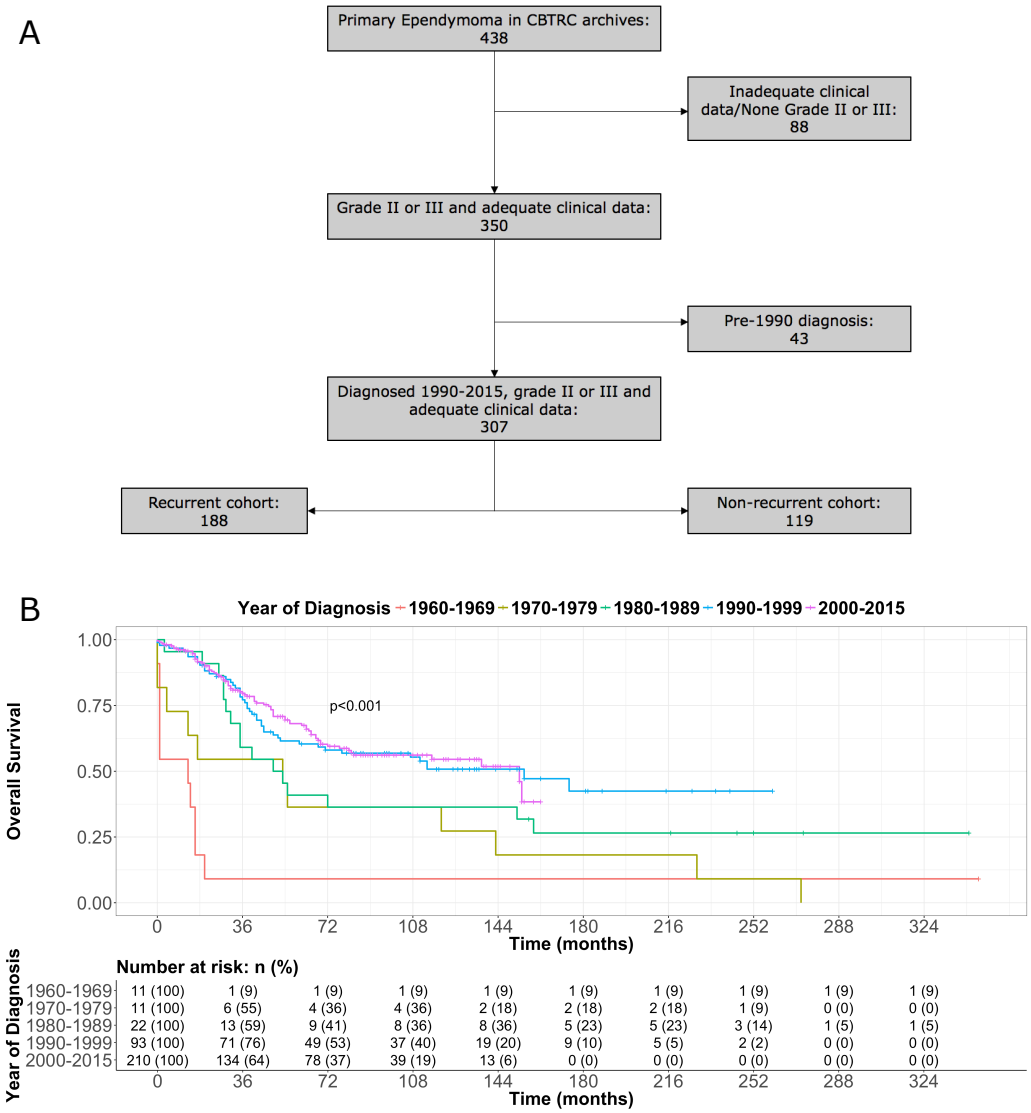


Figure 3-1: (A) Patient flow through the study. Recurrent and non-recurrent cohorts included patients with grade II or III ependymoma, diagnosed after 1989, with sufficient clinical data for analysis. (B) Children diagnosed after 1989 had significantly improved OS compared with those diagnosed earlier. Children diagnosed in the 1960s had the worst outcomes.

Children diagnosed before the 1990s had worse outcomes than those diagnosed later ($p<0.001$). In particular, a high proportion died prior to recurrence (Figure 3-1B). There was no difference in OS for those diagnosed between 1990 and 2000 compared with after the year 2000 ($p=0.415$). The cohort was therefore restricted to diagnoses from 1990 onwards, giving a total of 307 children, including 188 recurrences.

3.3.2 Baseline characteristics of the recurrent cohort

The recurrent cohort included 188 patients who recurred at least once between 1990 and 2015. 65 patients were treated on the International Society of Paediatric Oncology (SIOP) 1992 'Baby Brain' protocol (Grundy et al., 2007) and 47 on the SIOP 1999 protocol (unpublished). The remaining 76 were treated outside the confines of a clinical trial. 157 patients were treated in the UK across 16 centres, 16 in the USA, 9 in Holland, 5 in Dublin, 2 in Denmark and 2 in Spain.

114 (60%) patients had died and 70 (37%) were alive at last follow up. Median OS for the group was 61 months (95% confidence interval 48-71 months). 5 and 10-year OS was 50% and 33% respectively (Figure 3-4A). Median follow up duration was 50 months for all patients (range 1-260 months) and 97 months for patients still alive at the end of the study (range 13-260 months).

136 (72%) tumours occurred in the posterior fossa, 44 (23%) in the supratentorium and 7 (4%) in the spine. Location was unknown for one patient. There was a 1.3:1 male:female ratio, with boys accounting for 56% of the cohort ($n=105$) and girls for 41% ($n=77$) ($p=0.050$). Gender was unknown in 6 cases (3%) (Table 3-1).

Median age of diagnosis of primary ependymoma was 35 months (3 to 199 months) and exhibited a unimodal distribution (Figure 3-2A). PF, ST and spinal tumours presented at median ages of 30, 67 and 136 months respectively. There were significant differences for all tumour locations. PF versus ST $p=0.014$; PF versus spinal $p<0.001$; ST versus spinal $p=0.03$ (Figure 3-2B).

85 (52%) primary tumours were WHO grade II, 78 (48%) WHO grade III and 25 had no information about grade. 76 (45%) patients had GTR, whilst 93 (55%) had STR. Extent of resection was unavailable for 19 patients. 77 (43%) patients received radiotherapy. Radiotherapy data was unavailable in 10 cases. 125 (73%) patients received chemotherapy. Chemotherapy data was unavailable for 17 (Table 3-1).

Parameter		Recurrent Cohort (n=188)		Non-recurrent Cohort (n=119)		Chi-Square P Value
		Number	%	Number	%	
Age	<3 years	94	51	47	40	0.076
	3+ years	91	49	71	60	
	NK	3	-	1	-	
Gender	Male	105	58	65	55	0.817
	Female	77	42	53	45	
	NK	6	-	1	-	
Extent of Resection	GTR	76	45	65	59	0.028
	STR	93	55	46	41	
	NK	19	-	8	-	
Location	PF	136	73	75	64	0.405
	ST	44	23	31	27	
	SP	7	4	10	9	
	NK	1	-	2	-	
Grade	WHO II	85	52	58	62	0.153
	WHO III	78	48	36	38	
	NK	25	-	25	-	
Radiotherapy at diagnosis	Yes	77	43	32	50	0.382
	No	101	57	32	50	
	NK	10	-	55	-	
Chemotherapy at diagnosis	Yes	125	73	67	58	0.008
	No	46	27	49	42	
	NK	17	-	3	-	

Table 3-1: Baseline characteristics of the recurrent and non-recurrent cohorts. Patients in the recurrent cohort were significantly less likely to have received GTR but significantly more likely to have received chemotherapy. NK: Not Known. Percentages calculated as the proportion of those with data available.

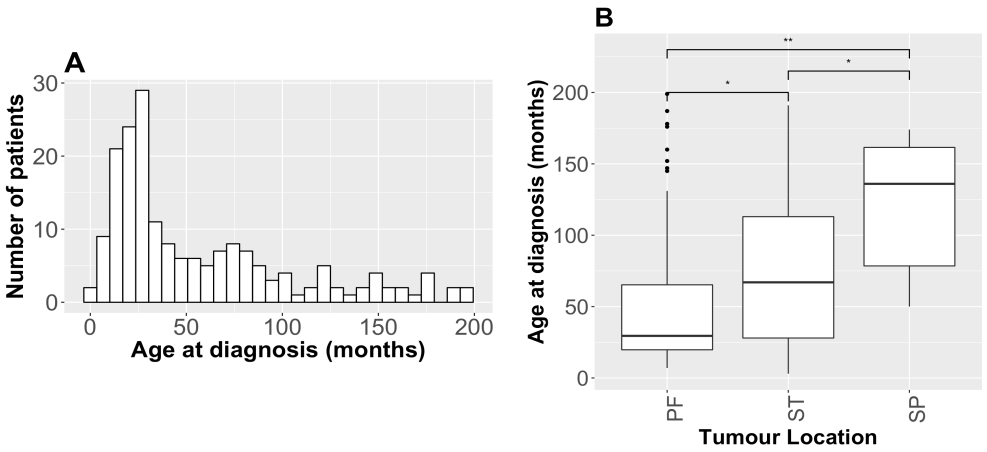


Figure 3-2: (A) Age at presentation for all ependymomas had a unimodal distribution. This is consistent with expectations for paediatric ependymomas. n=185. (B) Median age at presentation compared between tumour locations. Boxes represent 25th centile, median and 75th centiles. *: p<0.05, **: p<0.001.

3.3.3 Baseline characteristics of the non-recurrent cohort

Median OS for the non-recurrent cohort was not calculable as only seven (6%) patients had died. Median follow up duration for all patients was 82 months (0-182 months) (Figure 3-4A).

75 (64%) tumours occurred in the posterior fossa, 31 (27%) in the supratentorium and 10 (9%) in the spine. Location was unknown for two patients. Boys accounted for 55% of the cohort (n=65) and girls for 45% (n=45) (p=0.07). Gender was known for all patients (Table 3-1). Median age of diagnosis was 50 months (0 to 225 months), which was significantly older than the recurrent cohort (p=0.002).

58 (62%) primary tumours were WHO grade II, 36 (38%) WHO grade III and 25 had no information about grade. 65 (59%) of patients had GTR whilst 46 (41%) had STR. Extent of resection was unavailable for eight patients. 32 (50%) patients received radiotherapy. Radiotherapy data was unavailable in 55 cases. 67 patients received chemotherapy (58%). Chemotherapy data was unavailable for three children (Table 3-1).

Comparing baseline characteristics between the recurrent and non-recurrent cohorts, the recurrent group more likely to have STR (p=0.028) and chemotherapy (p=0.008) (Table 3-1).

3.3.4 Factors predicting risk of recurrence

To establish whether any factors were associated with risk of recurrence, a competing risks analysis was performed. In a univariate analysis WHO grade III ependymomas were associated with increased risk of recurrence when compared to grade II ependymomas (p=0.003), whilst older age at first diagnosis was associated with a decreased risk (p=0.026). Gender and tumour location were not associated with risk of recurrence.

GTR was associated with a lower recurrence risk (p=0.004) as was receipt of radiotherapy (p=0.020). Chemotherapy was associated with a significantly increased risk (p=0.001).

When all significant factors were included in a multivariate analysis only resection status remained significant (Subhazard Ratio = 0.636, p=0.013) (Table 3-2). Whilst GTR was associated with a decreased risk of recurrence, it did not completely prevent it. Of the 213 patients with GTR in the multivariate analysis, 115 (54%) experienced at least one recurrence.

	Sub Hazard Ratio	95% CI	P Value
Grade (III v II)	1.78	1.238-2.548	0.058
Age (continuous)	1	0.995-1.004	0.917
Resection (GTR Vs STR)	0.636	0.445-0.910	0.013
Radiotherapy (yes vs no)	0.670	0.409-1.097	0.111
Chemotherapy (yes vs no)	1.307	0.784-2.179	0.304

Table 3-2: Multivariate competing risks analysis for risk of recurrence. Death included as a competing risk. Only resection remained significantly associated with increased recurrence risk. n=213.

3.3.5 Patterns of recurrence

There was a median of two recurrences (range 1-8). Following first recurrence, 96 (51%) had further relapses, 45 (24%) died and 44 (23%) survived with no disease progression. Following each subsequent recurrence, the proportion of patients who recurred again was around 50%; this figure was fairly constant even at fifth, sixth and seventh recurrence (Figure 3-3). 89 (47%) patients had only one recurrence, 46 (24%) two, 22 (12%) three, 14 (7%) four, 7 (4%) five, 3 (2%) six, 1 (0.5%) seven and 1 (0.5%) eight (Figure 3-3). For some children, ependymoma became a chronically relapsing disease with 26 (14%) having in excess of three recurrences.

Data on metastasis was available for 82 patients at first recurrence, of these 20 (24%) had metastatic disease and 62 had isolated local disease (76%). Of those with metastatic disease, 10 had isolated distant disease and 10 had local and distant disease. The proportions of patients with metastatic disease were fairly constant for relapses two (31%), three (22%) and four (25%).

Patients who received radiotherapy after initial diagnosis had a median of one recurrence, compared with two recurrences for those who did not (p=0.04). There were no significant differences in number of recurrences for other baseline characteristics (Table 3-3).

Parameter		**Median recurrences**	**P Value**
Age	<3 years	2	0.07
	3+ years	1	
Gender	Male	2	0.10
	Female	1	
Extent of Resection	GTR	2	0.22
	STR	1	
Location	PF	1	0.80
	ST	1	
Grade	WHO II	2	0.06
	WHO III	1	
Radiotherapy at diagnosis	Yes	1	0.04
	No	2	

Table 3-3: Number of recurrences experienced for patients with different baseline characteristics. Only radiotherapy was associated with fewer recurrences. N=188.

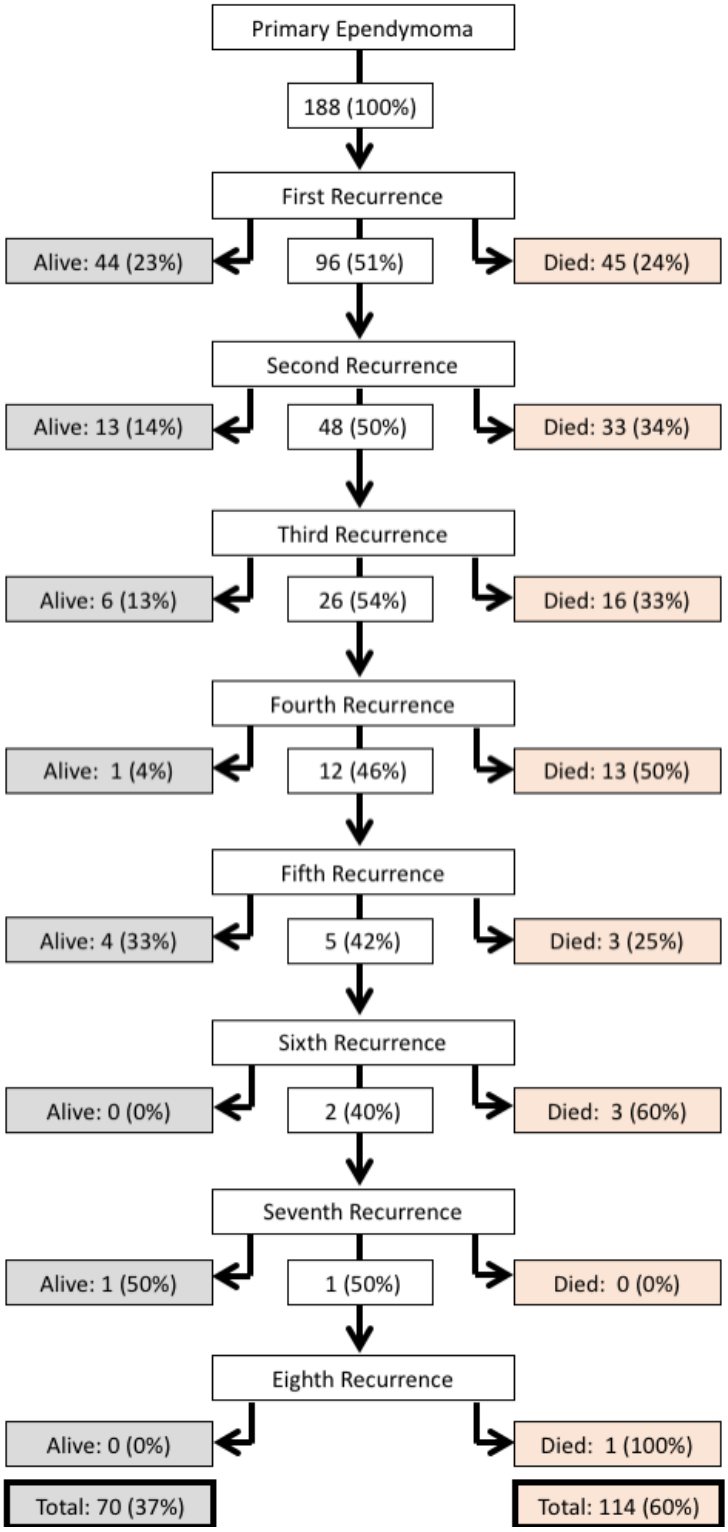


Figure 3-3: Flow diagram illustrating all recurrences for patients with tumours in any location. Between 40% and 50% of patients experienced further relapses after each relapse. 60% of the cohort were not alive at the end of the follow up period. N=188. Five patients were lost to follow up; three after recurrence 1 and two after recurrence 2. One patient was known to be alive and had at least one recurrence but with an unknown final number of recurrences. This patient was included in the total number of surviving patients.

3.3.6 Recurrence and overall survival

Survival analysis, stratified by recurrence status, demonstrated that recurrence was a key factor associated with OS. In the non-recurrent cohort only seven (6%) patients died. In the recurrent cohort 114 (60%) died (p<0.001) (Figure 3-4A). However, OS, from primary diagnosis, did not differ with increasing numbers of recurrences (p=0.066) (Figure 3-4B).

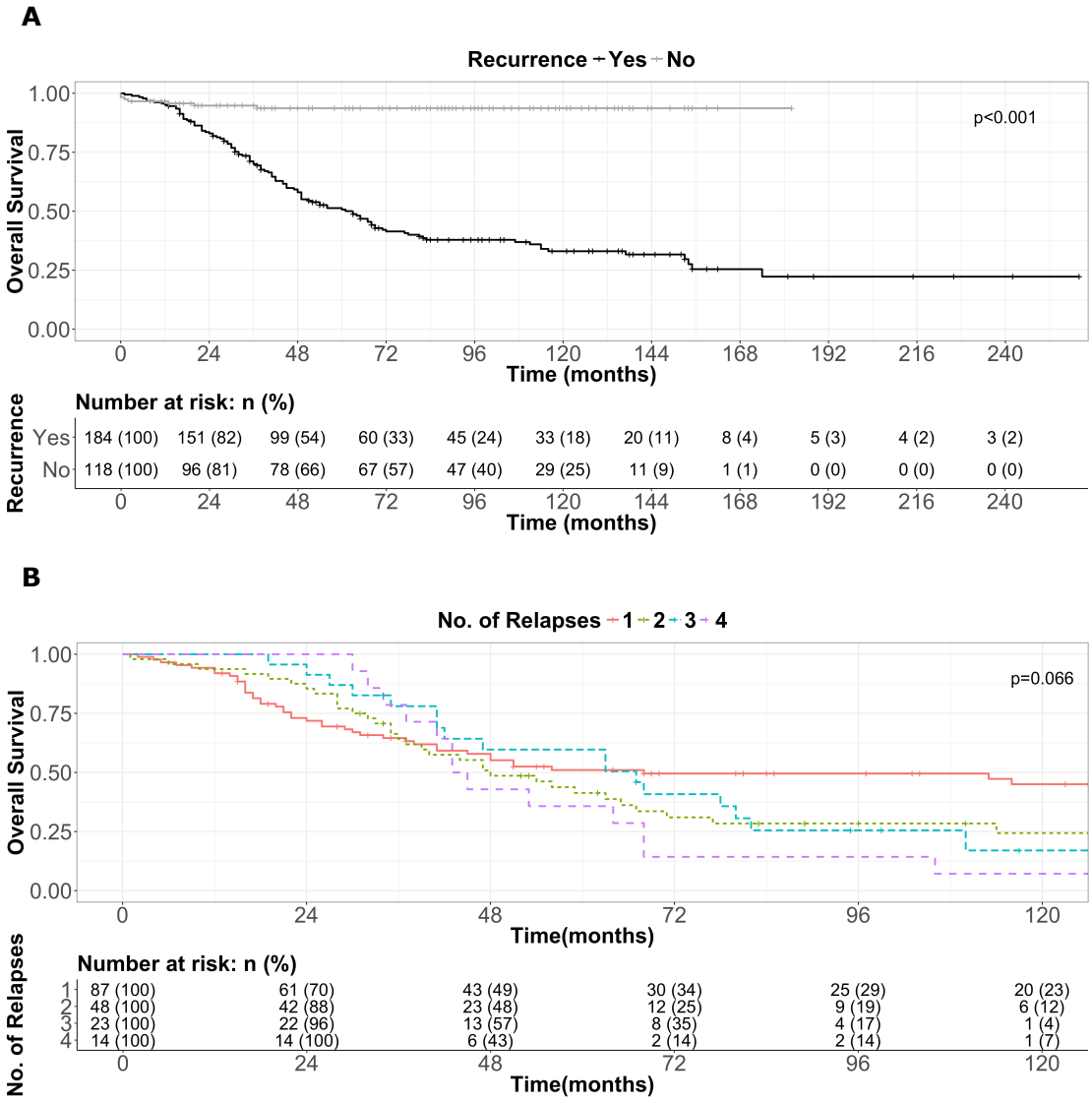


Figure 3-4: (A) Patients who experienced recurrence had worse OS than those who did not (p<0.001). (B) When stratified by number of recurrences, there were no significant differences in OS between those experiencing 1, 2, 3 or 4 episodes (p=0.066).

3.3.7 Overall survival in the recurrent cohort

3.3.7.1 At primary diagnosis

Age, gender, extent of resection, tumour location, grade and therapy were not associated with OS (Table 3-4). Following each recurrence, OS shortened (Table 3-5), falling to 5 months after the fifth recurrence.

Parameter		Cases tested	Median survival (months)	5-year survival (%)	10-year survival (%)	Supremum P Value
Age	<3yr	93	53	46	29	0.896
	>3yr	91	63	55	37	
Gender	M	103	53	48	29	0.548
	F	76	64	55	38	
Extent of Resection	STR	93	47	43	33	0.157
	GTR	76	67	59	31	
Location	PF	136	51	46	27	0.143
	ST	40	68	56	44	
	SP	7	-	100	67	
Grade	II	84	68	57	38	0.181
	III	76	48	44	23	
Radiotherapy at diagnosis	Yes	77	63	52	32	0.343
	No	101	53	48	31	
Chemotherapy at diagnosis	Yes	125	54	47	31	0.515
	No	46	63	59	34	

Table 3-4: OS from primary diagnosis stratified by epidemiological and treatment variables for the recurrent cohort. No factors were significantly associated with outcome.

Tumour Episode	Number of patients	Median OS (months)	95% CI (months)	5-year OS (%)	10-year OS (%)
P	184	61	48-71	50	33
R1	182	31	24-38	34	30
R2	92	17	13-25	19	18
R3	45	14	9-27	18	-
R4	24	11	4-20	14	-
R5	11	8	-	23	-
R6	4	7	-	0	-
R7	2	3.5	-	0	-
R8	1	4.0	-	0	-

Table 3-5: Median, 5- and 10-year OS following each recurrence. P: Primary, R: Recurrence.

3.3.7.2 At first recurrence

At first recurrence, 133 (95%) patients underwent further resection, data was unavailable for 48. This left only seven patients in the non-surgery group, limiting statistical power. There was no difference in OS for those who underwent further surgery compared to those who did not (p=0.059). No difference in OS was identified for extent of resection at first recurrence (p=0.126).

95 (66%) patients received radiotherapy, with data unavailable for 45. Those who received radiotherapy had a better OS (33 months vs 10 months, p=0.001). However, this survival benefit only lasted for the first six years following recurrence, raising the question as to whether radiotherapy at recurrence can lead to a sustained remission (Figure 3-5A). When patterns of radiotherapy were compared, there was no difference in OS depending on whether patients received reirradiation or first irradiation at recurrence (p=0.091) (Figure 3-5B). Treatment with chemotherapy was not associated with OS (p=0.829).

OS after first recurrence was associated with metastatic status and relapse location. 20/83 patients suffered a metastatic recurrence and had significantly worse median OS than those with local recurrence (20 months versus 44 months p=0.025) (Figure 3-5C). Those with just local disease fared best (median OS 44 months), those with distant and local disease had intermediate survival (median OS 20 months) and those with isolated distant disease had the worst outcomes (median OS 9 months) (p=0.013) (Figure 3-5D).

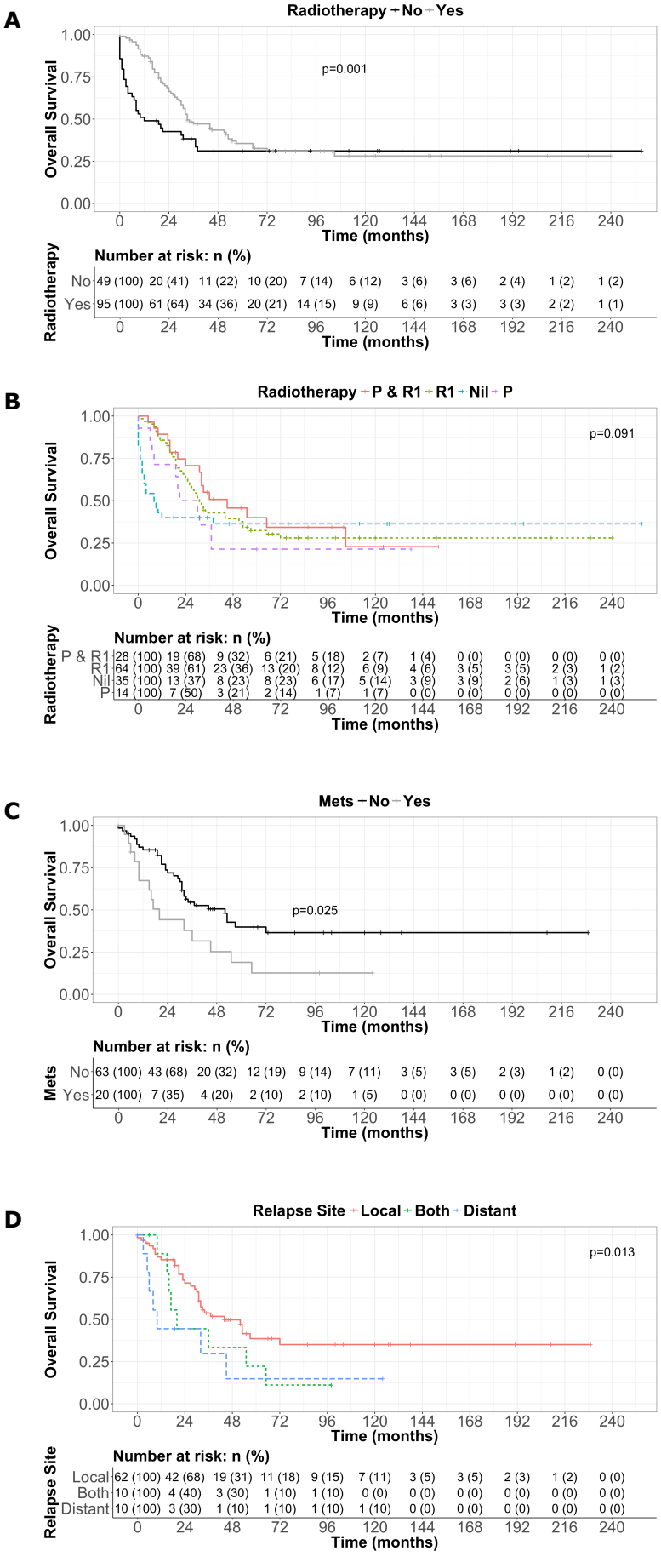

A
Radiotherapy No Yes
Overall Survival
p=0.001
Time (months)
Number at risk: n (%)
Radiotherapy
No 49 (100) 20 (41) 11 (22) 10 (20) 7 (14) 6 (12) 3 (6) 3 (6) 2 (4) 1 (2) 1 (2)
Yes 95 (100) 61 (64) 34 (36) 20 (21) 14 (15) 9 (9) 6 (6) 3 (3) 3 (3) 2 (2) 1 (1)
B
Radiotherapy P & R1 R1 Nil P
p=0.091
P & R1 28 (100) 19 (68) 9 (32) 6 (21) 5 (18) 2 (7) 1 (4) 0 (0) 0 (0) 0 (0) 0 (0)
R1 64 (100) 39 (61) 23 (36) 13 (20) 8 (12) 6 (9) 4 (6) 3 (5) 3 (5) 2 (3) 1 (2)
Nil 35 (100) 13 (37) 8 (23) 8 (23) 6 (17) 5 (14) 3 (9) 3 (9) 2 (6) 1 (3) 1 (3)
P 14 (100) 7 (50) 3 (21) 2 (14) 1 (7) 1 (7) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0)
C
Mets No Yes
p=0.025
Mets
No 63 (100) 43 (68) 20 (32) 12 (19) 9 (14) 7 (11) 3 (5) 3 (5) 2 (3) 1 (2) 0 (0)
Yes 20 (100) 7 (35) 4 (20) 2 (10) 2 (10) 1 (5) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0)
D
Relapse Site Local Both Distant
p=0.013
Relapse Site
Local 62 (100) 42 (68) 19 (31) 11 (18) 9 (15) 7 (11) 3 (5) 3 (5) 2 (3) 1 (2) 0 (0)
Both 10 (100) 4 (40) 3 (30) 1 (10) 1 (10) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0)
Distant 10 (100) 3 (30) 1 (10) 1 (10) 1 (10) 1 (10) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0)

Figure 3-5: (Previous page). (A) OS after first relapse was significantly associated with treatment with radiotherapy at relapse (p=0.001) (B) but not with whether this was first irradiation or reirradiation (p=0.091). (C) The presence of metastases was associated with poorer OS (p=0.025) (D) and patients with distant metastases had worse outcomes than those with isolated local disease (p=0.013).

3.3.8 Progression in the recurrent cohort

3.3.8.1 Time to first and subsequent recurrences

The median time to first recurrence was 17 months (0-149 months). The data was skewed towards early recurrence with 66% of tumours recurring less than two years after diagnosis. Within five years, 93% had recurred, rising to 98% by 10 years. Four recurrences occurred beyond 10 years after diagnosis, with the latest at 12.5 years (Figure 3-6A). Of these four, two were subsequently confirmed as ependymoma at recurrence by molecular profiling. The other two did not undergo molecular profiling but were confirmed by histopathological analysis (Chapter 4).

Following the first relapse subsequent recurrences occurred more rapidly, with a shortened median time to the next relapse (Figure 3-6B). Median time to first relapse was 17 months compared to 11, 7, 5, 11, 4, 2 and 1 for relapses 2, 3, 4, 5, 6, 7 and 8 respectively ($p<0.001$).

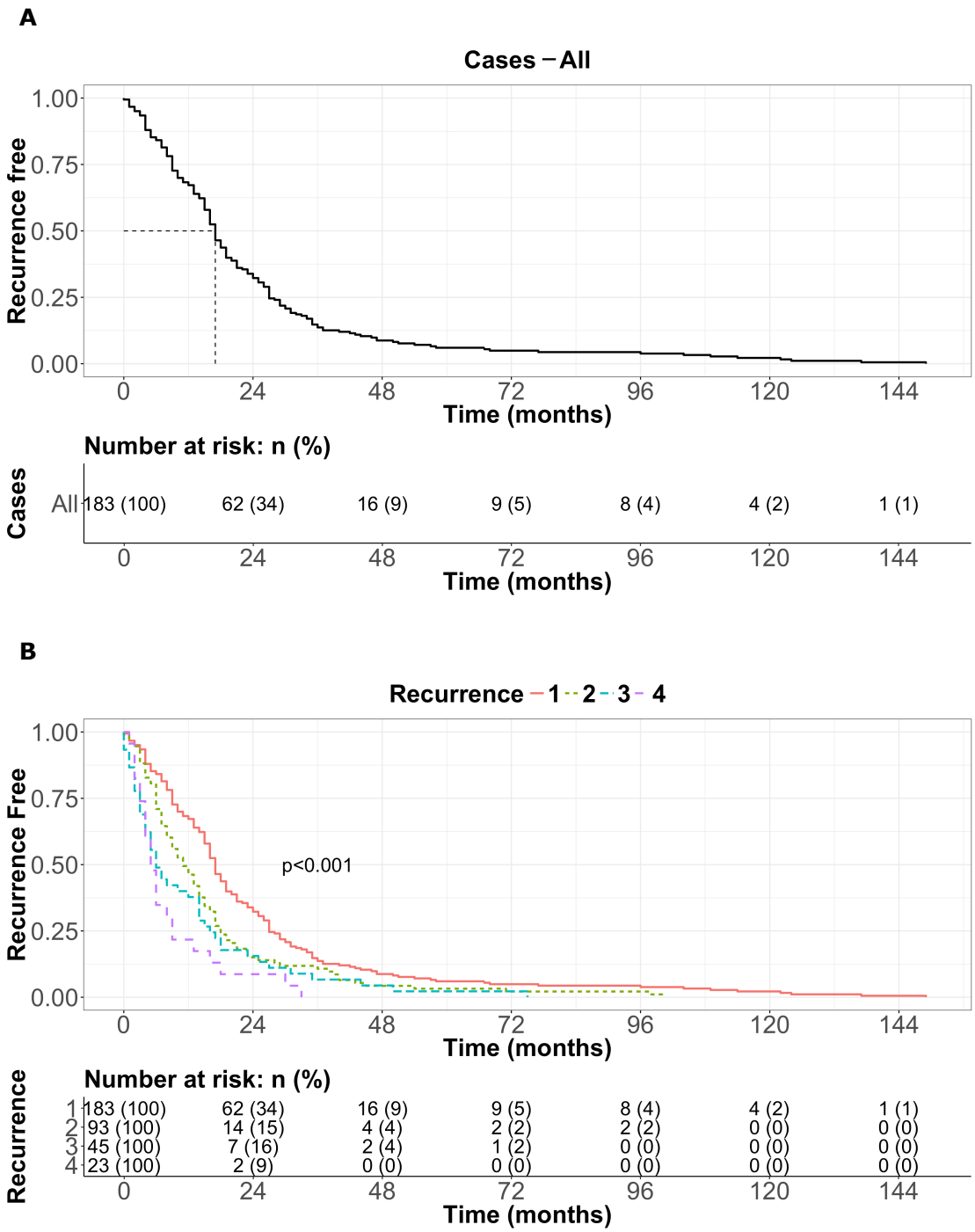


Figure 3-6: (A) Median time to first recurrence for the relapsed cohort was 17 months with a long tail of late recurrences. Dotted lines represent the median. (B) Median time to progression for recurrences 1, 2, 3 and 4 significantly shortened for each subsequent relapse ($p<0.001$). Recurrences beyond the fourth were not included in the statistical analysis in view of the small numbers available.

3.3.8.2 Factors associated with time to first recurrence

To further delineate the behaviour of recurrent ependymoma, time to progression data was stratified based on age, gender, extent of resection, receipt of radiotherapy or chemotherapy, tumour location and grade (Table 3-6, Figure 3-7).

Factors associated with a more rapid first recurrence were: tumour grade (WHO II 21 months versus WHO III 16 months, p=0.002) (Figure 3-7A), tumour location (PF 18 months versus ST 14 months, p=0.010) (Figure 3-7B), extent of initial resection (STR 15 months versus GTR 19 months, p=0.002) (Figure 3-7C), no radiotherapy at primary diagnosis (radiotherapy 20 months versus no radiotherapy 16 months, p=0.025) (Figure 3-7D) and chemotherapy at primary diagnosis (chemotherapy 16 months, no chemotherapy 22 months, p=0.049) (Figure 3-7E).

Parameter		**Median time to recurrence (months)**	**Median Difference (months)**	**95% CI (months)**	**P Value**
Age	<3yr	17	1	-3 to +5	0.658
	>3yr	17			
Gender	M	17	0	-3 to +5	0.862
	F	17			
Extent of Resection	STR	15	6	+2 to +10	0.002
	GTR	19			
Location	PF	18	6	+1 to +10	0.010
	ST	14			
	SP	11			
Grade	II	21	6	+2 to +11	0.002
	III	16			
Radiotherapy at diagnosis	Yes	20	5	+1 to +9	0.025
	No	16			
Chemotherapy at diagnosis	Yes	16	5	-10 to 0	0.049
	No	22			

Table 3-6: Univariate analysis of median difference in time to first recurrence for patients categorised by epidemiological features and therapy. Median differences compared by Mann-Whitney U test. Number of patients included in each comparison is indicated in the life tables in Figure 3-7.

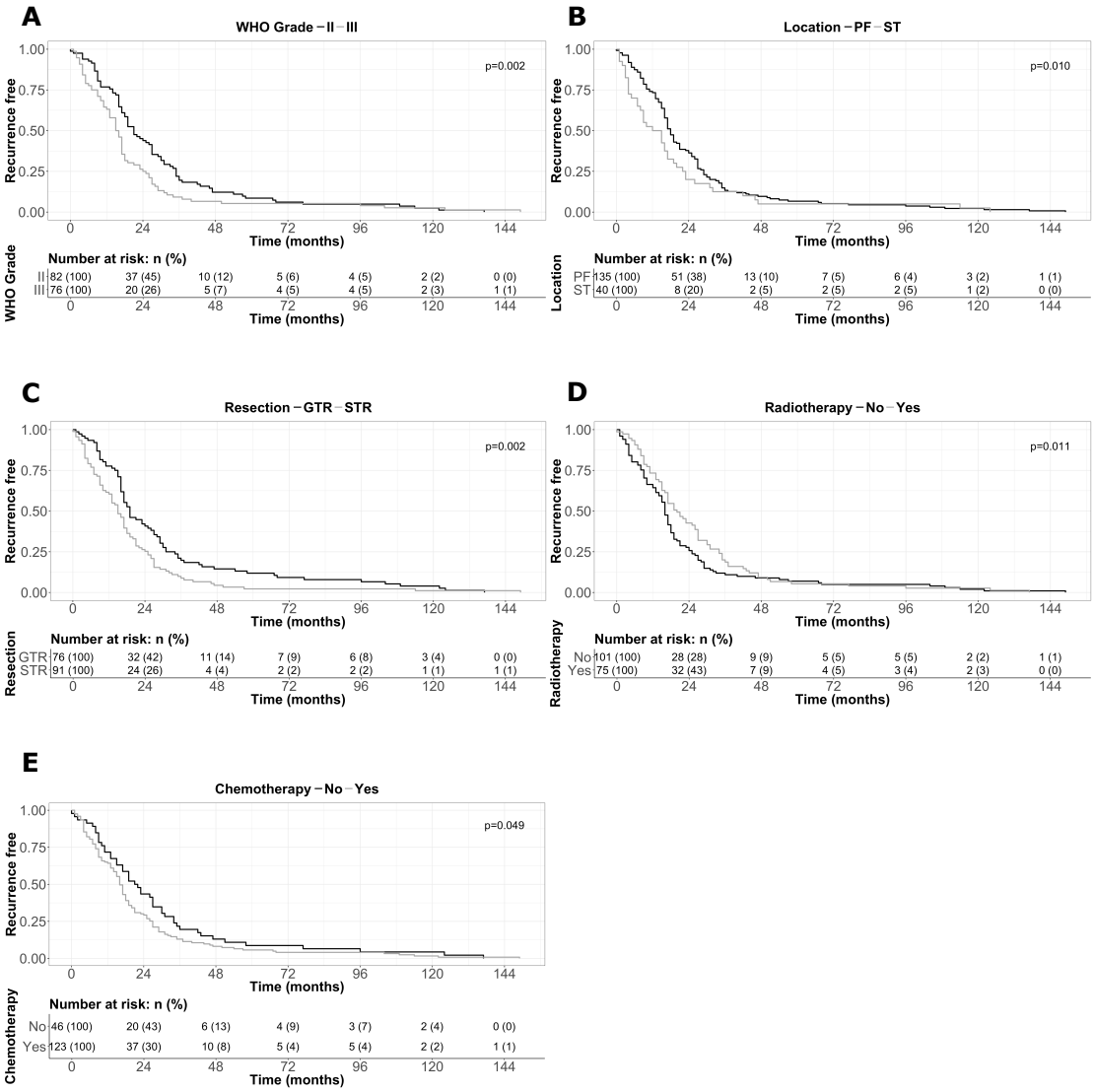


Figure 3-7: (A) Grade III ependymomas relapsed significantly faster than grade II tumours (p=0.002). (B) PF tumours relapsed significantly more slowly than ST tumours (p=0.010). (C) GTR was associated with a significantly lower relapse compared to STR (p=0.002). (D) Patients who received radiotherapy relapsed significantly more slowly than those who did not (p=0.025). (E) Patients who received chemotherapy relapsed significantly more quickly than those who did not (p=0.049).

Factors showing significance in the univariate analyses were included in a multivariate analysis using the CPH model. The assumption of proportional hazards was fulfilled (global Schoenfeld test, p=0.368).

The factors that remained associated with a faster time to first recurrence were grade III versus grade II tumour (p=0.002) and not receiving radiotherapy following initial diagnosis (p=0.048). Extent of resection, tumour location and receipt of chemotherapy all lost significance (Table 3-7).

Factor		Hazard Ratio	95% CI	P Value
Extent of resection	GTR	1.000	0.981-2.241	0.062
	STR	1.482		
Location	PF	1.000	0.867-2.037	0.192
	ST	1.328		
Tumour grade	II	1.000	1.328-1.258	0.002
	III	1.780		
Radiotherapy at primary diagnosis	Yes	0.598	0.358-0.996	0.048
	No	1.000		
Chemotherapy at primary diagnosis	Yes	1.000	0.472-1.658	0.701
	No	0.884		

Table 3-7: Multivariate analysis of risk factors associated with a more rapid time to first recurrence. Only tumour higher tumour grade and lack of treatment with radiotherapy remained significant. 132 cases included.

3.3.8.3 Event free survival after first recurrence

GTR and receipt of radiotherapy were associated with better EFS following first recurrence (GTR vs STR, median 21 versus 9.5 months, p=0.009 and radiotherapy versus none, median 19 versus 7 months, p<0.001) (Figure 3-8A and B). Chemotherapy was not associated with any difference (p=0.652).

The absence of metastatic disease was associated with a longer median EFS of 20 versus 12 months, p=0.008 (Figure 3-8C). Isolated distant disease had worse median EFS than either combined local and distant, or isolated local disease (20 vs 18 vs 5 months, p<0.001) (Figure 3-8D).

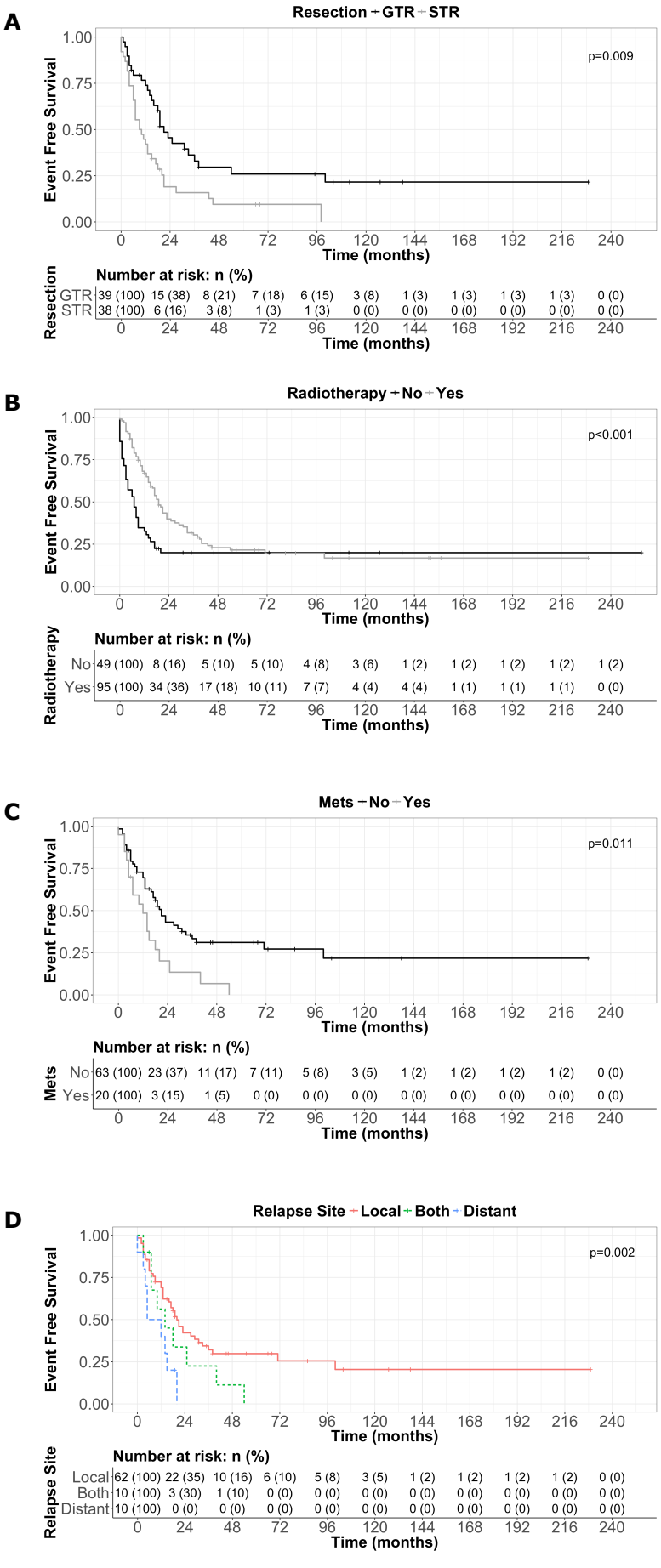

A
Resection GTR STR
Event Free Survival
1.00
0.75
0.50
0.25
0.00
p=0.009
0 24 48 72 96 120 144 168 192 216 240
Time (months)
Number at risk: n (%)
Resection
GTR 39 (100) 15 (38) 8 (21) 7 (18) 6 (15) 3 (8) 1 (3) 1 (3) 1 (3) 1 (3) 0 (0)
STR 38 (100) 6 (16) 3 (8) 1 (3) 1 (3) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0)
0 24 48 72 96 120 144 168 192 216 240
Time (months)
B
Radiotherapy No Yes
Event Free Survival
1.00
0.75
0.50
0.25
0.00
p<0.001
0 24 48 72 96 120 144 168 192 216 240
Time (months)
Number at risk: n (%)
Radiotherapy
No 49 (100) 8 (16) 5 (10) 5 (10) 4 (8) 3 (6) 1 (2) 1 (2) 1 (2) 1 (2) 1 (2)
Yes 95 (100) 34 (36) 17 (18) 10 (11) 7 (7) 4 (4) 4 (4) 1 (1) 1 (1) 1 (1) 0 (0)
0 24 48 72 96 120 144 168 192 216 240
Time (months)
C
Mets No Yes
Event Free Survival
1.00
0.75
0.50
0.25
0.00
p=0.011
0 24 48 72 96 120 144 168 192 216 240
Time (months)
Number at risk: n (%)
Mets
No 63 (100) 23 (37) 11 (17) 7 (11) 5 (8) 3 (5) 1 (2) 1 (2) 1 (2) 1 (2) 0 (0)
Yes 20 (100) 3 (15) 1 (5) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0)
0 24 48 72 96 120 144 168 192 216 240
Time (months)
D
Relapse Site Local Both Distant
Event Free Survival
1.00
0.75
0.50
0.25
0.00
p=0.002
0 24 48 72 96 120 144 168 192 216 240
Time (months)
Number at risk: n (%)
Relapse Site
Local 62 (100) 22 (35) 10 (16) 6 (10) 5 (8) 3 (5) 1 (2) 1 (2) 1 (2) 1 (2) 0 (0)
Both 10 (100) 3 (30) 1 (10) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0)
Distant 10 (100) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0) 0 (0)
0 24 48 72 96 120 144 168 192 216 240
Time (months)

Figure 3-8: (Previous page). EFS after first relapse was significantly associated with extent of resection (p=0.009) (A) and radiotherapy (p<0.001) (B). The presence of metastases was associated with poorer EFS (p=0.011) (C) and patients with distant metastases had worse outcomes than those with isolated local disease (p=0.002) (D).

3.3.9 Outcomes stratified by tumour location in the recurrent cohort

3.3.9.1 Baseline characteristics and relapse patterns

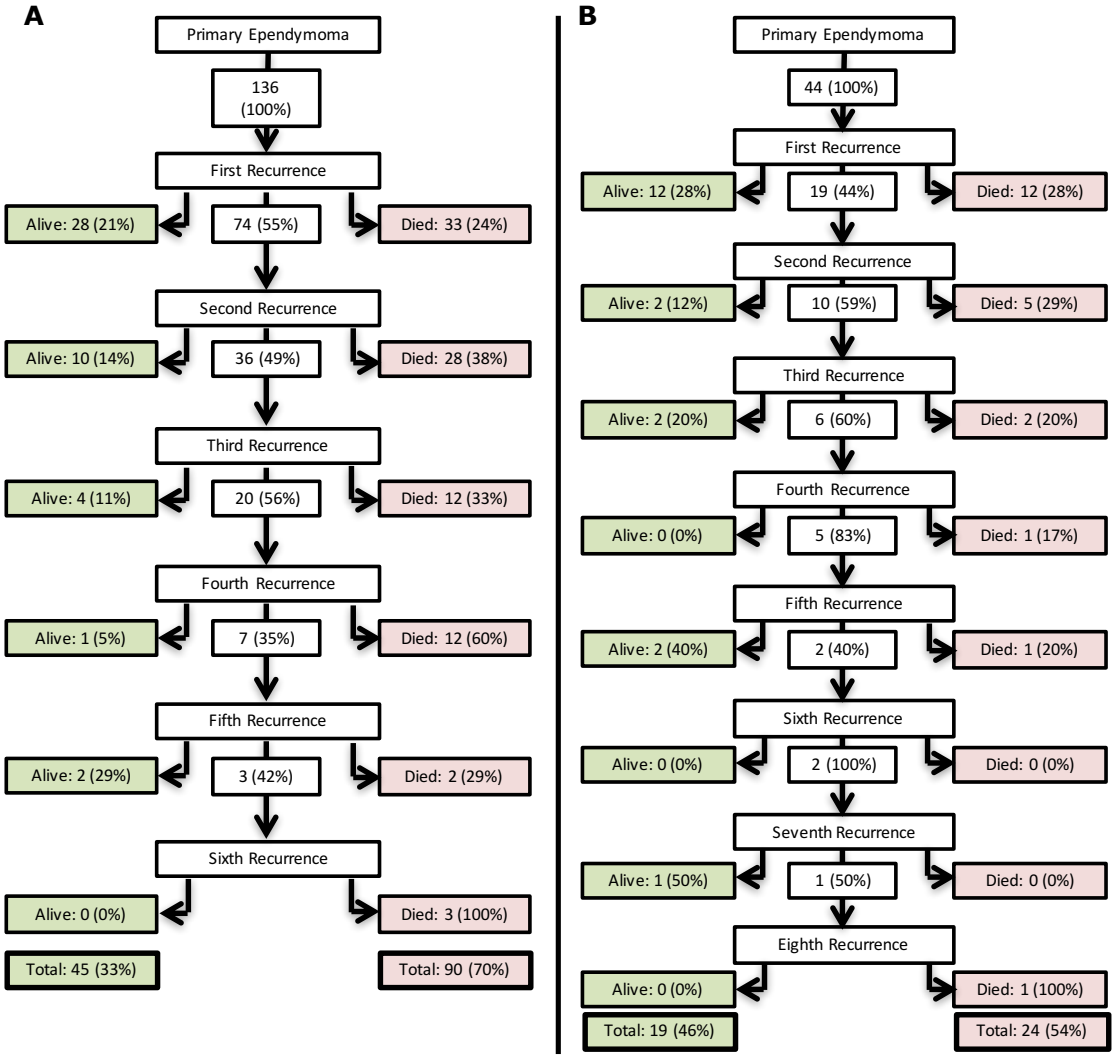


Figure 3-9: Flow diagram of patient outcomes for (A) posterior fossa (N=136) and (B) supratentorial ependymomas (N=44).

Different outcomes have been ascribed to tumours appearing in differing CNS locations (Pajtler et al., 2015; Ramaswamy et al., 2016), hence the need to consider tumours in different locations as separate cohorts. Flow diagrams were generated for those patients with PF (N=136) and ST (N=44) tumours (Figure 3-9). Numbers were insufficient to provide illustrations for spinal tumours. PF

tumours had a maximum of six recurrences compared to eight for ST tumours. The median number of recurrences for the PF cohort was two compared to one for the ST cohort (p=0.493). For initial recurrences, both groups demonstrated an approximately 50% recurrence rate after each relapse, consistent with the data presented for the overall cohort.

Recurrent tumours in the posterior fossa compared to the supratentorium had different baseline characteristics at primary diagnosis, which may have been masked in the analysis of the combined cohort. PF tumours that recurred were more likely to be from younger children than ST tumours (p<0.001). The recurrent ST tumours were less completely resected than the PF tumours (p=0.027) and were more likely to be of a higher grade (p=0.041). There was no difference in whether patients received chemotherapy or radiotherapy, or in the gender mix of the groups (Table 3-8).

Parameter		**Posterior Fossa (n=136)**		**Supratentorial (n=44)**		**Chi-Square P Value**
		Number	%	Number	%	
Age	<3 years	82	60	12	29	<0.001
	3+ years	54	40	29	71	
	NK	0	-	3	-	
Gender	Male	81	61	20	48	0.151
	Female	51	39	22	52	
	NK	4	-	2	-	
Extent of Resection	GTR	59	49	11	28	0.027
	STR	62	51	29	72	
	NK	15	-	4	-	
Grade	WHO II	66	56	14	36	0.041
	WHO III	51	44	25	64	
	NK	19	-	5	-	
Radiotherapy at diagnosis	Yes	52	40	21	53	0.201
	No	79	60	19	47	
	NK	5	-	4	-	
Chemotherapy at diagnosis	Yes	96	76	28	74	0.830
	No	30	24	10	26	
	NK	10	-	6	-	

Table 3-8: A comparison of the recurrent posterior fossa cohort with the recurrent supratentorial cohort. Significant differences were identified in the age, extent of resection achieved and tumour grade.

3.3.9.2 Risk of recurrence

Using both recurrent and non-recurrent cohorts the risk of recurrence was established for intracranial tumour locations. In the PF cohort, all variables associated with recurrence risk within the whole cohort remained significant, except for resection status and radiotherapy. For ST tumours, only the extent of resection remained significant (p=0.011) (Table 3-9).

		Sub Hazard Ratio	95% CI	P Value	Cases tested
PF	Age (continuous)	0.996	0.992-1.000	0.050	203
	Gender (female vs male)	0.846	0.597-1.198	0.347	200
	Grade (III vs II)	1.767	1.207-2.584	0.003	165
	Resection (GTR vs STR)	0.702	0.491-1.002	0.051	187
	Radiotherapy (yes vs no)	0.723	0.513-1.016	0.062	199
	Chemotherapy (yes vs no)	2.224	1.391-3.557	0.001	182
ST	Age (continuous)	0.998	0.993-1.003	0.450	74
	Gender (female vs male)	1.107	0.591-1.941	0.821	74
	Grade (III vs II)	1.295	0.690-2.432	0.421	62
	Resection (GTR vs STR)	0.424	0.218-0.824	0.011	72
	Radiotherapy (yes vs no)	0.633	0.346-1.158	0.138	72
	Chemotherapy (yes vs no)	1.542	0.751-3.166	0.238	67

Table 3-9: Univariate competing risks analysis for tumours based on intracranial location. Comparison between relapsed and non-recurrent cases with death before relapse as a competing risk.

3.3.9.3 Progression and survival

For the PF tumours, low tumour grade (p=0.044) (Figure 3-10A); GTR (p=0.001) (Figure 3-10B); receipt of radiotherapy (p=0.026) (Figure 3-10C); and non-receipt of chemotherapy (p=0.016) (Figure 3-10D), all remained significantly associated with a slower time to first recurrence. Age and gender were not significantly associated with time to first recurrence (p=0.806 and 0.278 respectively). For the ST group, only lower tumour grade was significantly associated with a slower time to first relapse (p=0.047) (Figure 3-11A).

No factors were associated with OS in the PF cohort. In the ST cohort, children under three years of age had significantly better OS (p=0.036) (Figure 3-11B).

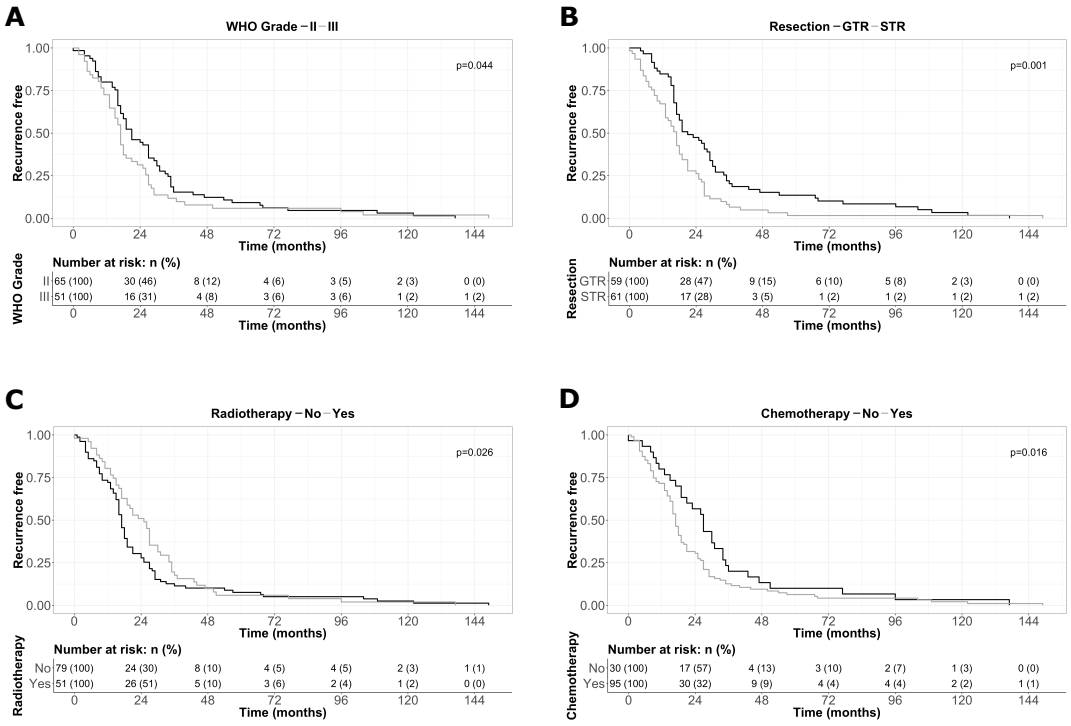


Figure 3-10: Factors associated with more rapid time to first relapse in posterior fossa tumours. (A) WHO grade III (p=0.044). (B) STR (p=0.001). (C) No radiotherapy (p=0.026). (D) Chemotherapy (p=0.016).

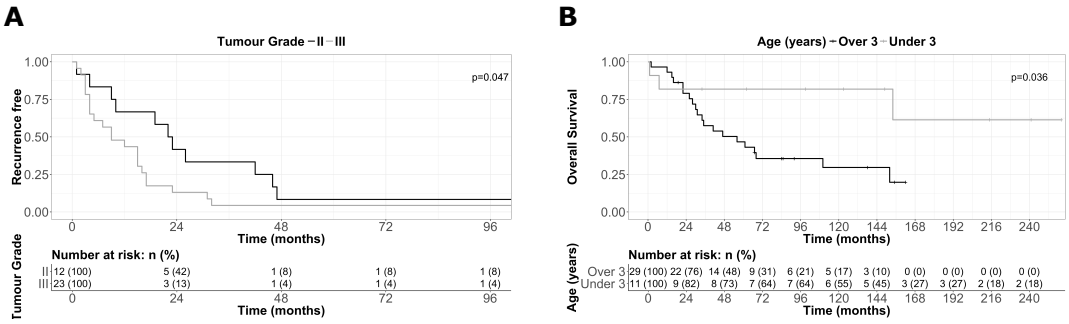


Figure 3-11: (A) Faster time to first relapse was significantly associated with higher grade in the supratentorial cohort. (B) OS for the supratentorial cohort stratified by age at diagnosis. Younger children had better survival (p=0.036).

3.4 Discussion and conclusions

This chapter presents a large analysis of recurrent paediatric ependymoma in order to better understand the natural history of the disease and the impact of various therapies. The large cohort, coupled with long follow up, has permitted a more comprehensive assessment of outcomes than previous studies. The cohort size has also allowed the investigation of the pattern of relapse stratified by tumour location and, as will be demonstrated in Chapter 4, molecular subgroup. The data highlights the dismal prognosis for children who recur and that lack of

early recurrence does not guarantee long-term disease-free survival. A number of epidemiological and disease factors were associated with risk and speed of recurrence; in particular, tumour grade and treatment with radiotherapy. However, once a relapse occurred, only radiotherapy significantly impacted OS in the short term. The poor overall outcomes shown and the lack of sustained therapeutic response to current interventions support a call for a better understanding of the underlying biology driving recurrent disease, hopefully leading to the development of new treatments.

In order to investigate a contemporaneous cohort, outcomes were initially analysed by decade of diagnosis. It was found that children treated after 1989 had better OS than those treated earlier. This allowed the cohort to be refined to take into account improvements in healthcare over time, including the development of intensive care techniques. It was encouraging that outcomes had improved and this was consistent with both recent cancer statistics (Cancer Research UK, 2015a) and another study that analysed outcomes by treatment decade (Snider et al., 2017). This highlights that epidemiological studies must be updated over time and that the conclusions of older studies, whilst scientifically valid when published, should not be relied upon indefinitely.

The survival data for children with recurrent ependymoma was sobering; recurrence was the key feature conferring poor prognosis. Only 34% and 30% of children were alive five and ten years after their first recurrence, in comparison with 94% in the non-recurrent cohort. These figures are consistent with other published work (Antony et al., 2014; Messahel et al., 2009; Zacharoulis et al., 2010). Given that the median age of diagnosis of first recurrence was under three years, with the OS described above, many children with recurrence will not reach adulthood. In view of this poor outlook, it is perhaps surprising that such little research has investigated recurrent paediatric ependymoma.

Previously published studies have usually focussed on the first, second and occasionally third relapse (Messahel et al., 2009; Vinchon et al., 2005). Within the recurrent cohort, children had a median of two, with a maximum of eight, recurrences. For this small proportion of children, ependymoma became a chronically relapsing disease. The risk of further recurrence remained high, following each relapse, at around half. Whilst a 50% relapse rate has previously been described for the primary to first recurrence (Messahel et al., 2009), this data has shown that it also applies to later relapses. Interestingly, there was no

significant difference in OS between children who relapsed only once and those that recurred multiple times, suggesting that recurrent disease itself confers a poorer prognosis rather than the number of episodes.

Whilst the majority of recurrences occurred within two years, a lack of early recurrence does not provide reassurance about long term prognosis; 34% of patients in the recurrent cohort relapsed after two years and four children relapsed beyond ten years after initial diagnosis. This finding has important implications for the duration of follow up for these children, including transition to adult services.

The time interval between relapse decreased with each recurrence. This was particularly significant for the time from first to second recurrence compared to primary to first recurrence, which was in direct contrast to a previously reported smaller case series (Hoffman et al., 2014b). There could be a number of reasons for this including: increasing biological aggressiveness; increased resistance of the tumour to therapy; decreasing host physiological reserve; or a lack of effective, evidence based therapy at recurrence.

GTR and radiotherapy at diagnosis were both significantly associated with decreased recurrence risk and delayed time to first recurrence. Although radiotherapy lost its association with risk of recurrence in multivariate analysis. However, neither treatment prevented recurrence. In fact, approximately half of patients who received GTR or radiotherapy still recurred, suggesting that close monitoring of these patients is required. This was consistent with another published study (Marinoff et al., 2017). Additionally, neither GTR nor radiotherapy were associated with improved OS from diagnosis in the recurrent cohort. However, when used at first recurrence, there was evidence that both radiotherapy and GTR were associated with better EFS, and radiotherapy with better OS. This supports other authors who have suggested improved outcomes with these interventions at relapse (Messahel et al., 2009; Vinchon et al., 2005).

Studies have suggested that reirradiation at recurrence may be of benefit (Bouffet et al., 2012; Lobón et al., 2016; Merchant et al., 2008). This study showed the importance of radiotherapy at first recurrence, irrespective of whether this was reirradiation or first irradiation, which differed from previously reported research (Zacharoulis et al., 2010). Disappointingly, none of these survival benefits persisted beyond the first few years post therapy. It is well described that radiotherapy can have significant adverse neurocognitive effects

on the developing brain (Spiegler et al., 2004; Yock et al., 2014). Its use must therefore be cautiously considered.

Unsurprisingly, the presence of metastatic disease at first relapse was associated with significantly worse outcomes than with local disease only. However, it is of interest that children with only metastatic disease had slightly worse outcomes than those with local and metastatic disease. This may suggest a difference in the biology of local versus distant disease and warrants further consideration.

There were associations between receipt of chemotherapy, recurrence risk and reduced time to progression, although this was lost in multivariate analyses. It is unlikely that chemotherapy directly caused tumour progression but more probable that it was given to children at higher risk of poor outcomes, for example those ineligible for radiotherapy or not achieving GTR. This finding was similar to that identified in another study (Zacharoulis et al., 2010).

Tumour grade was also associated with both risk of, and time to first, recurrence. Grade III tumours had an increased chance of recurrence and relapsed more quickly compared to grade II tumours. This is in agreement with one study (Goldwein et al., 1990), but in disagreement with another, where grade made no difference in time to first relapse (Messahel et al., 2009). There is much discord in the literature as to the reliability of histological grading and its association with outcome (Ellison et al., 2011). However, this study suggests that grade is associated with outcome, and is likely to be more reliable in view of the large number of samples included.

In univariate analysis, younger age was associated with increased recurrence risk. This could have been influenced by the different treatments given to younger children, or by the molecular composition of tumours in different age groups. The majority of children were treated in Europe, where radiotherapy is generally only given at primary diagnosis to children over three years of age. The younger children often receive chemotherapy (Grill et al., 2001; Grundy et al., 2007; Massimino et al., 2011). This was reflected in the loss of significance for age and treatment with chemotherapy in multivariate analysis which included treatment with radiotherapy.

In contrast, age at diagnosis did not affect time to first recurrence. This is surprising given that young children have been previously thought to have poorer outcomes (Jaing et al., 2004; Perilongo et al., 1997). This may be in part explained by the recurrent cohort having poorer outcomes thus masking any

difference between age groups. Even more surprising was the association between younger age and better outcome in ST ependymomas. This may be explained, to some extent, by the molecular subgroups present in the supratentorium. EPN_YAP tumours have been described as occurring in younger children and are associated with better outcomes (Pajtler et al., 2015).

Previous research has suggested ependymomas arising within different CNS compartments should be considered as biologically distinct groups (Johnson et al., 2010; Pajtler et al., 2015; Taylor et al., 2005). When PF and ST ependymomas were considered as separate cohorts, increased risk of recurrence was associated with higher grade and chemotherapy for PF tumours, and extent of resection for ST tumours. The location based cohorts did not have significantly different survival outcomes or recurrence risk. However, tumours arising in the posterior fossa relapsed at a slower rate than those occurring in the supratentorium. This is in disagreement with a study which identified no difference in time to first relapse based on tumour location (Messahel et al., 2009). The present study is based on a significantly larger dataset which may account for this discrepancy.

The strengths of this chapter of work were the large size of the dataset, not only of the cohort as a whole, but also when sub-classified by tumour location. It supported a number of previously published studies, and some of the differing findings may be related to increased statistical power from a larger cohort. Novel contributions included the recurrence patterns for tumour locations and the consistent 50% risk of relapse beyond first and second recurrences.

At recurrence, paediatric ependymoma is a highly aggressive disease with extremely poor outcomes. Many children who recur will not reach adulthood. Current treatments fail to provide sustained control of this tumour and cause significant morbidity. Given the developing knowledge of location-based subgroups, a better understanding of the biological basis of recurrent ependymoma is needed to guide targeted therapies.

4 DNA Methylation Analysis of the Recurrent Clinical Cohort

4.1 Introduction

The understanding of paediatric brain tumours is benefitting from the advent of a 'molecular era' of cancer research (Louis et al., 2016). Molecular profiling techniques are playing an enhanced role in discovery and definition of tumour biology as demonstrated by recent descriptions of seven medulloblastoma subgroups, BRAF fusions in pilocytic astrocytomas, and distinct molecular phenotypes of CNS primitive neuroectodermal tumours (CNS PNETs) (Jones et al., 2008; Schwalbe et al., 2017; Sturm et al., 2016; Taylor et al., 2012).

When this study began, in February 2015, ependymoma was defined by tumour location and grade. Subgroups had been proposed based on genomic imbalances (Dyer et al., 2002), and the concept of two PF groups had been introduced (Wani et al., 2012; Witt et al., 2011). In May 2015, Kristian Pajtler and colleagues published work on the molecular classification of paediatric ependymoma using DNA methylation profiling, including the profiling of nearly 50 matched primary and recurrent ependymomas (Pajtler et al., 2015). On the basis of this, the experimental protocol was adapted to include the generation of a molecular profile of the cohort, by extracting DNA from FFPE and FF tissue, from a subset of the primary and recurrent cohorts described in Chapter 3.

This chapter presents the DNA methylation profiling data from an independent cohort of matched primary and recurrent ependymomas. The aims of this chapter were to:

(1) Profile a subset of the recurrent paediatric ependymoma clinical cohort;
(2) Provide clinical annotation to a group of DNA methylation confirmed, recurrent paediatric ependymomas;
(3) Investigate the hypothesis that ependymomas do not change molecular subgroup at recurrence;
(4) Investigate whether tumours cluster by molecular profile or recurrence status.

4.2 Materials and methods

4.2.1 Generation of DNA methylation IDAT files

Processing of samples for DNA methylation profiling was conducted by collaborators at UCL genomics (London, UK). FFPE samples underwent a restoration process to repair fragmented ends of DNA sequences, before all samples were bisulfite converted. Samples were then run on Illumina Infinium 450k DNA methylation arrays (Illumina Inc, San Diego, USA), according to the UCL genomics summary protocol, as summarised below:

In a deep well plate, 500 ng of high quality bisulphite converted DNA was whole genome amplified overnight (37°C for 20-24 hours), then fragmented (37°C for one hour and fifteen minutes in a hybridisation oven), precipitated and resuspended in hybridisation buffer. Samples were hybridised onto BeadChips using a liquid handling robot (Freedom Evo, Tecan Ltd, Switzerland) and incubated at 48°C for 16-24 hours. The amplified and fragmented DNA samples anneal to locus specific 50mers (covalently linked to one of over 500,000 bead types) during hybridisation. Unhybridised and non-specifically hybridised DNA was washed away and the BeadChip was prepared for staining and extension. Single-base extension of the oligos on the BeadChip, using the captured DNA as a template, incorporated detectable labels on the BeadChip and determined the DNA methylation level of the query CpG sites. The process of single base extension and staining was carried out using the liquid handling robot. The staining procedure itself involved signal amplification by multi-layer immunohistochemical staining. Finally, the BeadChips were scanned using the iScan scanner with autoloader (Illumina Inc, San Diego, USA). Data was saved in the IDAT file format.

4.2.2 The DKFZ brain tumour classifier

The IDAT files generated by the 450k DNA methylation arrays were processed through the German Cancer Research Centre (DKFZ) brain tumour classifier version 11b2 (accessed at www.molecularneuropathology.org). Samples were compared to a reference cohort of brain tumour entities and a subgroup prediction was generated. The confidence of the prediction was indicated by a score between 0 and 1. DKFZ use a cut off of 0.9 to call a subgroup classification. However, this is very stringent and as this was a study aimed at generating new hypotheses, rather than making clinical diagnoses, a lower score of 0.5 was accepted.

4.2.3 The Chip Analysis Methylation Pipeline (ChAMP)

The Chip Analysis Methylation Pipeline (ChAMP) (Morris et al., 2014) was implemented from within the R statistical environment to cluster the data. Data was loaded into R using the Champ.load function (Aryee et al., 2014; Fortin et al., 2017), followed by quality control analysis using champ.QC (Morris et al., 2014). Data was then normalised with beta-mixture quantile normalisation (BMIQ) using the champ.norm function (Teschendorff et al., 2013). R base functions were then used to produce multi-dimensional scaling plots (R Core Team, 2014). The script used for this analysis can be found in Appendix 3.

4.3 Results

4.3.1 The recurrent methylation cohort

DNA methylation profiles were generated for 253 tissue samples in the recurrent clinical cohort; 120 primary tumours (64%) and 133 recurrences (80 first, 31 second, 13 third, 5 fourth, 2 fifth recurrences plus 1 sixth and 1 seventh recurrence). This represented 64% of all the primary tumours and 45% of all the tumours (primary or recurrent) in the recurrent clinical cohort (Chapter 3). The patients in the methylation cohort showed no significant baseline differences with respect to age (p=0.813), gender (p=0.905), extent of resection (p=0.460), tumour location (p=0.587), grade (0.627), treatment with radiotherapy (p=0.903) or chemotherapy (p=1.000) when compared to the recurrent clinical cohort at primary diagnosis.

4.3.2 Classifier scores

The median classifier score for all of the samples was 0.990 (range 0.05-1.00). There was a difference in median scores between the primary and recurrent samples, with the primary samples generally exhibiting higher scores (median 1.00 versus 0.980, p=0.025).

A score of 0.5 was used as a cut off for defining a subgroup prediction; excluding 14 primary and 23 recurrent tumours from further analysis.

4.3.3 DNA methylation subgroup predictions

After excluding samples with low classifier scores, the primary tumours consisted of 77 (73%) EPN_PFA tumours; 13 (12%) EPN_RELA tumours; 2 (2%) EPN_PFB tumours, 4 (4%) EPN_YAP tumours and 3 (3%) EPN_MPE tumours. There were 7 (7%) non-ependymoma molecular diagnoses (Figure 4-1A and B). The recurrent tumours showed a similar distribution of DNA methylation profiles. The proportion

of non-ependymoma entities rose towards the higher recurrence numbers; this was likely biased by one case of HGNET_MN1 in which the same patient experienced multiple relapses (Figure 4-1B).

A

Methylation Prediction N (%)	Primary	R1	R2	R3	R4	R5	R6	Total
EPN_PFA	77 (73)	48 (71)	12 (44)	7 (54)	3 (43)	1 (50)	0	148 (66)
EPN_RELA	13 (12)	10 (15)	8 (30)	2 (15)	2 (29)	0	0	35 (16)
EPN_PFB	2 (2)	1 (1)	1 (4)	0 (0)	0 (0)	0	0	4 (2)
EPN_YAP	4 (4)	3 (4)	2 (7)	0 (0)	1 (14)	0	0	10 (4)
EPN_MPE	3 (3)	1 (1)	0 (0)	0 (0)	0 (0)	0	0	4 (2)
EPN_SPINE_SE	0 (0)	1 (1)	0 (0)	0 (0)	0 (0)	0	0	1 (0)
Other	7 (7)	4 (6)	4 (15)	4 (31)	1 (14)	1 (50)	1 (100)	22 (10)
Total	106 (100)	68 (100)	27 (100)	13 (100)	7 (100)	2 (100)	1 (100)	224 (100)

B

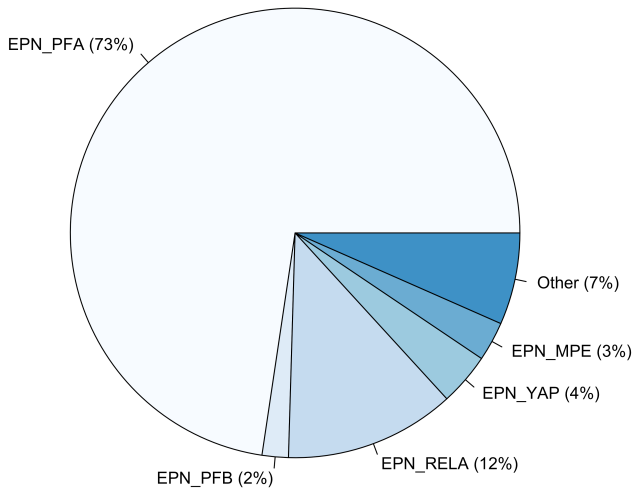


Figure 4-1: (A) Table illustrating the DNA methylation subgroup predictions at each episode of relapse (R). 'Other' includes all non-ependymoma entities and those samples which were predicted to be normal brain. (B) Distribution of DNA methylation subgroup predictions for the 106 primary paediatric ependymomas with adequate classifier scores (>0.5) in the dataset profiled by Illumina Infinium 450k methylation arrays. EPN_PFA was the most common tumour in this cohort.

4.3.4 Clinical correlates of the DNA methylation subgroups

DNA methylation subgroups have been associated with specific clinical features including location, age and gender (Pajtler et al., 2015). In order to validate the DNA methylation dataset described here and to describe the behaviour of the

DNA methylation subgroups when they recur, these clinical aspects were examined.

74 of the 77 (96%) EPN_PFA were located in the posterior fossa with the remaining three (4%) in the supratentorium. All 13 EPN_RELA tumours were in the supratentorium. EPN_PFA tumours were significantly more likely to occur in the posterior fossa and EPN_RELA tumours were significantly more likely to occur in the supratentorium ($p<0.001$). All four of the EPN_YAP tumours were in the supratentorium and both EPN_PFB tumours were located in the posterior fossa. All three EPN_MPE tumours were found in the spine.

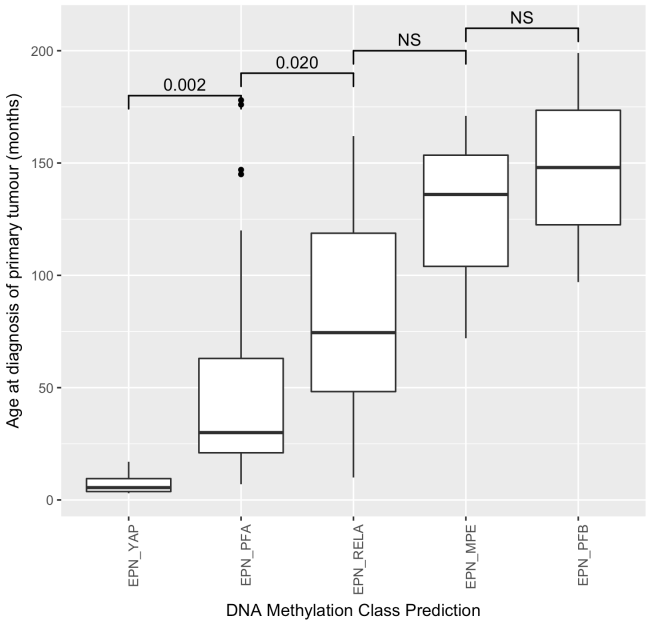


Figure 4-2: Age distributions for the tumours with DNA methylation subgroup predictions. Children with EPN_YAP tumours were the youngest, with increasing age for the other tumour types. Numbers indicate p values. NS: Not significant.

The median age for the EPN_PFA patients was 30 months, EPN_RELA 75 months, EPN_YAP 6 months, EPN_PFB 148 months and EPN_MPE 136 months. The children with EPN_YAP tumours were significantly younger than those with EPN_PFA tumours ($p=0.002$). The children with EPN_PFA tumours were significantly younger than those with EPN_RELA tumours ($p=0.020$). There were no significant differences between the ages of the children with EPN_RELA tumours and EPN_PFB or EPN_MPE tumours, but the numbers for this analysis were small (Figure 4-2).

There were no significant differences in the gender distribution of the DNA methylation subgroups. 42 (59%, p=0.211) of the EPN_PFA patients were male as were 8 (53%, p=0.290) of the EPN_RELA patients, 0 (0%, p=0.067) of the EPN_YAP patients and 1 (50%, p=0.500) of the EPN_PFB patients.

4.3.5 Outcomes for subgroups in the recurrent methylation cohort

Curves were generated to illustrate differences in OS between DNA methylation subgroups. Numbers were small for the EPN_PFB, EPN_MPE and EPN_YAP tumours and therefore, for these groups, statistics were not performed. However, the curves demonstrated that patients with EPN_PFB and EPN_YAP tumours appeared to have the best OS; EPN_MPE tumours had intermediate OS; and EPN_PFA and EPN_RELA tumours had very poor OS (Figure 4-3).

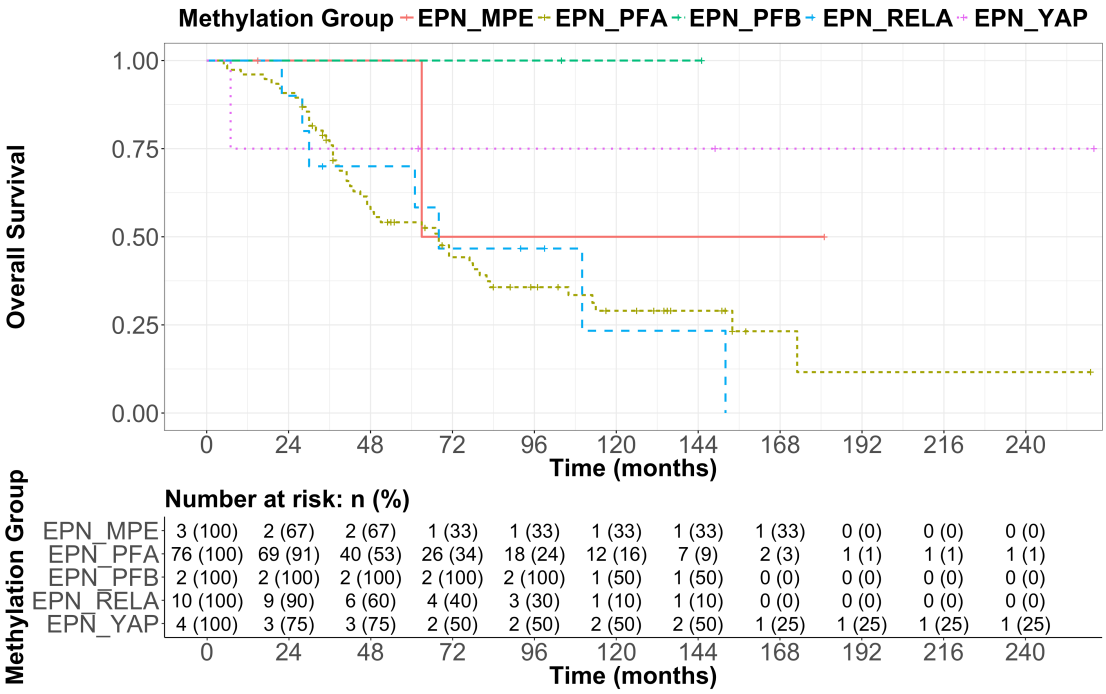


Figure 4-3: Overall survival of the ependymoma DNA methylation subgroups in the recurrent cohort. EPN_PFA and EPN_RELA were associated with poor outcomes, EPN_MPE with intermediate outcomes and EPN_YAP and EPN_PFB with better outcomes. Statistics not performed in view of the low numbers in three of the five groups.

Further analysis was performed on the two most common paediatric subgroups, EPN_PFA and EPN_RELA.

Median OS for EPN_PFA was 67 months, compared with 110 months for EPN_RELA. However, there was no statistically significant difference in OS between the two groups (p=0.763) (Figure 4-4A).

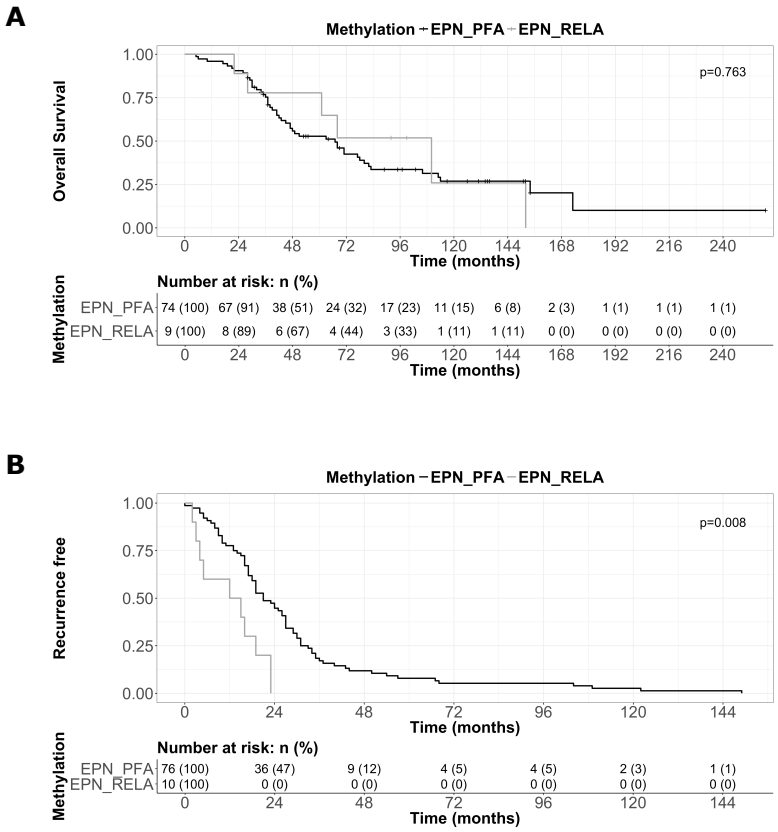


Figure 4-4: (A) Overall survival for the relapsed EPN_PFA and EPN_RELA groups was not significantly different (p=0.763). (B) Time to first relapse for the EPN_PFA and EPN_RELA groups demonstrated that EPN_RELA tumours recurred significantly more quickly (p=0.008).

EPN_RELA tumours recurred significantly more quickly than EPN_PFA tumours (median EPN_PFA 21 months versus median EPN_RELA 15 months, p=0.008) (Figure 4-4B).

EPN_PFA and EPN_RELA subgroups were analysed to determine whether grade, extent of resection, receipt of radiotherapy or chemotherapy, age or gender were associated with TTP or OS at primary diagnosis. The only positive association was seen between tumour grade and time to recurrence in EPN_PFA (median 17.5 months for grade III and 27 months for grade II tumours, p=0.043) (Table 4-1).

Parameter		EPN_PFA			EPN_RELA		
		Cases	TTP (months)	P-value	Cases	TTP (months)	P-value
Age	<3yr	43	23	0.562	2	4	0.239
	>3yr	34	21		8	16	
Gender	M	45	21	0.783	6	18	0.325
	F	31	25		5	5	
Extent of Resection	GTR	38	25	0.117	4	18	0.521
	STR	31	20		6	9	
Grade	II	38	27	0.043	4	19	0.424
	III	34	18		9	12	
Radiotherapy at diagnosis	Yes	32	26	0.637	5	15	1.000
	No	41	19		5	12	
Chemotherapy at diagnosis	Yes	51	21	0.900	5	5	0.623
	No	18	25		4	17	

Table 4-1: Factors associated with time to first progression for the EPN_PFA and EPN_RELA cohort. The only factor significant association was between higher grade and a more rapid first relapse in EPN_PFA tumours (p=0.043).

EPN_PFA and EPN_RELA tumours were next analysed to determine whether, consistent with data for the recurrent clinical cohort, surgery and radiotherapy were associated with better outcomes after first relapse. Using tumours with an EPN_PFA or EPN_RELA diagnosis at primary and available clinical information (n=60 for EPN_PFA and n=10 for EPN_RELA), it was found that radiotherapy at recurrence for EPN_PFA tumours was associated with better OS (median 10.5 vs 32 months, p=0.036) and EFS (median 10 vs 37 months, p= 0.013) (Figure 4-5). Extent of surgery was not associated with outcome in either subgroup and radiotherapy was not associated with outcome in the EPN_RELA group.

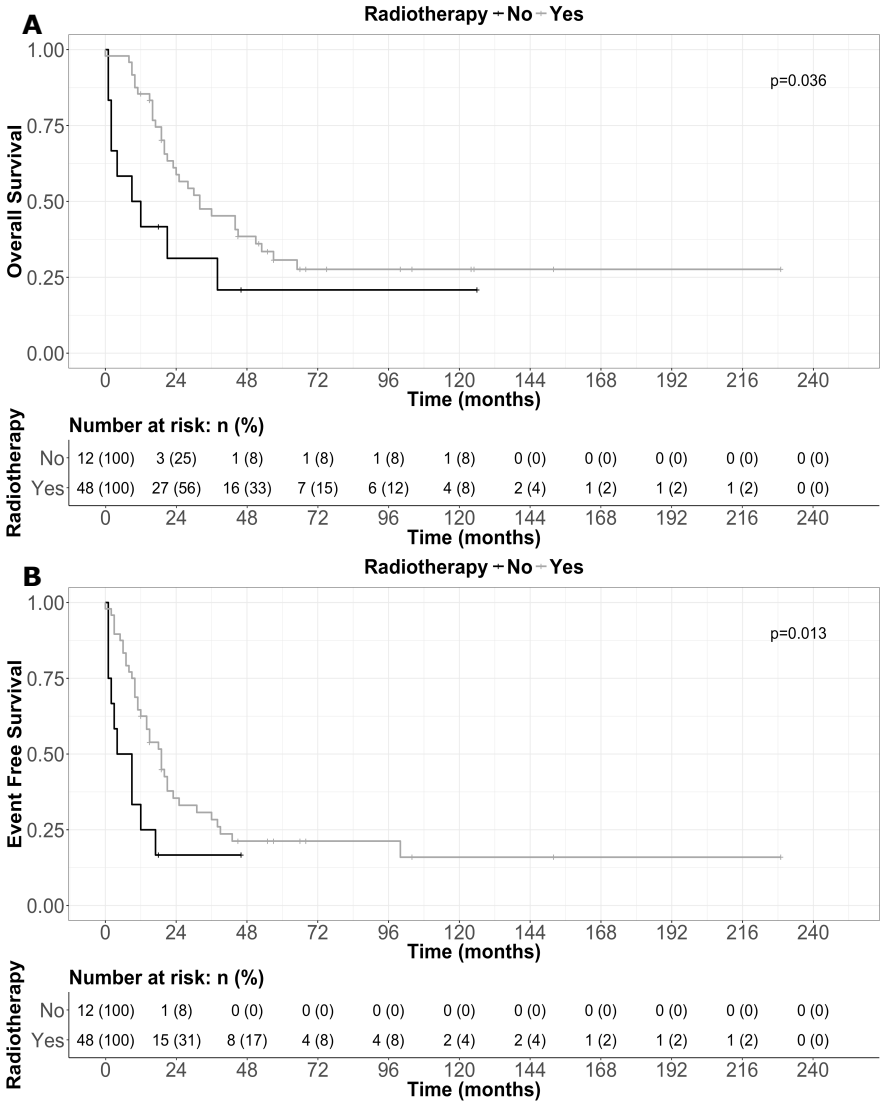


Figure 4-5: Survival outcomes at first recurrence for patients with EPN_PFA. OS (A) and EFS (B) were significantly worse for patients who did not receive radiotherapy at first recurrence

4.3.6 Recurrence in 'good prognosis' subgroups

Studies have indicated that EPN_PFA and EPN_RELA have worse outcomes than EPN_PFB, EPN_YAP and EPN_MPE (Pajtler et al., 2015; Ramaswamy et al., 2016). In the recurrent cohort, all three of these better prognosis subgroups were represented: two EPN_PFB; three EPN_MPE; and four EPN_YAP. The clinical histories of these cases were reviewed (Table 4-2). Despite recurring, only two out of the nine patients had died by the end of follow up. Median follow up for the living patients was relatively long at 114 months. Six of the patients had STR, three had GTR and two were treated with radiotherapy. OS appeared to be better for these subgroups than EPN_PFA and EPN_RELA (5-year survival for EPN_PFB, EPN_MPE and EPN_YAP 100%, 50% and 75% respectively) but this was not tested for significance in view of the low numbers.

ID	Methylation Prediction	Gender	Age (months)	Relapses	Location	Grade	Resection	Radiotherapy	Chemotherapy	Outcome	Follow up (months)
Epend223	EPN_PFB	F	97	1	PF	II	STR	Yes	Yes	A	104
Epend316	EPN_PFB	M	199	1	PF	II	STR	Yes	Yes	A	188
Epend094	EPN_MPE	F	136	1	SP	III	GTR	No	No	A	181
Epend129	EPN_MPE	F	72	3	SP	II	GTR	No	No	D	63
Epend152	EPN_MPE	M	171	1	SP	II	GTR	No	No	A	14
Epend017	EPN_YAP	F	7	1	ST	III	STR	No	Yes	D	7
Epend103	EPN_YAP	F	17	1	ST	III	STR	No	Yes	A	260
Epend124	EPN_YAP	F	3	2	ST	III	STR	No	Yes	A	62
Epend171	EPN_YAP	F	4	1	ST	II	STR	No	Yes	A	114

Table 4-2: Summary of the clinical outcomes of the better prognosis DNA methylation subgroups. All but two patients were alive at the end of follow up despite all patients suffering from at least one recurrence. SP: Spinal, A: Alive, D: Dead.

4.3.7 DNA methylation predictions in matched primary and recurrent cases

Matched primary and first recurrent pairs were available for 56 cases, with DNA methylation subgroup matching in 54 (98%) (Appendix 1). Using the genotypes generated by the DKFZ classifier, it was confirmed that all paired samples were from the same patient. 39 cases were EPN_PFA, 9 EPN_RELA, 1 EPN_PFB, 2 EPN_YAP and 1 EPN_MPE. Two of the matching cases were non-ependymoma entities. In the two cases where the DNA methylation group changed, one switched from EPN_PFA to DNET and one from DNET to EPN_RELA. Paired results were available for two of the four cases that first recurred more than 10 years after initial diagnosis. In both cases the tumours were classified as EPN_PFA at primary and recurrence, consistent with ependymoma recurrence rather than a treatment induced second malignancy.

4.3.8 Clustering DNA methylation data

Given the lack of evidence of a change in DNA methylation subgroup at first recurrence in this and a previous study (Pajtler et al., 2015), it was hypothesised that samples would cluster according to DNA methylation subgroup assignment. This was confirmed by performing multidimensional scaling of all primary and first recurrent tumours (Figure 4-6). It was clearly demonstrated that, with the exception of one EPN_PFB tumour, samples clustered according to their DNA methylation subgroup prediction rather than recurrence status, thus supporting the hypothesis.

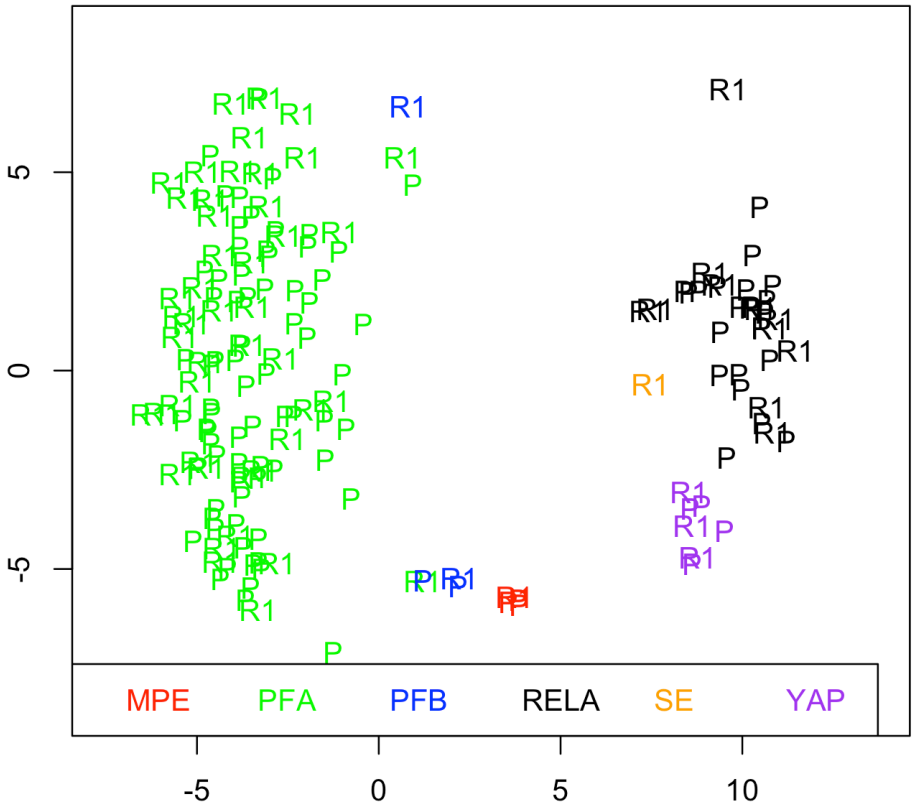


Figure 4-6: Multidimensional scaling plot based on the 1000 most variable DNA methylation probes of all primary and first recurrent ependymoma samples demonstrating clustering by subgroup rather than recurrence status. P:Primary. R1: Recurrence 1.

EPN_PFA represented the largest group of DNA methylation samples and formed the majority of the PF samples. Evidence for multiple EPN_PFA subgroups has been recently presented at an international conference (Pajtler et al., 2017). In order to investigate for evidence of EPN_PFA subgroups within this dataset, multidimensional scaling of the EPN_PFA samples was undertaken. Two clusters of samples were identified which may represent the two major PFA subgroups (EPN_PFA1 and EPN_PFA2) (Figure 4-7). Clustering into one of the two groups occurred irrespective of whether the tumour was primary or recurrent. Further investigation into these groups was not undertaken as it did not represent one of the main aims of the study; however extensive investigation of this dataset has been undertaken elsewhere (Pajtler et al., 2017).

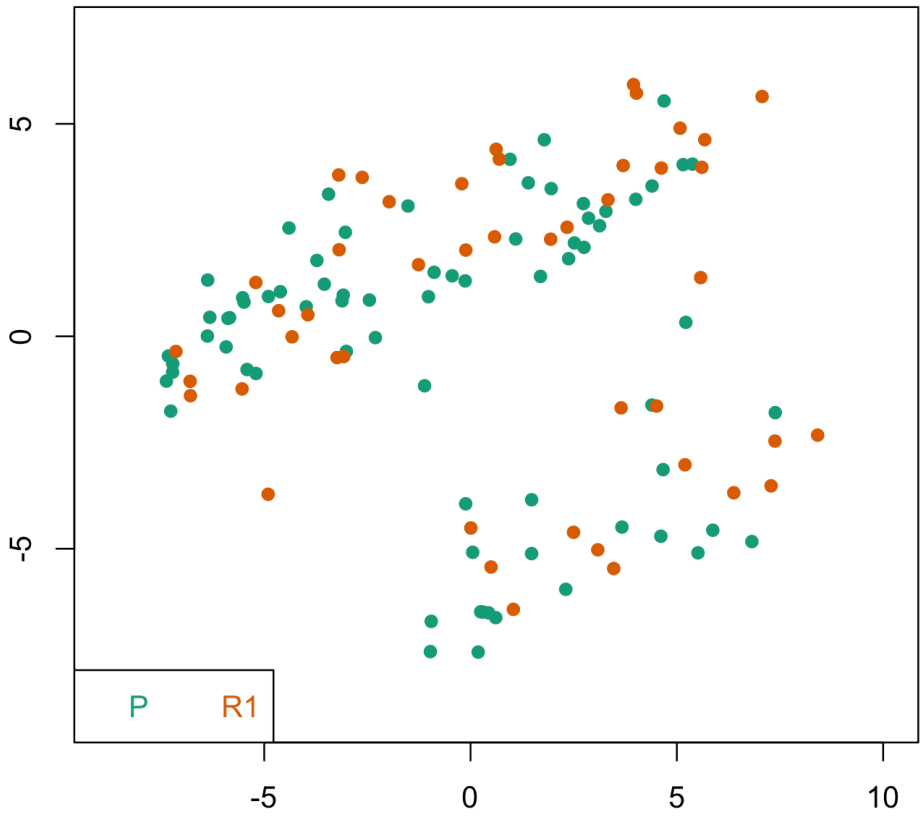


Figure 4-7: Multidimensional scaling based on 1000 most variable DNA methylation probes of all EPN_PFA primary (P - green) and first recurrent (R1 - orange) samples indicating two potential subgroups of EPN_PFA ependymoma.

4.4 Discussion and conclusions

Molecular subgrouping in ependymoma has advanced rapidly over the last few years and a shared understanding of how these subgroups behave clinically is required. Whilst one study has investigated a cohort with matched primary and recurrent DNA methylation profiles in both adults and children, this was limited to comparing subgroup assignment at recurrence with no additional clinical analysis specific to the relapsed group (Pajtler et al., 2015). This chapter is believed to be the first piece of work in which a substantial cohort of purely recurrent paediatric ependymomas have both clinical data and DNA methylation annotations.

The primary tumours for 56% (n=106) of the original clinical cohort were annotated with a DNA methylation prediction with adequate classifier score. This cohort was highly representative of the clinical recurrent cohort (Chapter 3) in terms of baseline characteristics. This cohort was also consistent with current knowledge of the DNA methylation subgroups, namely: age, location and gender, suggesting that the DNA methylation profiling was robust (Pajtler et al., 2015). The only exception to this was the finding that three (4%) EPN_PFA tumours

were reported to have occurred in the supratentorium. Possible explanations for this are that these tumours did arise supratentorially, perhaps originating close to the edge of the infratentorium or that these tumours were misclassified at the time of diagnosis. Because of the multicentre nature of the study and the access to imaging available it was not possible to perform further review on these cases.

The majority (98%) of ependymomas remained in the same subgroup from primary diagnosis to first recurrence, consistent with Pajtler et al. 2015. Importantly, tumours recurring after long time intervals did not change subgroup; late recurrences were still classified as ependymoma rather than treatment induced secondary malignancies. One possible reason for the 2% that changed subgroup is contamination of the sample with normal brain tissue, making accurate classification more difficult. Unfortunately, there was insufficient remaining tumour tissue for repeat analysis. Another explanation is that as the DKFZ classifier is a research tool not yet proven in a clinical setting, it is possible that misclassification could occur.

There was a small but significant decrease in classifier scores of recurrent tumours compared to primaries (median 1.000 in the primaries and 0.980 in the recurrences, $p=0.025$). The classifier was developed from primary tumours taken from patients who had received neither chemotherapy nor radiotherapy at the time of resection (www.molecularneuropathology.org - not yet published). It is possible that there were subtle therapy induced changes that affected the DNA methylation profile at recurrence. This hypothesis merits further consideration.

An analysis of the two major paediatric DNA methylation subgroups, EPN_PFA and EPN_RELA, demonstrated a significant difference in time to first recurrence. EPN_RELA recurred more rapidly than EPN_PFA, with all of the EPN_RELA cases relapsing within two years of primary diagnosis, and EPN_PFA cases taking up to 12.5 years. This finding may have implications for how the follow up of children with different DNA methylation subgroups is undertaken. Children with EPN_RELA may need more regular follow up immediately after diagnosis, whereas children with EPN_PFA may need longer term follow up.

EPN_PFA and EPN_RELA demonstrated no statistical difference in OS for children experiencing at least one recurrence. OS for the tumours previously associated with improved outcomes (EPN_PFB, EPN_MPE and EPN_YAP) appeared to be better than for EPN_PFA and EPN_RELA. However, numbers were inadequate to

confirm this statistically, reflecting their relative rarity. Whilst these tumours do still show evidence of a propensity for relapse, their OS was still generally good. This was supported by seven out of nine of these patients being alive at the end of follow up, but needs further investigation with substantially increased numbers.

A more detailed analysis of the EPN_RELA and EPN_PFA subgroups failed to demonstrate any association for factors previously thought to impact OS at primary diagnosis, including: treatment with radiotherapy or chemotherapy; extent of resection; tumour grade; and age at diagnosis. A positive association was seen between grade and EPN_PFA tumours for time to first progression; tumours with a higher grade recurred more quickly. This cohort contained only recurrent tumours and therefore some of the lack of survival associations may relate to the selection of a group with inherently poor outcomes. Alternatively, there may have been a lack of power to detect associations and therefore further research, with greater numbers, is needed to determine the effects of these factors in primary tumours that go on to recur.

Radiotherapy was associated with better OS and EFS in EPN_PFA tumours at first recurrence. Surgery was not associated with any benefit in either group and radiotherapy was not associated with outcome in the EPN_RELA group. Given that the EPN_RELA group was very small it is difficult to draw firm conclusions from this finding. However, the radiotherapy findings for EPN_PFA were consistent with the results of the overall clinical cohort. Future research needs to generate cohorts with larger numbers to be able to determine the true behaviour of these molecular subgroups after therapy.

The multidimensional scaling demonstrated that samples clustered by their DNA methylation subgroup rather than primary or recurrence status, suggesting minimal change in DNA methylation profiles from primary to recurrence. This indicates that DNA methylation subgroup is the most important factor in determining how these tumours relate to one another. This may also suggest that DNA methylation modifying therapies that have failed on primary tumours are unlikely to be successful at recurrence. However, an analysis of differentially methylated regions at primary versus recurrence may help to identify any subtle changes in more detail and will form part of the ongoing work following this doctoral research.

Whilst it was not the primary aim of the study to perform in-depth analysis of ependymoma DNA methylation profiles, this data provided an opportunity to look, for the first time, at the clinical outcomes of a recurrent ependymoma cohort with confirmed DNA methylation subgroups. The data also provided an additional validation tool for the RNA sequencing performed in subsequent chapters.

5 RNA Sequencing of 106 FFPE Ependymomas

5.1 Introduction

Given that paediatric ependymoma is relatively rare, with an incidence in the United Kingdom of 2-10 cases per million for children aged 0-14 years (Cancer Research UK, 2015b), obtaining access to high quality tissue samples is challenging. In the UK, the CCLG tissue bank (CCLG, 2017) provides a resource from which to identify archival tumour specimens, particularly those collected from uncommon malignancies. The difficulty of accessing sufficient tissue specimens was compounded by the fact that this study required matched primary and recurrent samples to perform a paired analysis. Given that only two, single institution studies have conducted paired analysis of gene expression patterns in recurrent paediatric ependymoma (Hoffman et al. 2014a; Peyre et al. 2010), it is likely that this is not a unique problem. It therefore required an innovative solution.

A number of approaches were taken to increase the number of available samples. Firstly, collaborations were developed with the research group of Dr Nicholas Foreman in Denver, USA and with Dr Thomas Jacques at the UCL Institute of Child Health in London. In addition to utilising samples provided through the CCLG biobank, these collaborations increased the number of samples available for profiling. Secondly, the spectrum of potential samples available for use was widened by investigating the use of RNA sequencing of FFPE tissue. Presently, gene expression profiling tends to be limited to high quality fresh frozen (FF) tissue specimens. There is emerging evidence that it is feasible to perform whole transcriptome RNA sequencing on archival FFPE tissue specimens. In order to identify the level of evidence available with regards to FFPE RNA-seq, a PubMed search was undertaken with the terms “FFPE”, “RNA” and “Sequencing”.

221 results were returned and abstracts reviewed. Only studies of whole transcriptome RNA-seq on archival FFPE specimens were retained, resulting in 19 studies, 16 on human and three on animal tissue (Table 5-1). Of the human studies; 13 were based on cancer samples. None of these investigated any type of paediatric brain tumour. The only study using brain tissue was a study of glioblastoma multiformae (GBM) in adults, which consisted of just four samples (Esteve-Codina et al., 2017). This is important as gene expression patterns have

been reported to be different in the brain compared to other tissues, in particular due to the presence of nascent RNAs (Ameur et al., 2011).

This study provided the opportunity to (1) validate the use of FFPE RNA-seq in paediatric brain tumours and (2) add a substantial amount of data to developing knowledge about the benefits and pitfalls of this approach.

This chapter describes the FFPE RNA-Seq analysis of a cohort of primary and recurrent ependymomas, from tissue with storage times of up to 30 years. The validity of this approach was assessed by comparing the results to: previously published datasets; the DNA methylation profiles generated in Chapter 4; and matched fresh frozen (FF) specimens sequenced using the same library preparation techniques. In addition to the matched FF specimens, a larger, unmatched FF cohort of 67 samples was sequenced to make more general comparisons between sequencing outcomes for FFPE and FF brain tumour samples. Recommendations were also made to assist further research in this field and beyond.

The primary aims were to establish:

(1) Is RNA-seq, from FFPE tissue, feasible on a large scale?
(2) How does the quality of the data compare to FF samples?
(3) Is the data of adequate quality to include when investigating other research questions?
(4) What are the potential pitfalls of this approach and can recommendations be generated to advise future research?

(A) Human Studies

Reference	PubMed ID	Species	Read Type	Read Length	Read Depth	Samples	Matched FF/FFPE	Sample Age	Sample Origin	Cancer Study
(French et al., 2017)	28818508	Human	SE	50	NK	3	0	NK	Liver	No
(Haile et al., 2017)	28570594	Human	PE	75	20M	4	0	4 yrs	Lymphoma	Yes
(Jovanović et al., 2017)	28376728	Human	NK	NK	13M-80M	21	21	4 >10 yrs, 17 < 10 yrs	Breast	Yes
(Esteve-Codina et al., 2017)	28122052	Human	PE	76	54M-65M	4	4	NK	GBM	Yes
(Vukmirovic et al., 2017)	28081703	Human	PE	50	50M	12	0	6 yrs	Lung	No
(Guo et al., 2016)	27774452	Human	PE	90	20M	4	0	8 yrs	Breast	Yes
(Just et al., 2016)	26998913	Human	SE	80	42M	1	0	<2 yrs	Kidney	Yes
(Graw et al., 2015)	26202458	Human	NK	NK	50M	6	6	<1.5 yrs	Ovarian	Yes
(Li et al., 2014)	25495041	Human	SE	35	10M	2	2	2 yrs	Kidney	Yes
(Zhao et al., 2014)	24888378	Human	PE	48	200M	17	17	NK	Breast	Yes
(Hedegaard et al. 2014)	24878701	Human	PE	100	20M-50M	73	38	< 20 yrs	Bladder, Prostate, Colon, Tonsil	Yes
(Morton et al., 2014)	24735754	Human	PE	100	198M	18	0	3yrs	Lung	Yes
(Norton et al., 2013)	24278466	Human	PE	50	NK	9	9	4 yrs	Breast	Yes
(Xiao et al., 2013)	23180419	Human	SE	76	235M	2	0	94 yrs	Lung	No
(Morlan et al., 2012)	22900061	Human	SE	50	20-50M	4	0	NK	Breast	Yes
(Sinicropi et al., 2012)	22808097	Human	SE	50	43M	136	0	8.5 yrs	Breast	Yes

(B) Animal Studies

Reference	PubMed ID	Species	Read Type	Read Length	Read Depth	Samples	Matched FF/FFPE	Sample Age	Sample origin	Cancer Study
(Amini et al., 2017)	28835206	Canine	SE	125	NK	4	0	NK	Breast	Yes
(Hester et al., 2016)	27562560	Mouse	PE	50	30-70M	40	40	<2 vs >20yrs	Liver	No
(Auerbach et al., 2015)	25378103	Rat	PE	100	100M	8	8	4 yrs	Liver	No

Table 5-1: Summary of previously published studies investigating whole transcriptome RNA sequencing from archival FFPE tissues. (A) Studies using human tissue. (B) Studies using animal tissue. NK: Not Known. BP: Base pairs. M: Million.

5.2 Materials and methods

FF and FFPE cohorts were formed from a subset of the clinical recurrent cohort described in Chapter 3 based on tumour tissue availability. RNA sequencing was performed by Exiqon (Denmark), as described in section 2.9. Following the generation of the raw data, all analysis was conducted by the author. The file formats encountered in the analysis (fastq, SAM and BAM) are described in Appendix 2. The theory underpinning RNA-seq methodology and the key terminology is outlined in section 1.6.

5.2.1 RNA sequencing data analysis pipeline

An analytical pipeline was developed to ensure a uniform approach to all samples. This consisted of:

- FastQC data quality control (Andrews, 2010) (section 5.2.1.1);
- Read trimming of adapter sequences and low quality bases using Trimmomatic (Bolger et al., 2014) (section 5.2.1.2);
- Removal of abundant sequences, particularly rRNA, by alignment with TopHat2 (Kim et al., 2013) (section 5.2.1.3);
- Aligning remaining reads to the transcriptome and genome using TopHat2 (section 5.2.1.4);
- Counting reads and assigning to exons using FeatureCounts from the RSubRead package (Liao et al., 2014) (section 5.2.1.5).

The output of the pipeline was a matrix of raw gene expression levels which was then normalised before downstream analysis. The computer scripts used to perform the analysis can be found in Appendix 3. The results generated from the data analysis pipeline can be found in Appendix 4.

5.2.1.1 FastQC data quality control

FastQC is a tool written in JavaScript to perform basic quality control of high throughput sequencing data (Andrews, 2010). It was designed for DNA sequencing, but provides insight into RNA-seq data quality. It can assist with decisions about quality control interventions prior to downstream analysis, but does not have specific cut-offs for poorly performing samples in RNA-Seq.

Paired fastq files were run through the FastQC tool using the desktop module. Details of the parameters analysed and summaries of their interpretation for RNA-seq, as derived from the FastQC documentation (Andrews, 2010), follow:

- *Basic Statistics (*Figure 5-1*A):*
  This module provided details of the input file, number of reads generated and sequence length. It was used to check the basic parameters of the sequencing run.

- *Per base sequence quality (*Figure 5-1*B):*
  This module detailed the Phred quality score for each base position in the read across each sample. Reads on a green background were good quality; amber, an acceptable quality; and red, poor quality. The blue line indicated the mean quality score; the yellow box, the interquartile range; and the whiskers, the top and bottom 10% of scores.

- *Per base sequence content (*Figure 5-1*C):*
  For DNA sequencing, the relative composition of the four nucleotides should be the same at each base position in the read. This assumption does not hold true for RNA-seq, as library preparation begins with random hexamer priming that introduces a technical bias. This explains the erratic lines to the left of the illustrated plot and is why the module generates a warning in many RNA-seq libraries. It is not thought to impact downstream analyses (Andrews, 2010).

- *Per tile sequence quality (*Figure 5-1*D):*
  This module detailed the Phred quality of the reads at different positions on the Illumina flow cell. It can identify losses in quality in a particular area of the flow cell. The plot should be blue; other colours signify problems with the sequencing run.

- *Per base N content (*Figure 5-1*E):*
  When the sequencer is unable to accurately call a base, it replaces it with an 'N' in the output fastq file. This plot demonstrated the average proportion of 'Ns' at each base position in the read. An increase in 'Ns' can occur towards the end of sequencing runs if reads are short, or as chemicals used in the process deteriorate.

- *Per sequence quality scores (*Figure 5-1*F):*
  This module showed the number of reads per Phred score for each sample. Normally, most reads have high scores generating a unimodal distribution, peaking at the right of the graph (Andrews, 2010).

- *Per sequence GC content (*Figure 5-1*G):*
  The Guanine-Cytosine (GC) content from each sample should follow an approximately normal distribution to indicate a lack of bias or contamination within the data. A total RNA library would be expected to have a shifted distribution to an mRNA enriched library as RNA species have differing GC content (Aissani and Bernardi, 1991). Additionally, as every organism has a unique GC composition profile, the analysis can highlight evidence of contamination with non-human RNA.

- *Sequence length distribution (*Figure 5-1*H):*
  This module plotted the length of each read and could be affected by library preparation methods and read trimming.

- *Sequence duplication levels (*Figure 5-1*I):*
  This module estimated the level of duplication of reads within each fastq file. Less complex libraries produce less information about expressed genes, and have high levels of duplicated sequences. High levels of duplicated sequences can be caused by rRNA contamination and excessive PCR amplification during library preparation, and are indicated by a rise in the line at the right of the graph illustrated.

- *Adapter content (*Figure 5-1*J):*
  This module detects adapter sequences in the fastq files which can be removed by trimming to prevent negative impact on downstream analyses. Adapter sequences are bound to both ends of each cDNA fragment, with the proximal adapter attached to the flow cell. The sequencer reads along the fragment, proximally to distally. As the read length is fixed, if the fragment is short then the sequencer will begin to read into the adapter sequence attached to its distal end. This results in adapter sequence read-through in the raw read.

Following FastQC processing, results from individual samples were inspected, normalised and then combined within the R statistical environment (R Core Team, 2014), in order to assess quality differences between FFPE and FF cohorts.

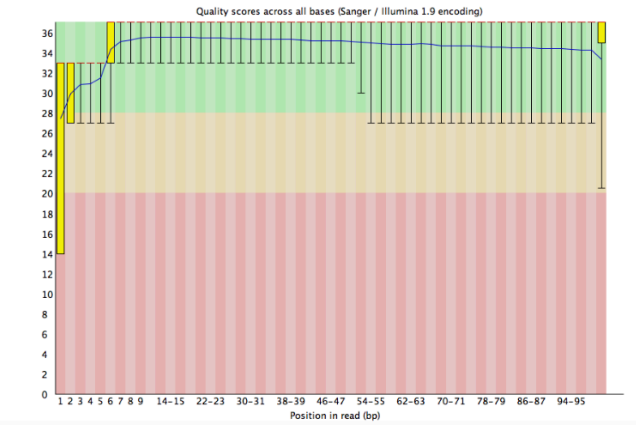
A

Basic Statistics

Measure	Value
Filename	4589-072.R1.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	49651561
Sequences flagged as poor quality	0
Sequence length	101
%GC	45

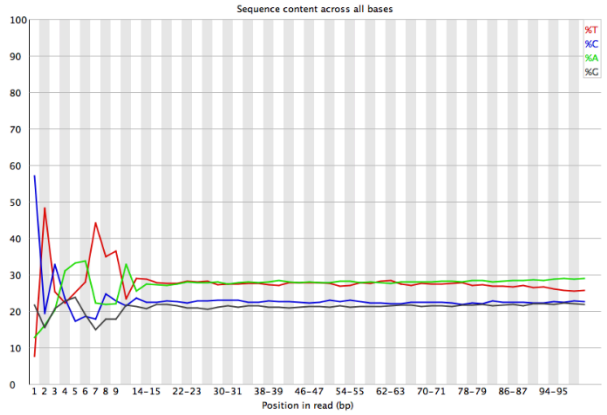
B

Per base sequence quality



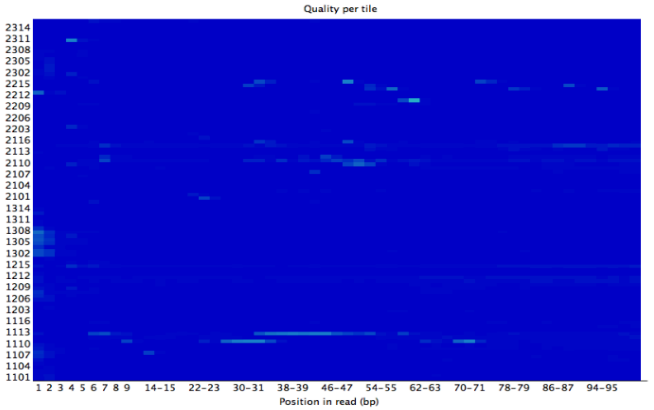
C

Per base sequence content



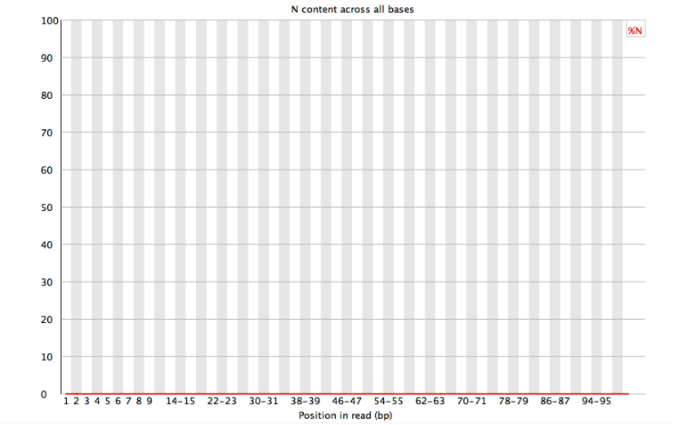
D

Per tile sequence quality



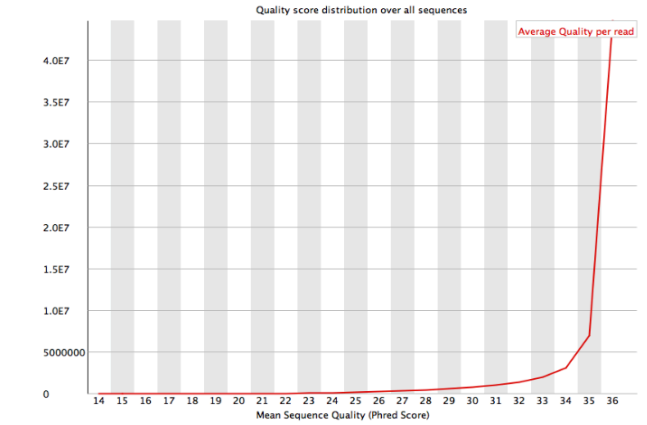
E

Per base N content



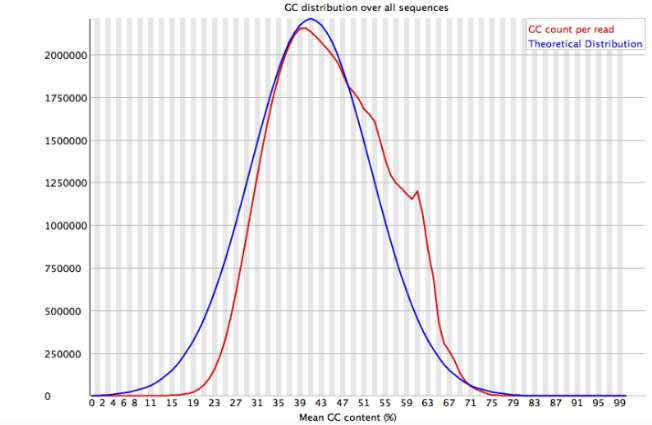
F

Per sequence quality scores



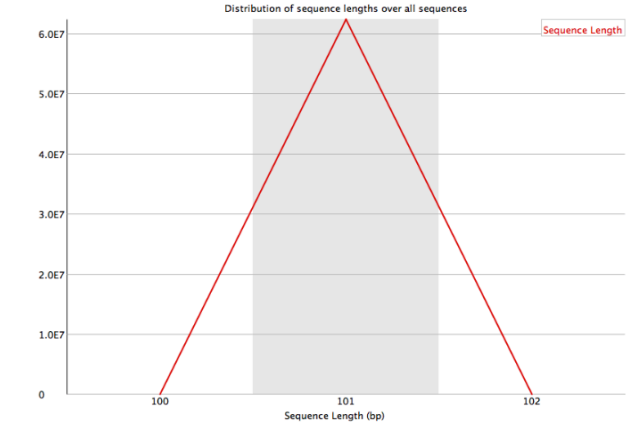
G

Per sequence GC content

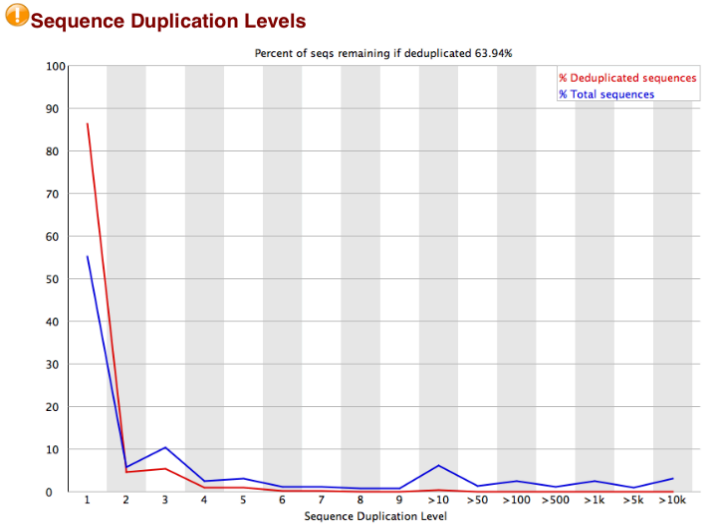


H

Sequence Length Distribution



I



J

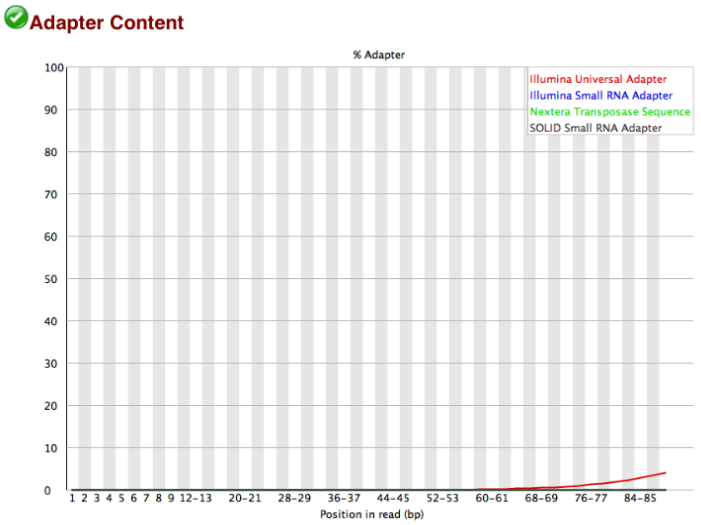


Figure 5-1: Summary of FastQC modules used in quality control of the FF and FFPE raw data (Andrews, 2010). Results from fastq file 4589-072 (Forward sequences) used as an example. (A) Basic statistics. (B) Per base sequence quality. (C) Per base sequence content. (D) Per tile sequence quality. (E) Per base N content. (F) Per sequence quality scores. (G) Per sequence GC content. (H) Sequence length distribution. (I) Sequence duplication levels. (J) Adapter content.

5.2.1.2 Read trimming of adapter sequences and low-quality bases

Reads were trimmed to remove adapter sequences and low-quality bases. There is debate about the benefits of trimming but there is some evidence that light trimming improves alignment rates (Del Fabbro et al., 2013; Williams et al., 2016).

Adapter sequences were removed using the ILLUMINACLIP command in Trimmomatic (Bolger et al., 2014) and paired-end adapter sequences provided by Illumina. Low quality bases were removed from the beginning and end of reads based on Phred scores below five. The SLIDINGWINDOW command was used to scan the full lengths of reads to remove sections where the mean Phred score fell below 15 over four bases. If read length fell below 30 base pairs the read was discarded; excessive numbers of short reads can result in changes in differential expression due to ambiguous mapping (Williams et al., 2016).

5.2.1.3 Removal of abundant sequences

All samples were filtered to remove abundant sequences (predominantly ribosomal RNA but also transfer RNA and mitochondrial RNA), by aligning all reads in each sample to an abundant sequences index file and retaining the unaligned reads.

The abundant sequences were obtained from the UCSC genome browser in FASTA format (www.genome.ucsc.edu) and a Bowtie 2 index file was generated. A Bowtie 2 index is a way of rearranging the genomic sequence, in order to minimise the amount of computer memory used to perform the alignment.

Alignment was then performed using Tophat 2. This required two fastq files of raw reads (forward and reverse) and a Bowtie 2 index as input. The output of the alignment was a BAM file of reads aligned to abundant sequences (discarded) and a BAM file of unaligned reads (retained).

The default parameters of TopHat 2 were modified to take into account the inner mate distance for the paired-end reads (Figure 1-6). Fragment length was estimated by the sequencing provider (Exiqon, Denmark) on the basis of Bioanalyser analysis of the sequencing libraries. The estimated average fragment length was 300 bases, from which adapter lengths were subtracted, leaving a median insert length of 155 base pairs. From the insert length, the inner mate

distance was calculated by subtracting the read lengths (2*100). This calculation gave a median inner mate distance of -45.

The calculated inner mate distance was confirmed by examining the BAM files using Picard tools (Accessed at: http://broadinstitute.github.io/picard, December 2015), which indicated that the median insert length was 156 base pairs. Subtracting the read lengths from this (2*100) indicated an inner mate distance of -44, which was consistent with the value predicted by the Bioanalyser.

5.2.1.4 Alignment to transcriptome and genome

The unaligned BAM files were first converted back to fastq files, using the bam2fastx converter within the TopHat 2 package. They were then aligned, by TopHat 2, to the human transcriptome (Gencode GrCh37 Version 11, obtained from www.genecodegenes.org) and genome (version Hg19). The output of the genome/transcriptome alignment was a BAM file of reads aligned to human sequences and a BAM file of unaligned reads for each sample. The BAM files of aligned reads were used to count reads overlapping with exons and genes.

5.2.1.5 Counting reads using FeatureCounts

FeatureCounts, a programme within RSubRead (Shi, 2014), summarises aligned reads within a BAM file (Liao et al., 2014), and was used to count aligned fragments overlapping with exons or genes. This resulted in a counts matrix indicating the number of reads aligning to each exon or gene.

Inputs were a BAM file for each sample and the transcriptome reference sequence (Gencode GrCh37 Version 11). The commands for the algorithm were changed from default, to:

- facilitate paired-end reads;
- only count a fragment when both reads were mapped;
- specify a minimum mapping quality score of 10.

In order to summarise the read counts and make comparisons between FF and FFPE cohorts, metrics were generated for proportions of reads falling on exons; introns; intergenic regions; or reads not assigned to any feature. The proportion of reads falling on exons was calculated by working out the overall percentage of reads assigned to exons. The proportion of intronic reads was calculated by subtracting the number of reads assigned to exons from the number of reads assigned to genes. Intergenic reads were assumed to be those which could not be assigned to any features in the GTF file. Those reads which were either from

chimeric fragments, were ambiguous or unassigned for any other reason were included in the 'technical' category.

5.2.2 Differential expression analysis

EdgeR (Empirical Analysis of Digital Gene Expression Data in R) (Robinson et al., 2010) was selected to perform differential expression analysis in view of its extensive documentation, widespread use and ability to perform a paired analysis.

Raw, untransformed, data was loaded into the R statistical environment as a table of counts, one column per sample and one gene per row, with a sample sheet to identify the correct column for each sample. Gene names were converted from Ensembl gene IDs to Entrez gene IDs and official gene symbols using the R packages org.Hs.eg.db v3.4.1 (Carlson, 2017) and AnnotationDbi v1.3.82 (Pages et al., 2017).

In order to maximise the utility of the False Discovery Rate (FDR) corrections, data was filtered to remove genes with a count of less than two fragments per million amongst half of the dataset. This approach aimed to remove genes that would be unlikely to show differential expression, over and above the noise of lowly expressed genes in the data.

Following filtering, EdgeR corrected for varying library size between the samples and estimated the dispersion of the data using common, trended and tagwise approaches (Robinson et al., 2010). Differential expression analysis was then performed, generating lists of differentially expressed genes.

5.2.3 Clustering and data visualisation

For comparison of FFPE and FF cohorts, data was normalised and transformed using the R-log transformation in DESeq2 (Love et al., 2014). The R-log transformation normalises samples based on the number of reads generated, but also takes into account the distribution of lowly and highly expressed genes to prevent them from distorting the final results (Love et al., 2014).

Following transformation, samples underwent unsupervised hierarchical clustering using the HClust package in R (R Core Team, 2014). Clustering was performed using all of the genes and subsets of the most highly expressed and variable genes; these approaches were compared to assess the clustering stability. For all clustering, Euclidean distance measures were used and Ward's algorithm was

implemented. Additionally, clusters identified as possible PF subgroups were subjected to supervised hierarchical clustering using published signature genes of the PF subgroups (Pajtler et al., 2015; Witt et al., 2011), to check for overlap with the unsupervised approach.

5.2.4 The hypergeometric test

To compare the similarity of FFPE, FF and previously published datasets, hypergeometric tests were used to assess whether the level of overlap of significantly differentially expressed genes was likely to have occurred by chance. This test used the principle of random sampling, without replacement, from a population of known size. When comparing two sets of genes (sets A and B) the calculation in the R statistical environment took the format:

```
phyper(q, m, n, k, lower.tail=FALSE)
```

where:
q = number of significant genes appearing in set A and in set B;
m = number of significant genes in total in set A;
n = Total number of genes in set A minus number of significant genes in set A;
k = Number of significant genes in total in set B;
lower.tail=FALSE represents a need to look at the upper tail of the distribution.

The calculation was performed using the R base statistics package and the example above is derived from its associated documentation (R Core Team, 2014). A p-value of <0.05 was considered significant.

5.2.5 Gene ontology and enrichment analyses

Changes in related groups of genes can infer change in biological function. In order to investigate this, gene ontology and enrichment analyses were performed.

A threshold free approach was used by implementing the Gene Ontology enRichment anaLysis and visuaLizAtion tool (GOrilla) (Eden et al., 2009). This approach meant that groups of significant genes did not need to be arbitrarily specified by selecting a cut-off p-value, but that the enrichment algorithm identified groups of genes over-represented at the top compared to the bottom of ranked lists. There is evidence that gene ontology analyses are more sensitive to underlying biological changes if up- and down-regulated genes are considered

separately (Hong et al., 2013), therefore lists were ranked based both on fold change direction and statistical significance of all of the expressed genes.

In order to corroborate the GOrilla analyses, gene set enrichment analyses were performed using Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005). This approach used a modified Kolmogorov-Smirnov statistic as opposed to the multiple Hypergeometric test (mHG) test performed by GOrilla. Whilst GOrilla tested many more ontology terms than GSEA (14212 versus 4436 at the time of analysis), using both approaches meant that the analysis would be strengthened if two different statistical methods supported similar conclusions.

The GOrilla package was implemented through the web based interface (accessible at http://cbl-gorilla.cs.technion.ac.il). The entire ranked gene list was entered into the software. The GSEA package (version 3.0) was downloaded from the Gene Set Enrichment Analysis website (www.gsea-msigdb.org) as a java package. The enrichment analyses were run using the GSEA Preranked option. This allowed for the direct input of the gene lists generated by EdgeR. GSEA was run in 'classic' mode for the enrichment statistic, selected from the basic fields options, with otherwise default parameters.

The Database for Annotation, Visualisation and Integrated Discovery v6.7 (DAVID) (Huang et al., 2009a, 2009b) was used to compare the results with previous studies (Hoffman et al., 2014a; Wani et al., 2012). Genes with FDR <0.05 and fold change >2 were used as the target set. DAVID was not used for analyses other than validating the results in the context of these previous studies, and for comparing differential expression between matched FF and FFPE samples, where a statistical cut-off was desirable for an effective comparison.

5.3 Results

5.3.1 Generation of sequencing libraries

132 FFPE samples had RNA extracted, following which 26 were excluded for having an inadequate quantity of total RNA, leaving 106 samples for sequencing. The median input RNA concentration was 102.3 ng/μl (21.3 ng/μl to 877.5 ng/μl). Satisfactory libraries were created from all samples. The median age of the FFPE blocks was 11.7 years (0.54-27.85 years). All 67 FF samples had a sufficient quantity of RNA for sequencing with a concentration of greater than 100 ng/μl. Satisfactory libraries were created from all samples. 14 individual tumour

samples, including one matched primary and recurrent pair, appeared in both FF and FFPE cohorts.

5.3.2 Clinical summary of FFPE and FF tumours undergoing RNA-seq

FFPE cohort

106 samples from 50 (24%) patients in the clinical recurrent cohort were analysed; including 25 matched primary and first recurrences. The samples that were not part of matched pairs were either primary or recurrent samples from tumours which had recurred. There were no differences in parameters for the FFPE cohort versus the clinical recurrent cohort (Table 5-2).

Parameter		**Clinical Cohort (n=188)**		**FFPE Cohort (n=50)**		**P Value**
		Number	%	Number	%	
Age	<3 years	94	51	27	55	0.521
	3+ years	91	49	21	45	
	NK	3	-	2	-	
Gender	Male	105	58	28	57	1.000
	Female	77	42	21	43	
	NK	6	-	1	-	
Extent of Resection	GTR	76	45	24	52	0.409
	STR	93	55	22	48	
	NK	19	-	4	-	
Location	PF	136	73	35	70	0.710
	ST	44	23	13	26	
	SP	7	4	2	4	
	NK	1	-	0	-	
Grade	WHO II	85	52	23	46	0.519
	WHO III	78	48	27	54	
	NK	25	-	0	-	
Radiotherapy at diagnosis	Yes	104	59	30	64	0.617
	No	73	41	17	36	
	NK	11	-	3	-	
Median age		35 months		30 months		0.687
Median TTP		17 months		18 months		0.311
Median OS		61 months		106 months		0.087

Table 5-2: Comparison of key parameters between clinical and FFPE cohorts. NK: Not Known. P-values for clinical parameters by Chi-square test, for times by Wilcoxon and Supremum tests.

FF Cohort

67 samples from 30 (16%) patients in the clinical recurrent cohort were analysed. 29 had matched primary and recurrent pairs; two had only second recurrences available, leaving 27 primary and first recurrent pairs. There were no significant differences in clinical parameters for the FF cohort, compared to the clinical recurrent cohort (Table 5-3).

Parameter		**Clinical Cohort (n=188)**		**FF Cohort (n=30)**		**P Value**
		Number	%	Number	%	
Age	<3 years	94	51	13	43	0.556
	3+ years	91	49	17	57	
	NK	3	-	-	-	
Gender	Male	105	58	16	55	0.841
	Female	77	42	13	45	
	NK	6	-	1	-	
Extent of Resection	GTR	76	45	11	50	0.658
	STR	93	55	11	50	
	NK	19	-	8	-	
Location	PF	136	73	23	77	0.632
	ST	44	23	5	17	
	SP	7	4	2	7	
	NK	1	-		-	
Grade	WHO II	85	52	7	33	0.163
	WHO III	78	48	14	67	
	NK	25	-	9	-	
Radiotherapy at diagnosis	Yes	104	59	19	68	0.412
	No	73	41	9	32	
	NK	11	-	2	-	
Median age		35 months		40 months		0.936
Median TTP		17 months		17 months		0.718
Median OS		61 months		77 months		0.191

Table 5-3: Summary of the clinical features of the FF compared to the overall clinical cohort. NK: Not Known. P-values for clinical parameters by Chi-square test. P-values for times by Wilcoxon and Supremum tests.

5.3.3 Generation of raw reads

Libraries were sequenced with a target depth of 50 million paired-end reads. The mean number of raw reads generated by the FFPE samples was 45.9 million (median 46.3 million) with a range of 2.2 to 93.2 million. In comparison, the FF samples generated a mean of 54 million reads (median 53.4 million) with a range of 22 to 78.5 million (Figure 5-2A). There was a statistically significant difference between the median number of reads generated for the FF compared to FFPE group ($p<0.001$). The FF samples also had a smaller range of reads (91 million for FFPE, 56.5 million for FF). Seven of the libraries in the FFPE group generated particularly low numbers of reads with less than 20 million.

There was no correlation between the age of the FFPE block and the number of raw reads generated ($r=+0.06$, $p=0.546$). For the FFPE samples there was a weak positive correlation between the number of raw reads generated and the concentration of the input RNA ($r=+0.21$, $p=0.030$). The 260/280 and 260/230 spectrophotometer purity measurements had no association with the generation of raw reads ($r=+0.14$, $p=0.167$ and $r=+0.08$, $p=0.398$ respectively).

5.3.4 Read trimming

The median proportion of reads removed by trimming was 3.0% in the FFPE group and 4.7% in the FF group ($p<0.001$). The difference in distribution between the two groups was distinct; seven FFPE samples had in excess of 10% of reads removed, compared to none of the FF samples (Figure 5-2B).

5.3.5 Filtering abundant sequences

A median of 4.1% of reads were filtered from the FFPE samples, compared to 3.2% of FF reads. This did not represent a statistically significant difference ($p=0.099$). Both cohorts had a small minority of samples with significant rRNA contamination, eight FFPE and seven FF samples had in excess of 10% rRNA content (Figure 5-2C).

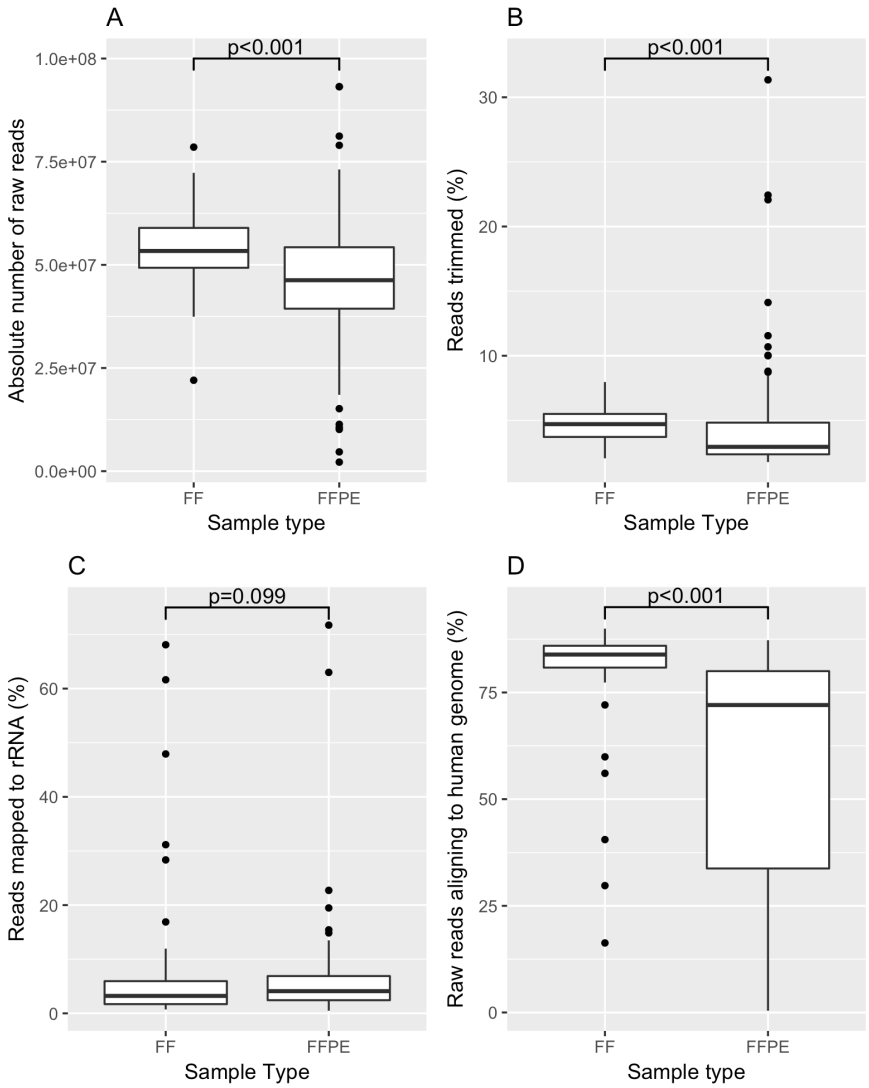


Figure 5-2: Basic mapping statistics for the FF and FFPE datasets. Numbers indicate p-values. Graphs demonstrate the variability of the FFPE cohort with wider boxes than the FF samples for every parameter investigated. (A) Number of raw reads generated by sequencer. (B) Levels of read trimming in each dataset. (C) Proportion of reads mapping to rRNA/abundant sequences. (D) Raw reads aligning to the human genome.

5.3.6 Reads aligning to the human genome

The median proportion of reads aligning to the human genome in the FFPE group was 72.1%, compared with 83.9% in the FF group ($p<0.001$). The FFPE reads exhibited a wider interquartile range of 46.2% compared with 5.1% in the FF group (Figure 5-2D).

FF samples with low human genome alignment were also noted to have higher levels of reads filtered out at the abundant sequences step, indicating rRNA contamination. This was not the case for FFPE, suggesting that in this cohort, the low alignment was not associated with rRNA contamination.

5.3.7 Bacterial sequence alignment in FFPE samples

To investigate the reason for low levels of alignment in some of the FFPE samples, reads were extracted at random from the BAM files of unaligned reads in samples with the lowest human alignment levels, and entered into the National Center for Biotechnology Information's Basic Local Alignment Search Tool (NCBI-BLAST). The search returned perfect alignment with numerous bacterial sequences, in particular the rRNA sequences of the Enterobacteriaceae family. As a result of this finding, the unaligned reads were realigned to the bacterial genome of e-coli, using TopHat 2, to establish the level of bacterial reads contained within the FFPE samples.

The proportion of reads aligning to bacterial sequences varied from 0.4% to 91.4% with a median of 7.9%. When the proportion of FFPE reads aligning to bacterial sequences was added to the proportion of reads aligning to human sequences, the overall alignment rate reached figures close to the overall alignment rate for the FF samples. The median proportion of aligned reads to any organism for the FFPE samples was 87.6% compared to 88.0% for the FFPE samples (p=0.078). There was a strong negative correlation between human and bacterial sequence alignment (r=-0.95, p<0.001) (Figure 5-3).

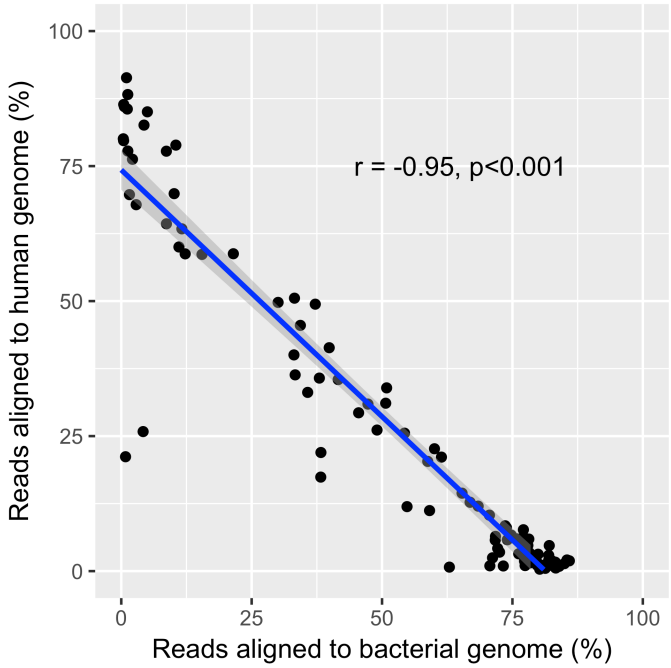


Figure 5-3: Proportion of reads aligning to human sequences plotted against the proportion of reads aligning to bacterial sequences for all the FFPE samples. The strong correlation indicates that as bacterial reads increased the proportion of reads aligning to the human genome fell.

5.3.8 FastQC analysis

The FastQC files were reviewed to look for differences in quality between the FFPE and FF cohorts, including evidence of contamination, which may have accounted for variability in alignment statistics.

5.3.8.1 FASTQC: Per sequence PHRED

The median number of reads with each Phred score were similar for both FFPE and FF cohorts, however there was a higher proportion of FFPE reads with the maximum Phred score (67% versus 58%) (Figure 5-4). FFPE samples had 0.14% of reads with low Phred scores (2 to 14), whereas the FF samples had no reads with a score less than 15. FFPE samples had more variable scoring than the FF samples but both cohorts still had a majority of reads with high scores.

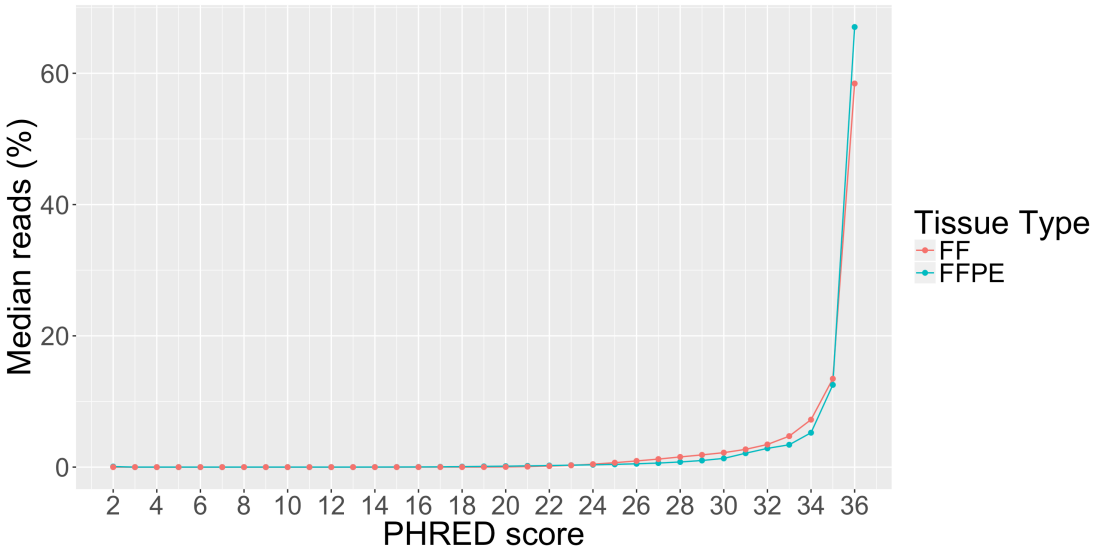


Figure 5-4: Illustration of the proportion of reads for each tissue type with Phred scores between 2 and 36. Each point represents the median percentage of reads at each score.

5.3.8.2 FastQC: GC content analysis

Combined GC profiles from the FFPE samples contained two peaks; one smooth peak at 35-42%, and one spike at 54% (Figure 5-5A). The FF data generated a similar smooth peak at 35-42% and a number of 'shoulders' at higher percentages. Given its similar nature in both cohorts, the smooth peak was thought to represent human mRNA (Figure 5-5B).

The sharp spike seen at 54% in the FFPE data correlated strongly with the samples that had very low (<20%) human genome, but high bacterial, alignment. Whilst many samples demonstrated evidence of reads contributing to

the bacterial spike, the worst affected had very few reads contributing to the human mRNA peak (Figure 5-5A).

Three FF samples generated large 'shoulders' vertically away from the normal distribution, at 47-63% (Figure 5-5B). These 'shoulders' had different morphology to the spike seen in the FFPE samples, and therefore represented a different source of contamination. The samples with 'shoulders' also demonstrated a lower percentage of reads contributing to the human mRNA peak. The alignment statistics for these three samples demonstrated significantly more rRNA than the other FF specimens (p=0.011), indicating rRNA contamination.

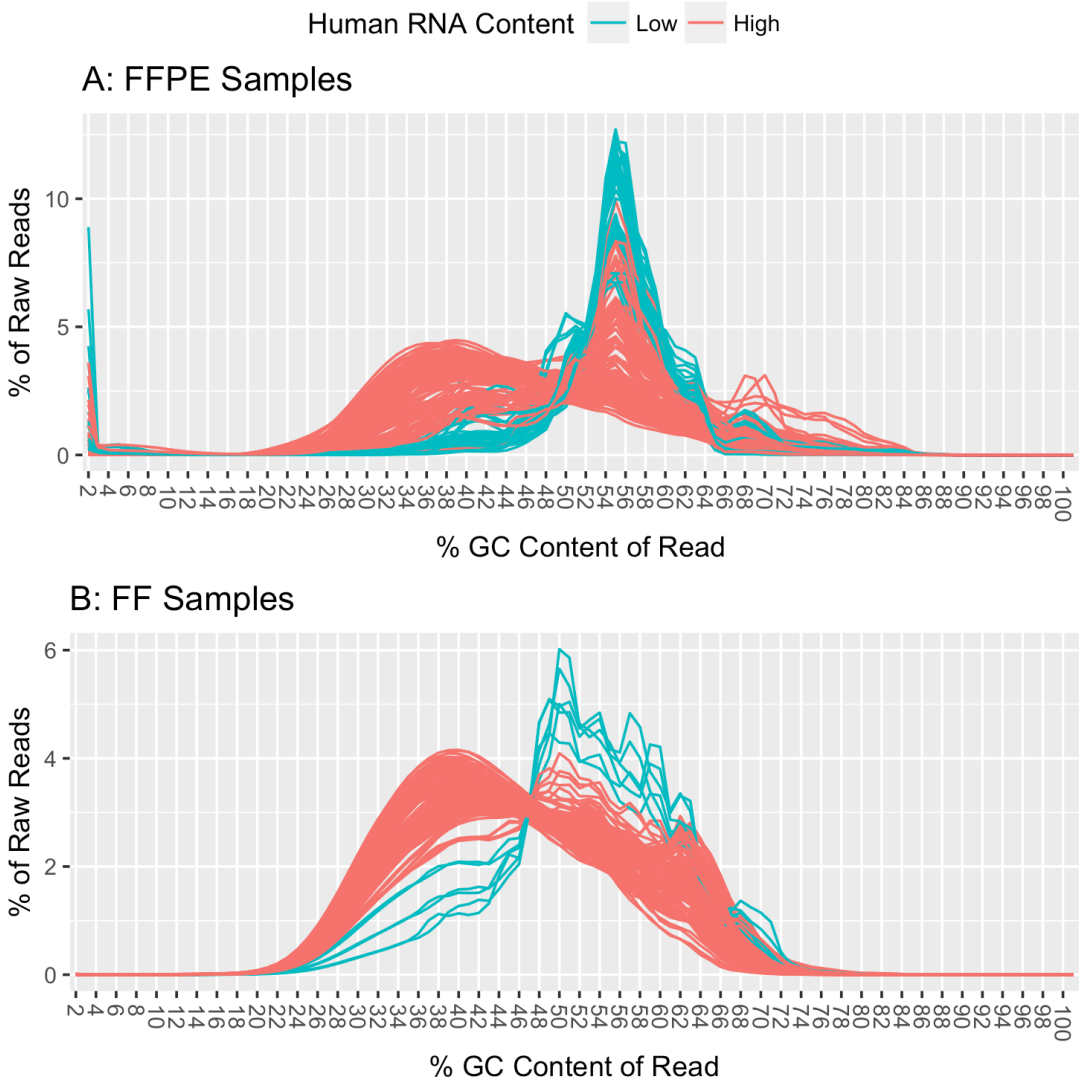


Figure 5-5: GC content curves. (A) FFPE samples demonstrating a smooth peak at 35-42% and a spike at 54%. The spike is formed by samples with high bacterial RNA and low human RNA (blue). (B) FF samples demonstrating a smooth peak at 35-42% and several 'shoulders' at 47-63%. The shoulders are formed by samples with high rRNA and low human RNA (blue).

5.3.8.3 FastQC: 'N' content analysis

The median proportions of 'Ns' increased along the read length in the FFPE samples, from 0.11% at the first base to 26.5% by base 100. In the FF samples the N content remained very low throughout (<0.08%) and showed no tendency to increase along the read length. At every base position, the N content in the FFPE reads was higher than N content in the FF reads ($p<0.001$) (Figure 5-6).

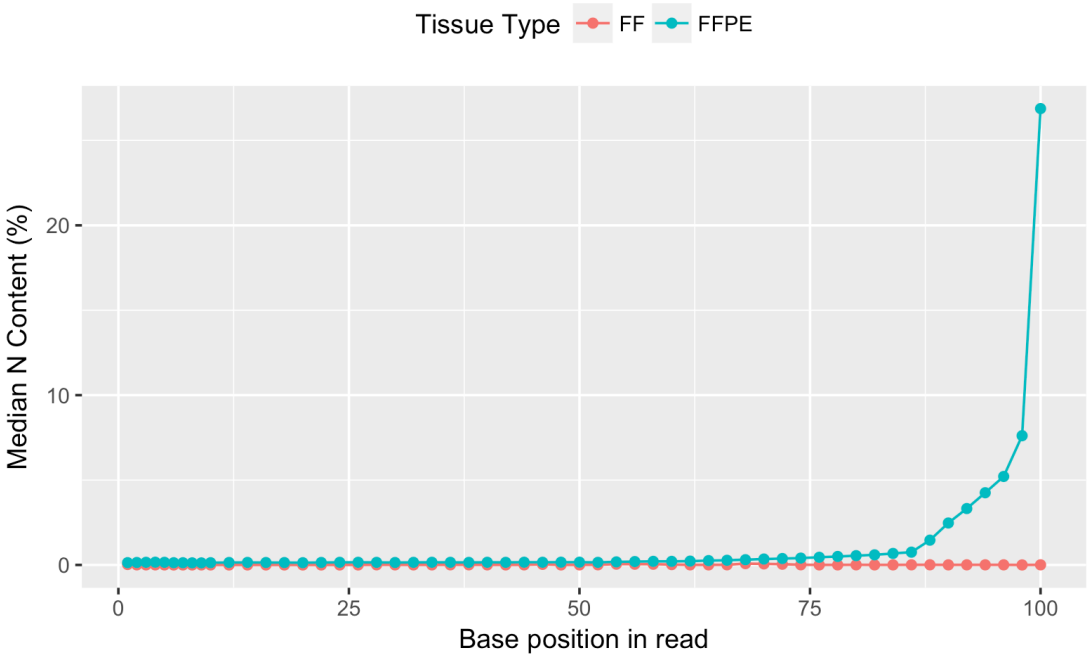


Figure 5-6: Median N content (% of total reads) by position of the base in read. FF tissue in red, FFPE tissue in blue.

5.3.8.4 FastQC: duplication levels

The FFPE samples had a significantly lower proportion of unique reads (median 9.6%, range 0.5% to 67.0%, $p<0.001$) than the FF samples (median 44.6%, range 13.1-70.3%) (Figure 5-7). The largest category of reads in the FF samples was formed from reads with no duplication (unique sequences) whereas the largest category in the FFPE samples was formed by those samples with a single duplicate.

In both cohorts, there was a small spike for sequences that were duplicated 10-50 times (FFPE: median 4.5%, range 1.9-64.8%; FF: median 7.1%, range 2.3-18.2%). This was largely consistent with that expected from an RNA sequencing library.

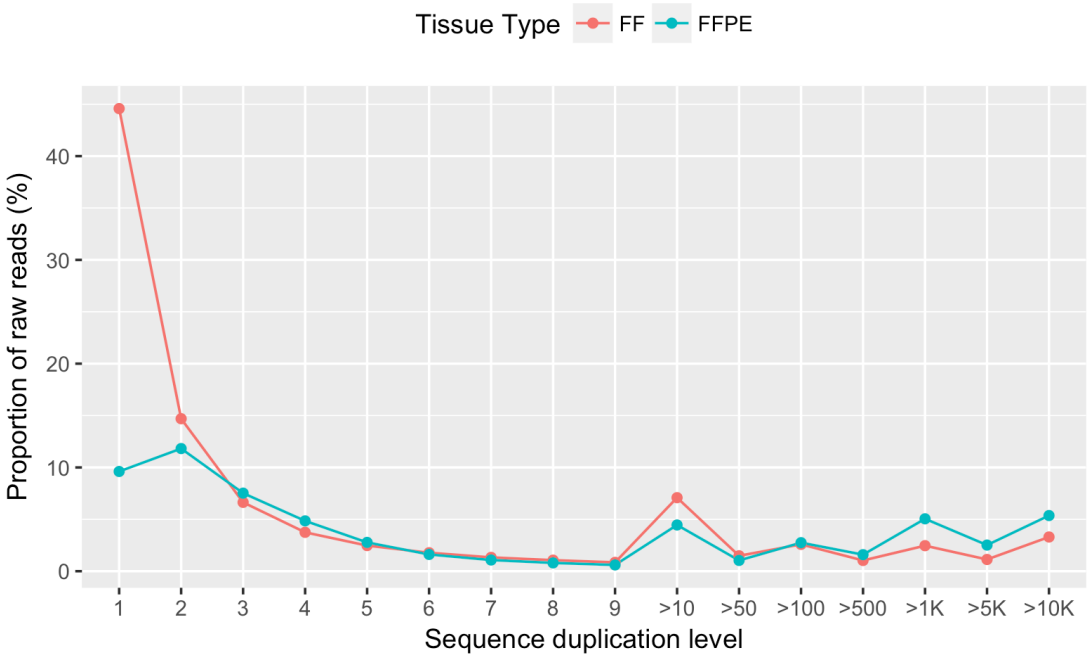


Figure 5-7: Median sequence duplication levels indicated by proportions of raw reads with different duplication levels across all samples. Sequences with no duplication (sequence duplication level = 1, unique sequences) were the most common finding in the FF cohort (44%). The FFPE samples demonstrated a significantly higher proportion of very highly duplicated sequences (sequence duplication level >10k) than the FF samples, $p<0.001$.

When examining sequences with extremely high duplication (greater than 10,000) the FFPE samples were more affected than the FF samples (FFPE median 5.4%, range 0.1-70.8% and FF median 3.3%, range 0.4%-47.2%) (Figure 5-7). The FF samples with these extremely high levels were those that had also demonstrated high levels of rRNA contamination in both the alignment and GC content analysis steps (Figure 5-8). Removal of rRNA during the alignment process removed this duplication.

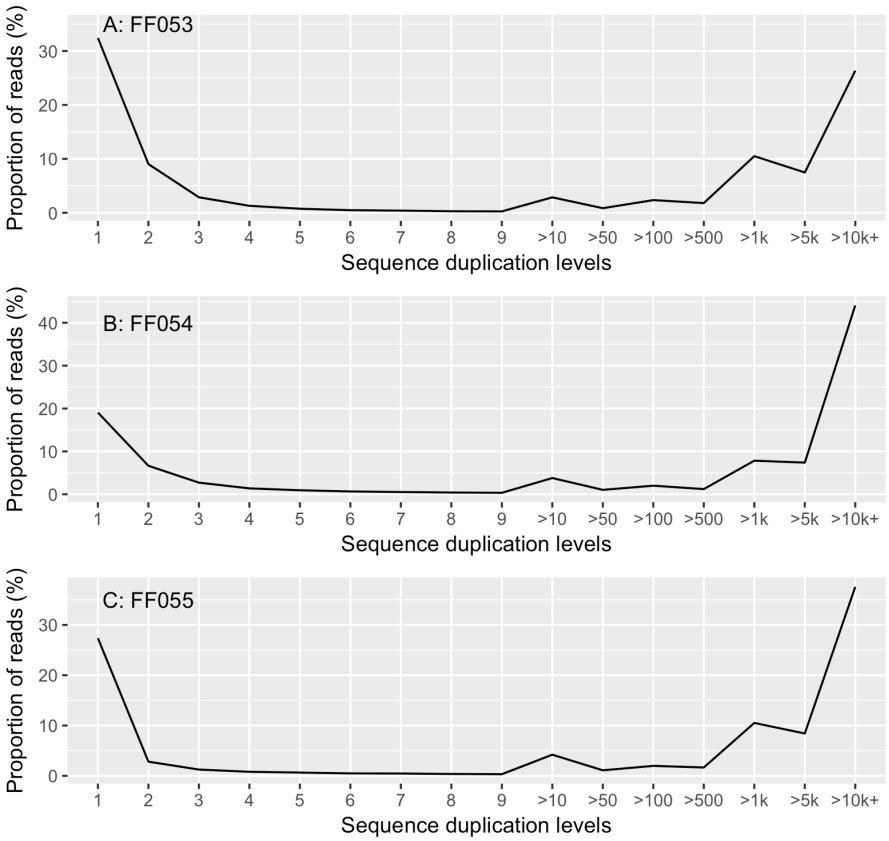


Figure 5-8: Sequence duplication levels for FF samples with high levels of rRNA contamination. FF053 – 47% rRNA alignment, FF054 – 62% rRNA alignment, FF055 – 68% rRNA alignment. High levels of rRNA sequence duplication identified by the rise in sequence duplication levels >10k.

The FFPE samples demonstrated a logarithmic relationship between low levels of unique reads (high duplication) and high levels of bacterial reads (Figure 5-9). The lack of library complexity arose from bacterial contamination, with many duplicated reads being contributed by bacterial rRNA.

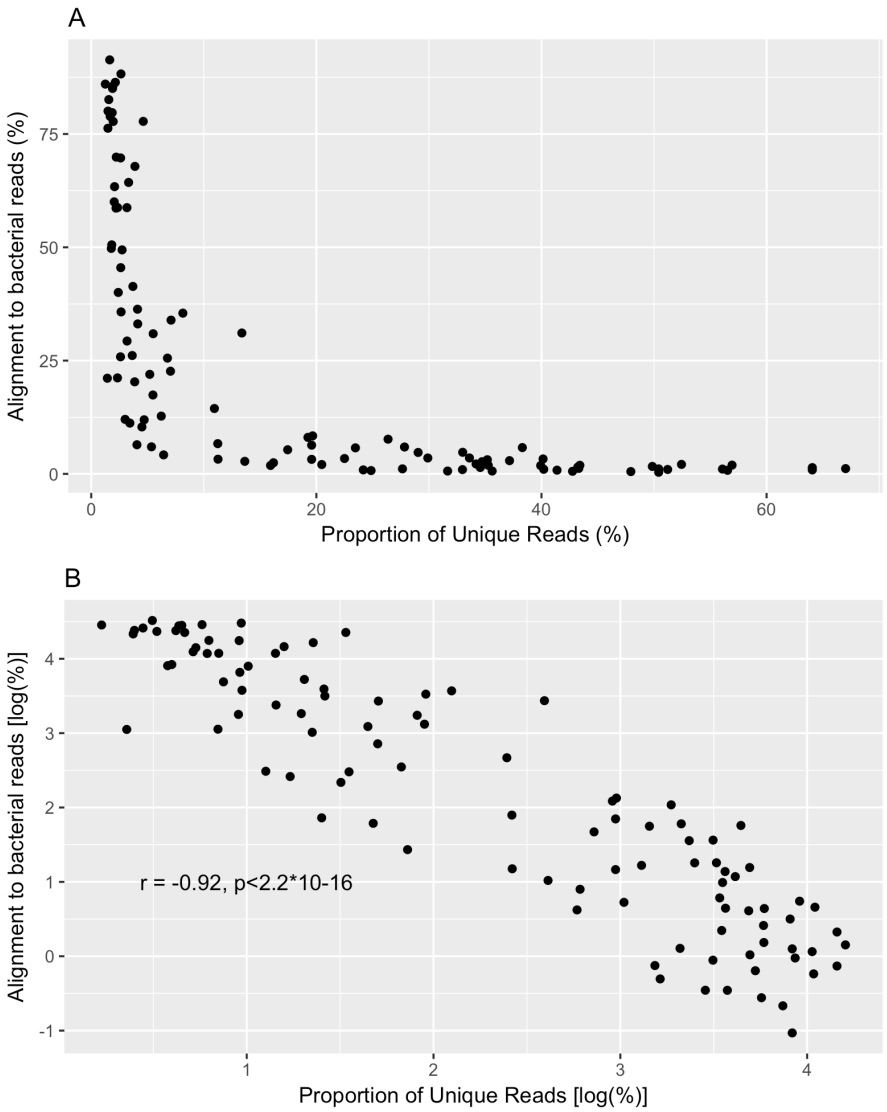


Figure 5-9: (A) Alignment of sequences to bacterial reads plotted against the proportion of unique reads. Each data point represents one FFPE sample. (B) Alignment of sequences to bacterial reads plotted against the proportion of unique reads following logarithmic transformation.

FastQC also generated data for the consequences of deduplication of the repeated sequences (Figure 5-10). This showed that once the repetitive sequences introduced by rRNA and bacterial RNA were removed, the duplication profiles were much improved for libraries of both tissue types.

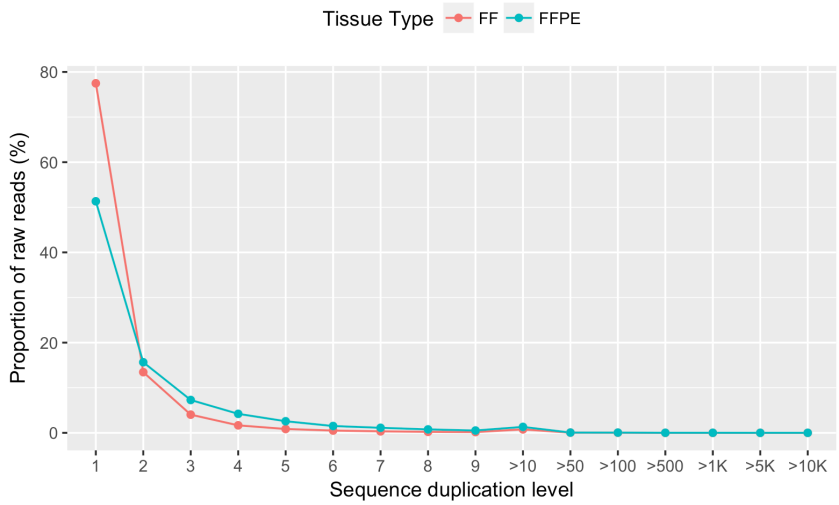


Figure 5-10: Median sequence duplication levels following FastQC deduplication. The graph demonstrates the proportion of raw reads at each sequence duplication level. This plot demonstrates that, after removal of duplicate sequences, the proportion of unique reads (sequence duplication level=1) was very high (50% for FFPE and 80% for FF samples).

5.3.8.5 FastQC: Adapter read-through and insert length

Adapter content increased towards the end of each read in both FFPE and FF samples. This was more evident in the FFPE than the FF samples, reaching statistical significance at every read position beyond the sixteenth base. By base position 100, the FFPE samples had a median of 6.7% reads containing adapter sequence compared to 4.3% in the FF samples ($p<0.001$) (Figure 5-11).

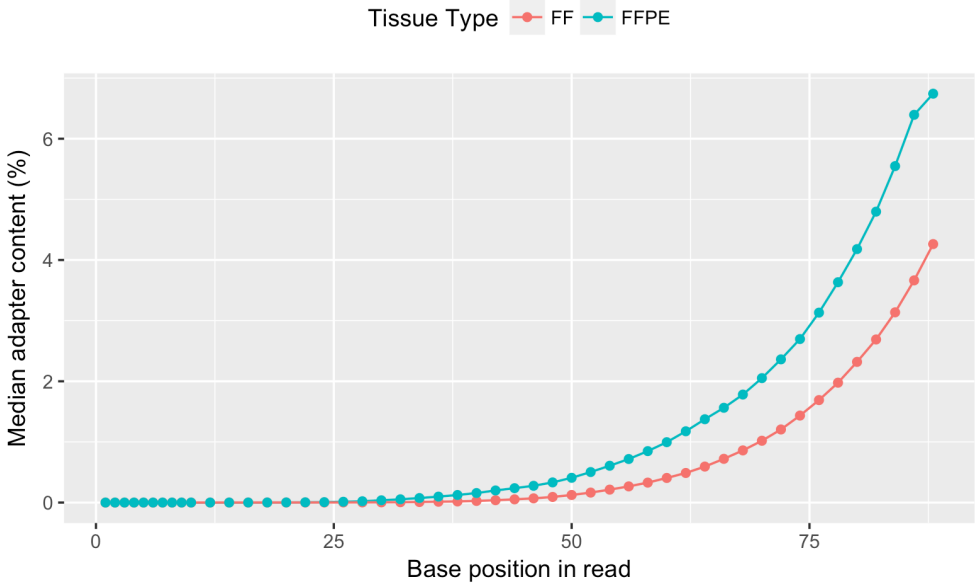


Figure 5-11: Median adapter content for FFPE and FF cohorts at each position in the read. As base position in the read increased, so did the median adapter content for both cohorts ($p<0.001$).

It was hypothesised that increased adapter content was seen more in the FFPE samples because the input RNA fragments were shorter. To test this, the insert length for each read was calculated using Picard tools. The median insert length for the FFPE samples was significantly shorter at 128 bases (111-146 bases) versus 155 bases (142-190 bases) ($p<0.001$) (Figure 5-12).

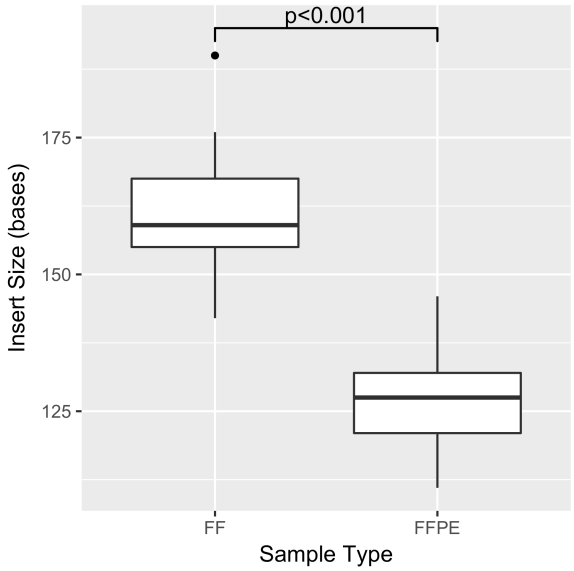


Figure 5-12: Box and whisker plot demonstrating the distribution of insert sizes for the FF versus FFPE samples. The FF cohort had significantly longer inserts ($p<0.001$), suggesting longer RNA fragments as input into the FF sequencing.

5.3.8.6 Other FastQC parameters

The other FastQC parameters tested were: per base sequence quality, per tile sequence quality, per base sequence content and sequence length; giving an indication of the quality of the sequencing run. All were acceptable, for all samples, showing no differences between FFPE and FF cohorts. This suggested that the sequencing runs were satisfactory and unaffected by RNA origin.

5.3.9 FFPE input material: scrolls versus cores

FFPE nucleic acids were extracted from a mixture of scrolls and cores. Scrolls of tissue were obtained for 44 (41.5%) samples and cores for 62 (58.5%). Samples extracted from scrolls and cores were compared to identify any key differences between these methods and to test the hypothesis that the level of bacterial read contamination could be affected by the input type.

There were no differences between scrolls and cores for the number of raw reads ($p=0.94$) (Figure 5-13A), input concentration of RNA ($p=0.95$) (Figure 5-13C) or age of FFPE block ($p=0.053$) (Figure 5-13B). There was a significant difference in proportion of bacterial reads between scrolls and cores (median 32.1% vs 4.12%,

p<0.001) (Figure 5-13D). Confirming the hypothesis that scrolls, taken from the surface of the block, were more likely than cores to be contaminated with bacteria.

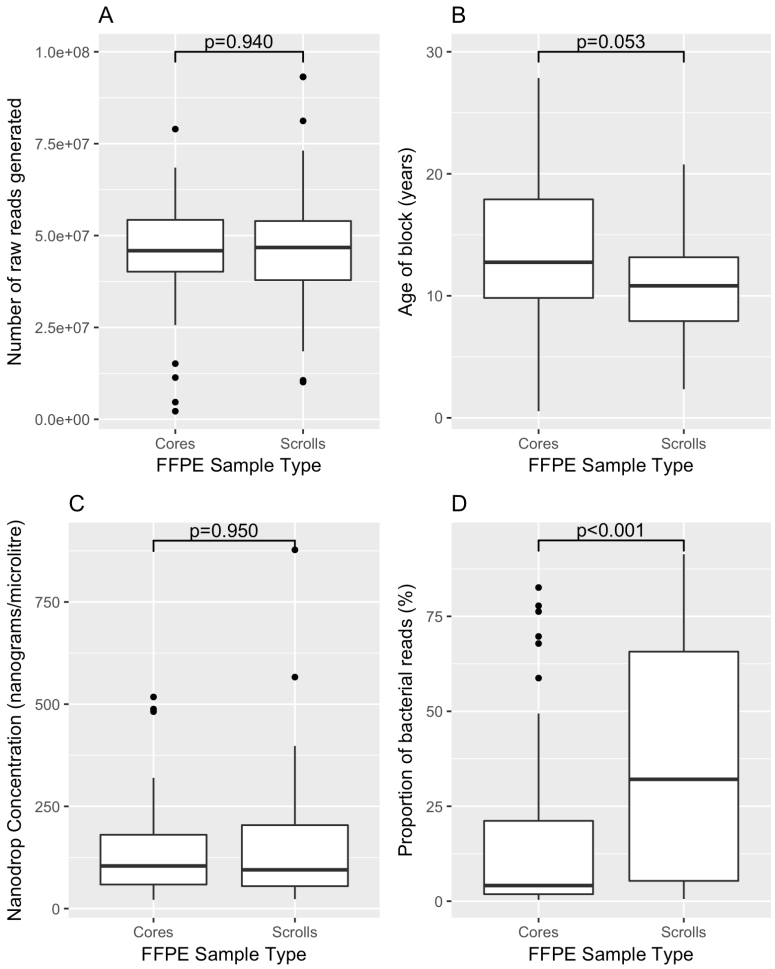


Figure 5-13: FFPE tissue parameters compared between samples extracted from cores versus scrolls of tissue with p-values indicated. (A) Number of raw reads generated. (B) Age of FFPE block. (C) Input RNA concentration. (D) Proportion of reads mapping to the bacterial genome. Scrolls were associated with significantly higher levels of bacterial read mapping compared to cores. No other significant differences were identified.

5.3.10 Counting aligned reads

The median proportion of aligned reads, for FFPE compared to FF samples, mapping to exons was 23.2% versus 43.7% (p<0.001); to introns, 51.6% versus 34.2% (p<0.001); and to intergenic regions, 12.4% versus 9.0% (p<0.001). There was no significant difference in the proportion of reads that were excluded for being chimeric, ambiguous or unmapped between the two tissue types (FFPE 11.3%, FF 11.4%, p=0.60). By only using reads that aligned to the human genome, bacterial reads and rRNA contamination were excluded.

5.3.11 Number of genes identified per sample

When counting genes identified by the presence of at least one RNA-Seq read, the FFPE samples identified a significantly lower median number of genes per sample than the FF cohort (24450 versus 28690, p<0.001). FFPE samples also had a much wider range of numbers of identified genes, suggesting greater sample to sample variability (FFPE range 1651-31270, FF range 22800-34460) (Figure 5-14).

A significant logarithmic relationship between number of genes and number of reads aligned to the human genome was evident in both cohorts, but much stronger for FFPE (r=+0.94, p<0.001) (Figure 5-15) than for FF samples (r=+0.52, p<0.001).

For the FFPE samples, greater RNA input concentration was weakly associated with the detection of more genes (r=+0.29, p=0.002). There was no correlation between the number of genes detected and the age of the sample (r=-0.16, p=0.13).

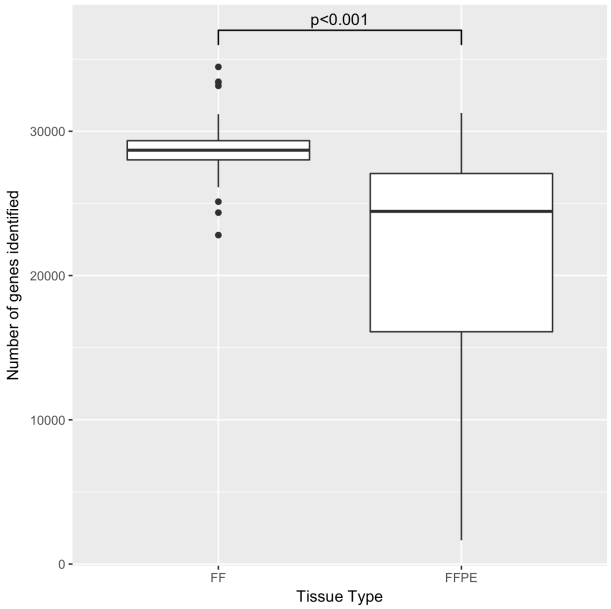


Figure 5-14: Distribution of numbers of identified genes with at least one read, compared between the fresh frozen (FF) and FFPE cohorts. Genes were identified more frequently and more reliably across samples in the FF cohort.

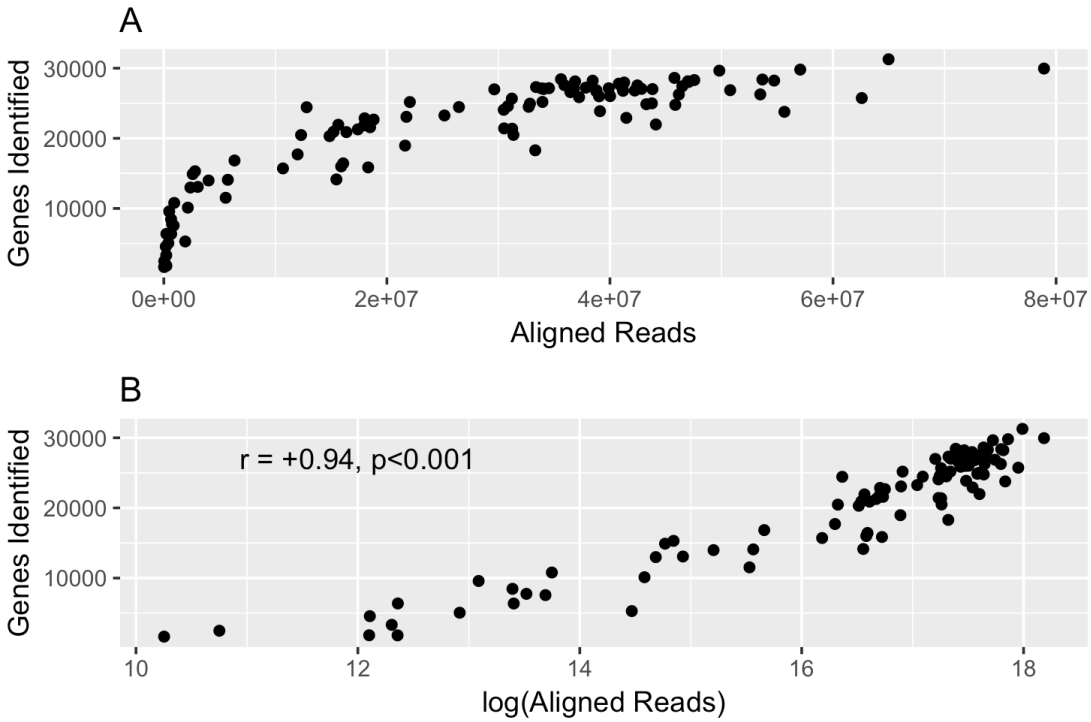


Figure 5-15: Relationship between number of genes identified with at least one read and number of reads aligned to the human genome for the FFPE samples. Each black dot represents one FFPE sample. A: Untransformed data. B: log transformed numbers of aligned reads demonstrating a linear relationship.

5.3.12 Correlations between biological replicates

The ENCODE RNA sequencing guidelines (ENCODE, 2016) recommend standards for correlation coefficients for biological replicates. The correlation is measured by comparing the expression levels for each gene with levels for other samples in the dataset. Correlation coefficients were calculated between all FFPE samples and between all FF samples to assess the reliability of the biological replicates.

The raw FFPE data had inter-sample correlation coefficients varying between 0.13 and 1.00, with a median of 0.85. The raw FF data, in contrast, had a narrower range of coefficients (0.51 – 1.00 with a median of 0.88). Both cohorts were skewed towards higher values. This suggested that the raw FFPE data demonstrated more variability between biological replicates than the FF samples (Figure 5-16).

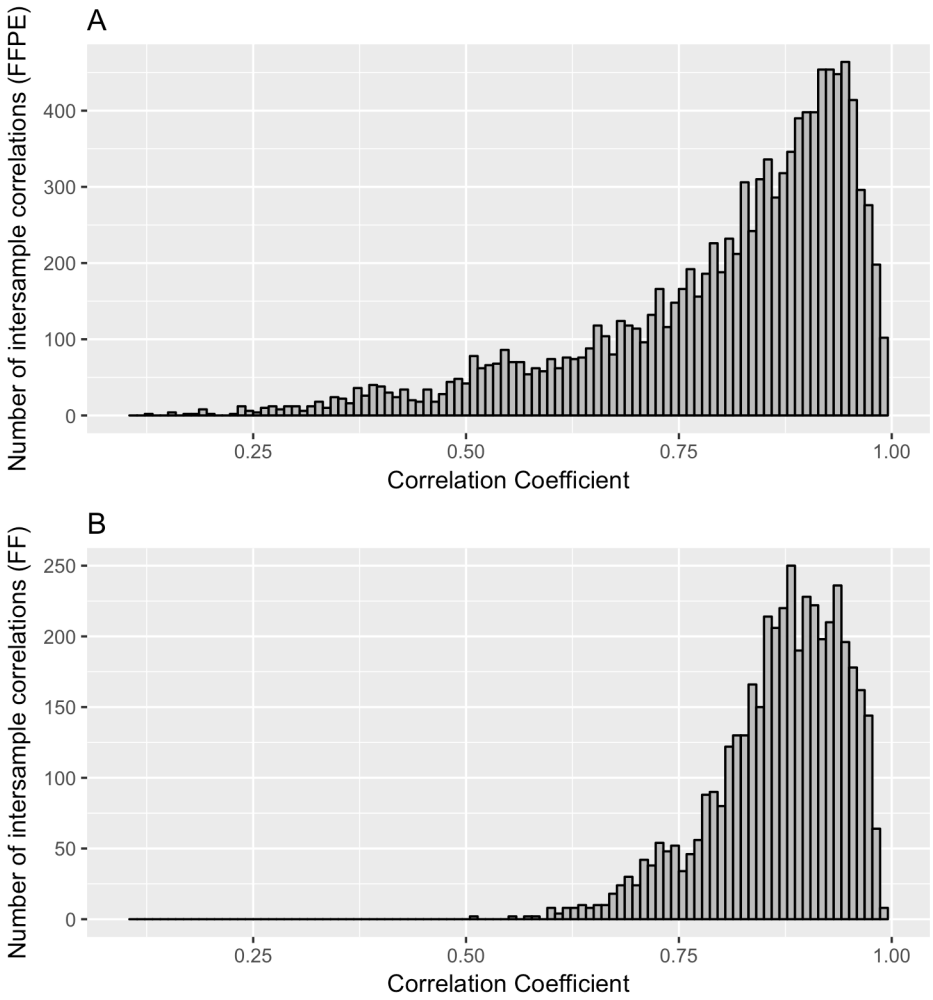


Figure 5-16: Histograms of inter-sample gene expression correlation coefficients for raw RNA-seq data for (A) FFPE samples and (B) FF samples.

When assessing the normalised and transformed data generated by the DESeq2 R-log transformation, the correlation coefficients were less variable in both sample types when compared with the raw data. The FFPE tissue again showed more variability than the FF samples, with the range of correlation coefficients being 0.57 – 1.00 (median 0.95) for FFPE and 0.91 – 1.00 (median 0.99) for FF. For both cohorts, the coefficients were skewed towards higher values (Figure 5-17).

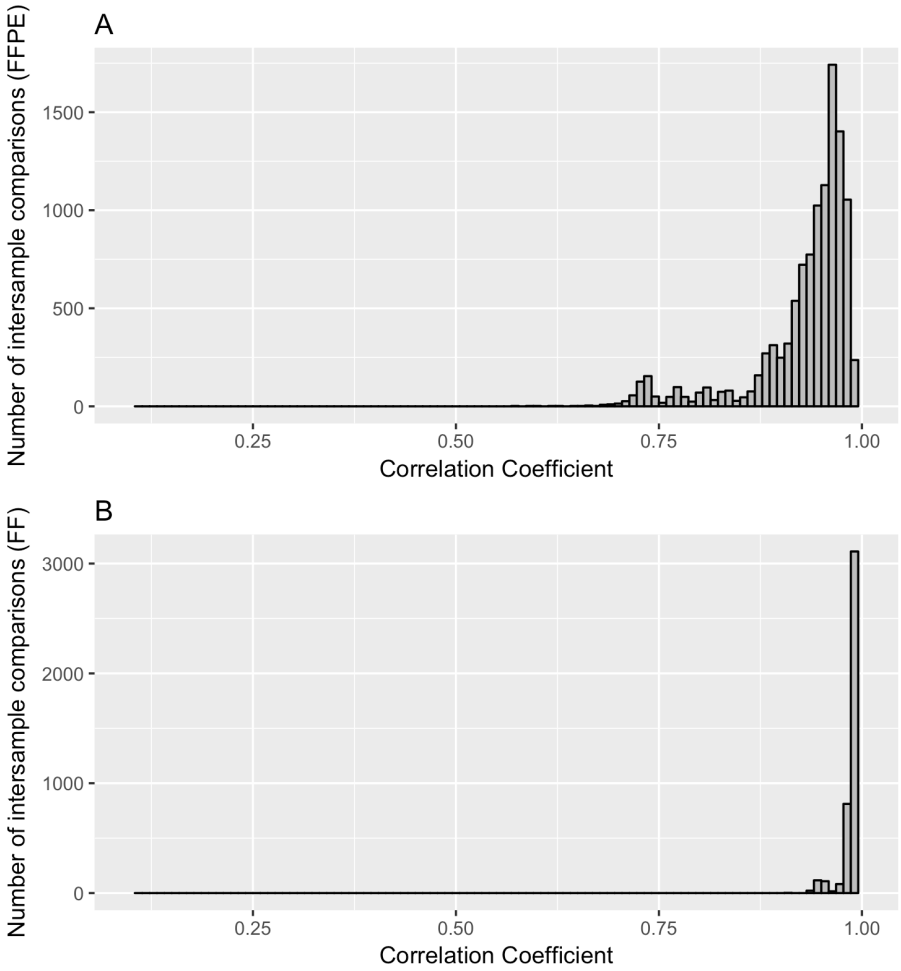


Figure 5-17: Histograms of inter-sample gene expression correlation coefficients for DESeq2 RLD transformed RNA-seq data for (A) FFPE samples and (B) FF samples.

5.3.13 Correlations between technical replicates

14 of the samples had matched FFPE and FF tissue and were used as technical replicates. These samples were compared using raw data and the DESeq2 R-log transformation. Overall the mean correlation coefficient for the raw data was relatively high at 0.85 (range 0.6 – 0.98). Once data had been normalised, in DESeq2, it improved to 0.98 (range 0.96-0.99); meaning that final correlations were very strong between matched FFPE and FF samples (Figure 5-18).

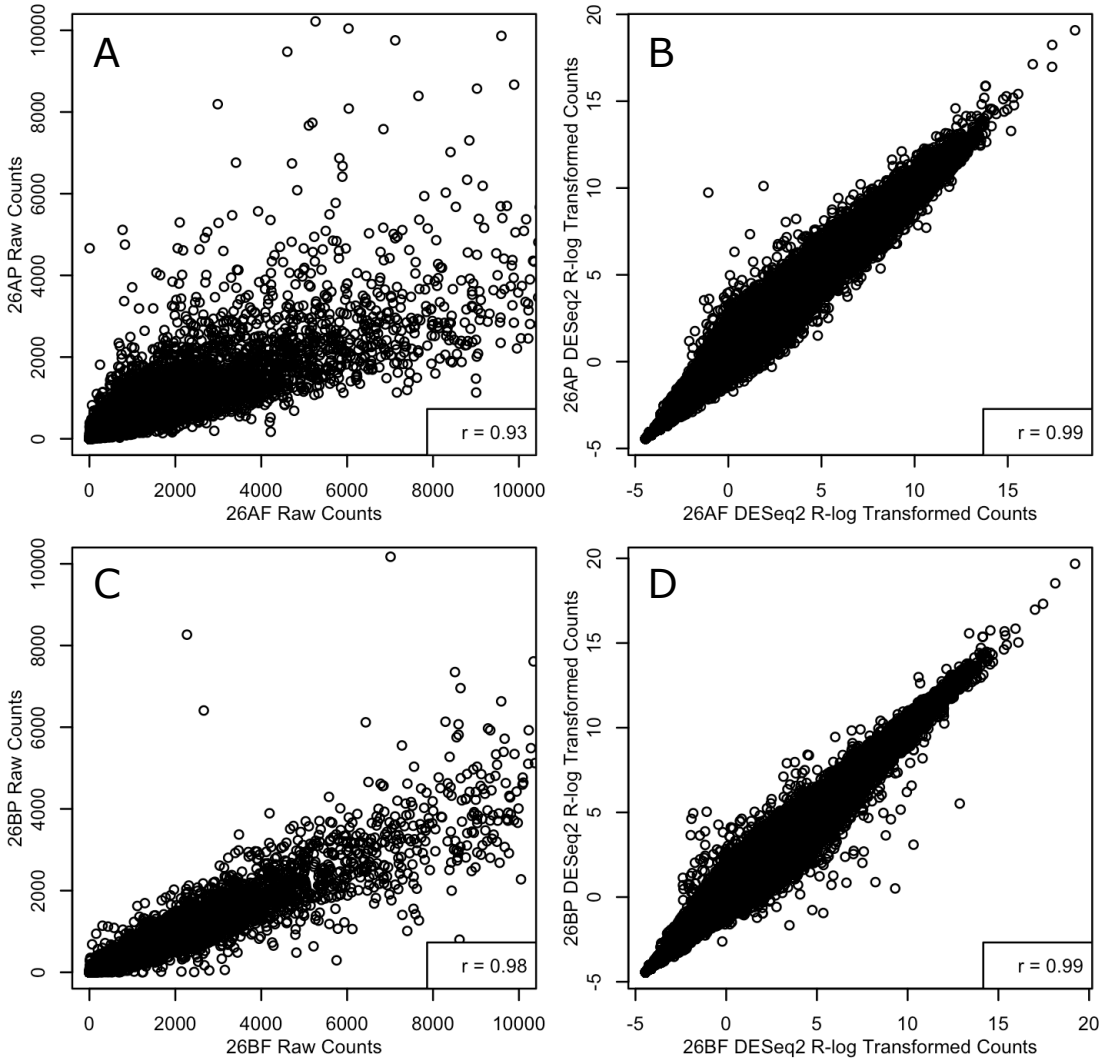


Figure 5-18: Representative examples of correlations between FF and FFPE technical replicates. (A) Raw counts from sample 26A FF (26AF) versus FFPE (26AP). (B) Normalised counts from sample 26A FF (26AF) versus FFPE (26AP). (C) Raw counts from sample 26B FF (26BF) versus FFPE (26BP). (D) Normalised counts from samples 26B FF (26BF) versus FFPE (26BP).

5.3.14 Differential expression between technical replicates

EdgeR was used to perform differential expression analysis between paired FFPE and FF samples. 1188 genes were differentially expressed at FDR <0.05 and fold change >2 level; this represented 6.4% of all the genes tested.

The differentially expressed genes were processed through the DAVID Gene Ontology database. For genes upregulated in the FFPE versus FF samples, the significant ontologies were "positive regulation of transcription from RNA polymerase II promoter" (FDR <0.001) and "transcription, DNA-templated' (FDR

0.048). For the genes downregulated in the FFPE versus FF samples, the significant ontologies were "ion membrane transport" (FDR <0.001), "chemical synaptic transmission" (FDR <0.001), "cell adhesion" (FDR <0.001), "potassium ion transmembrane transport" (FDR 0.002), "long-term synaptic potentiation" (FDR 0.006), "potassium ion transport" (FDR 0.011) and "homophilic cell adhesion via plasma membrane adhesion molecules" (FDR 0.024).

5.3.15 Hierarchical clustering

In view of the quality control differences described, unsupervised hierarchical clustering was performed separately on the FFPE and FF cohorts using R-log transformed data (Figure 5-19 and Figure 5-20).

FFPE Cohort Clustering

Three distinct groups were visible within the FFPE dendrogram (Figure 5-19).

The largest group contained 58 samples; 51 (88%) in the posterior fossa, 6 (10%) in the supratentorium, and 1 (2%) in the spine. The next largest group contained 27 samples; 23 (85%) in the supratentorium, 3 (11%) in the posterior fossa and 1 (4%) in the spine. These two groups were named PF and ST respectively, on the basis of the statistically significant difference in the locations of the tumours in each group ($p<0.001$, chi-square test). The age at diagnosis of primary disease for patients in the PF group was significantly lower than in the ST group (median 26.5 versus 73 months, $p=0.026$). The PF group was also noted to divide into two subclusters; which were named PF1 and PF2.

The third group contained 21 samples, 6 (29%) in the supratentorium and 15 (71%) in the posterior fossa. These were exclusively poor-quality RNA-seq samples, as judged by the proportions of reads aligned to the human and bacterial genomes. This was therefore called the Quality Control (QC) Fail group. The median number of human aligned reads was 35,913,105 in the PF group, 34,310,500 in the ST group, and 663,115 in the QC Fail group. This represented a significant difference between the PF and QC Fail group ($p<0.001$), ST and QC Fail group ($p<0.001$), but not the PF and ST group ($p=0.660$). The median percentage of reads aligned to the bacterial genome was 5.32% in the PF group, 4.48% in the ST group, and 77.79% in the QC Fail group (PF versus QC Fail group $p<0.001$, ST versus QC Fail group $p<0.001$, and PF versus ST group $p=0.679$).

Adequate DNA methylation class predictions were available for 54/58 (93%) tumours in the PF group, 22/27 (81%) tumours in the ST group and 14/21 (66.7%) in the QC fail group.

The PF group consisted of:

- 46 (85%) EPN_PFA;
- 4 (7%) EPN_YAP;
- 2 (4%) EPN_PFB;
- 1 (2%) EPN_MPE;
- 1 (2%) GBM_RTK_MYCN.

The ST group consisted of:

- 13 (59%) EPN_RELA;
- 7 (32%) HGNET_MN1;
- 1 (5%) HGNET_BCOR;
- 1 (5%) DIG1.

The QC Fail group consisted of:

- 10 (71%) EPN_PFA;
- 4 (29%) EPN_RELA.

The QC fail group was significantly less likely to produce an adequate class prediction compared to the PF and ST groups (p=0.001). Samples in the QC Fail group were subsequently removed from further analysis.

FF Cohort Clustering

Three distinct groups were evident within the FF dendrogram (Figure 5-20).

The first group contained 30 samples; 25 (83%) posterior fossa tumours, 2 (7%) supratentorial tumours, and 3 (10%) spinal tumours. The second group contained 28 samples; 24 (85%) posterior fossa tumours, 3 (11%) supratentorial tumours, and 1 (4%) spinal tumour. The third group contained 7 samples; 5 (71%) supratentorial and 2 (29%) posterior fossa tumours. There were statistically significant differences in tumour location between the first and third groups (p=0.001), second and third groups (p=0.004), but not between first and second groups (p=1.000). On the basis of the predominant tumour locations and statistical testing, groups one, two and three were named PF1, PF2 and ST respectively.

Two supratentorial samples (1C and 25C) clustered away from the rest of the samples and were removed from further analysis.

The two PF groups collectively contained DNA methylation predictions for 43/58 (74%) of tumours and consisted of:

- 37 (86%) EPN_PFA;
- 2 EPN_MPE (5%);
- 2 EPN_YAP (5%);
- 2 HGNET_MN1 (5%).

The ST group had few samples with DNA methylation predictions (3/7) but contained:

- 2 EPN_PFA;
- 1 EPN_RELA.

The only tumour with an EPN_RELA prediction clustered into the ST group. There were no EPN_PFB tumours in the dataset.

Comparisons between the FFPE and FF clusters

Both cohorts contained an ST group and two PF groups. No significant differences were identified when comparing the composition of tumour locations between the FF and FFPE PF groups (p=1.000, chi-square test) or between the FF and FFPE ST groups (p=0.268, chi-square test). Additionally, the FFPE cohort demonstrated a cluster of poor quality samples, which was not identified in the FF cohort.

When a comparison was made between the samples sequenced from both FFPE and FF material; 9/13 samples that clustered within one of the three main groups (PF1, PF2 or ST) clustered in the same group in FF and FFPE cohorts. This was unlikely to have occurred by chance (p=0.007).

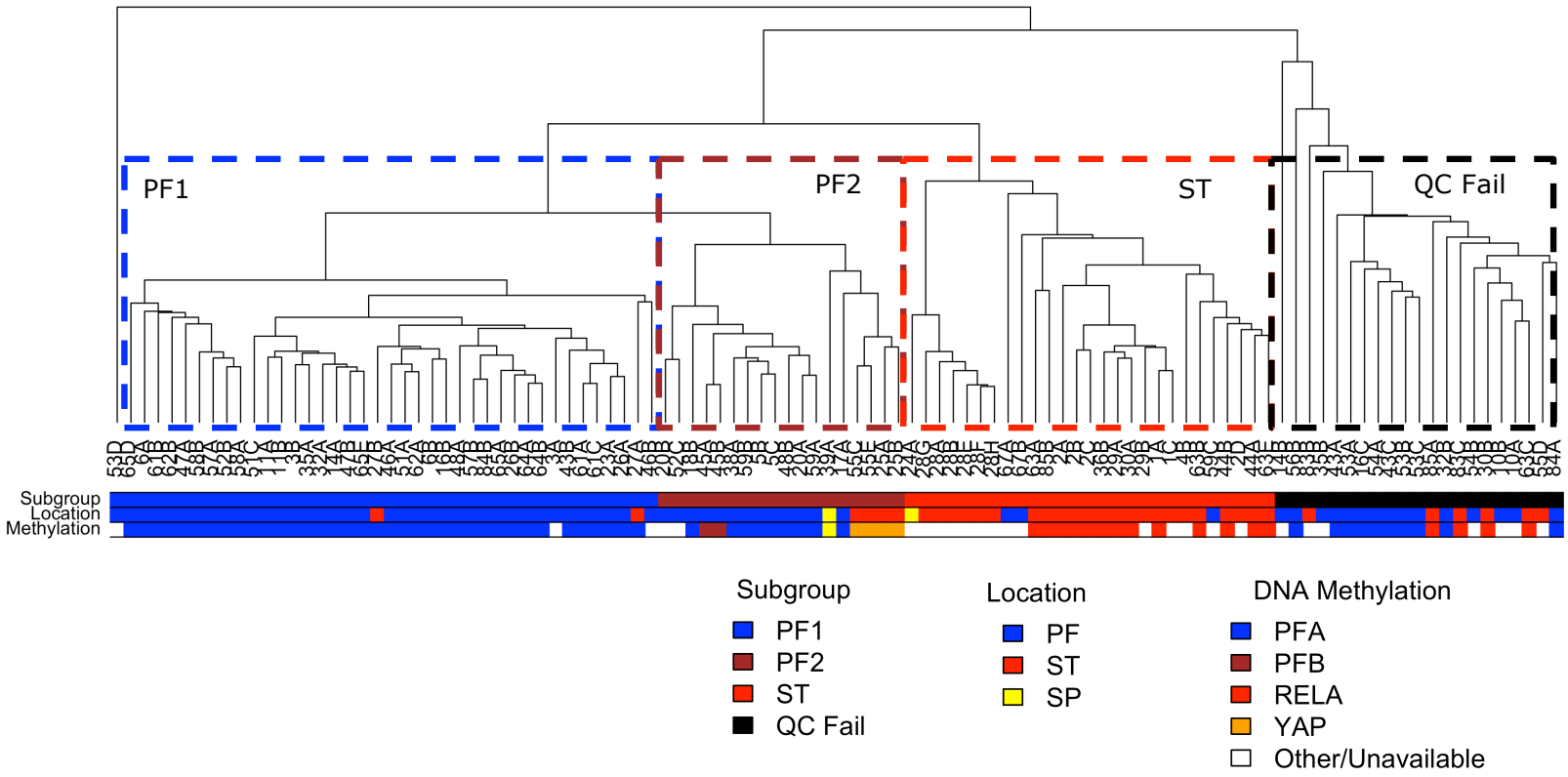


Figure 5-19: Unsupervised hierarchical clustering of all (mixed primary and recurrent) FFPE samples based on Euclidean distances, Ward's algorithm and all expressed genes. Three subgroups were initially identified; PF which subdivided into PF1 and PF2 containing significantly more PF location samples than the other groups; ST which contained predominantly ST samples and QC fail which contained samples with significantly lower human genome alignment than the other samples. PF1 contained exclusively EPN_PFA DNA methylation profiles and PF2 contained predominantly EPN_PFA profiles.

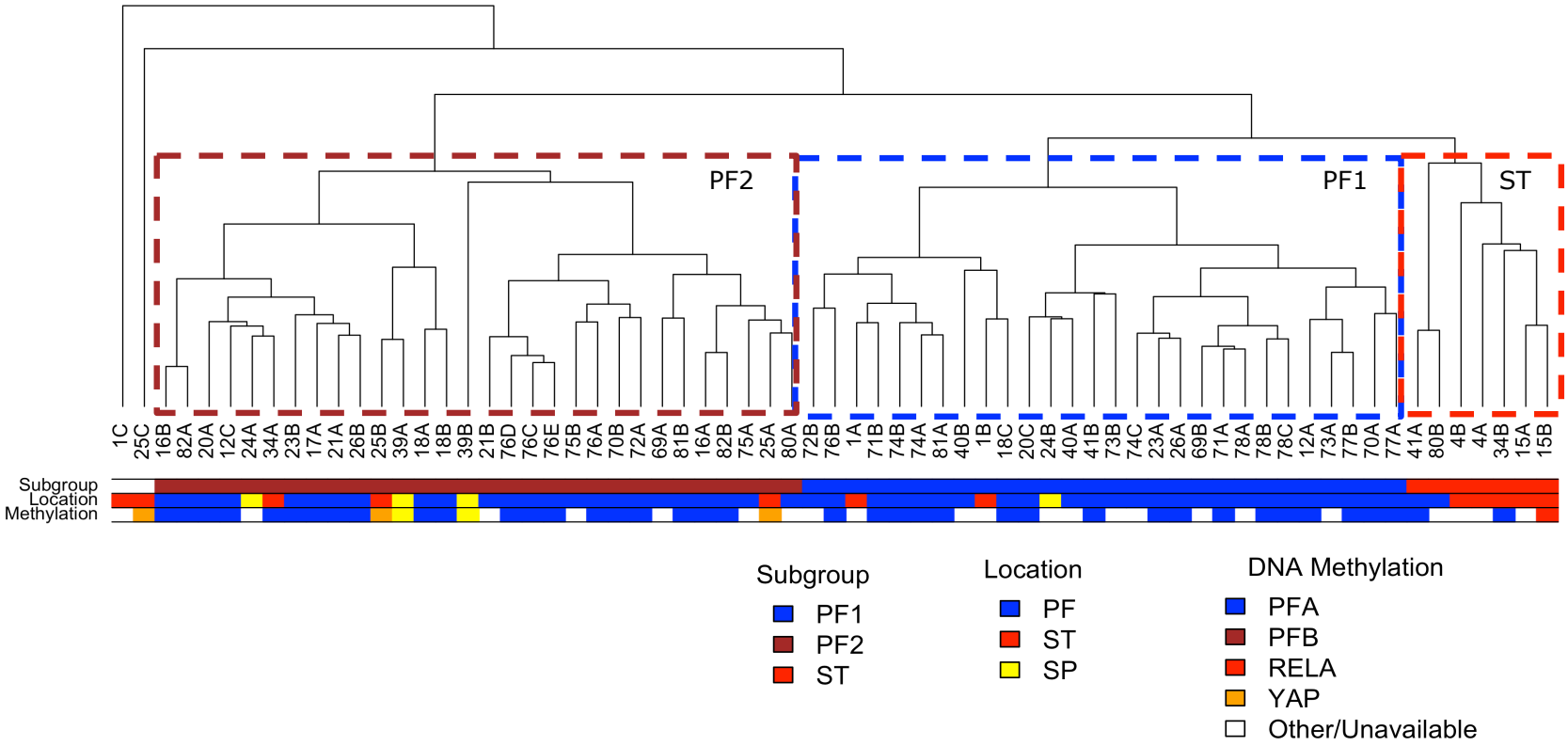


Figure 5-20: Unsupervised hierarchical clustering of all (mixed primary and recurrent) FF samples based on Euclidean distances, Ward's algorithm and all expressed genes. Three subgroups were initially identified; PF1 and PF2 containing significantly more PF location samples than the other groups; ST which contained predominantly ST samples. PF1 contained almost exclusively EPN_PFA DNA methylation profiles and PF2 contained predominantly EPN_PFA profiles. The groups reflected those seen in the FFPE sample clustering.

It was hypothesised that the two PF groups in both cohorts may demonstrate the EPN_PFA and EPN_PFB gene expression patterns described by previous authors (Pajtler et al., 2015; Wani et al., 2012; Witt et al., 2011). Further analysis of gene expression and DNA methylation patterns was therefore undertaken.

5.3.16 PF group features

DNA Methylation Profiles

Across both datasets PF1 contained almost exclusively EPN_PFA predictions, whilst PF2 contained EPN_PFA as the most common prediction (Table 5-4). On this basis it was evident that, PF1 and PF2 represented two mainly EPN_PFA groups, rather than separate EPN_PFA and EPN_PFB groups.

EPN_YAP tumours arise in the supratentorium (Pajtler et al., 2015). However, in both datasets, all of the EPN_YAP tumours clustered into the PF2 group rather than the ST group.

Methylation Prediction	PF1		PF2	
	FF Cohort (%)	FFPE Cohort (%)	FF Cohort (%)	FFPE Cohort (%)
EPN_PFA	17 (94)	36 (100)	20 (80)	10 (57)
EPN_RELA	-	-	-	-
EPN_PFB	-	-	-	2 (11)
EPN_YAP	-	-	2 (8)	4 (22)
EPN_MPE	-	-	2 (8)	1 (6)
Other	1 (6%)	-	1 (4)	1 (6)

Table 5-4: Illustration of composition by DNA methylation group of the PF1 and PF2 subgroups in the FF and FFPE datasets. Percentage calculated as a proportion of samples in each group with available DNA methylation data. DNA methylation data unavailable for 10 samples in FF PF1, five samples in FF PF2, three samples in FFPE PF1 and one sample in FFPE PF2.

Survival

OS was compared between the PF1 and PF2 groups across both datasets. The median OS for PF1 was 77 months versus 114 months for the PF2 group ($p=0.115$, log-rank test). There was no difference in time to first recurrence ($p=0.486$).

Differential Expression Analysis

Differential expression and gene ontology analyses were performed, comparing tumours with PF location in PF1 to those with PF location in PF2. In the FFPE dataset 4606 genes were differentially expressed at FDR <0.05 between the two

groups, representing 30.7% of all genes tested. In the FF dataset, 5518 genes were found to be differentially expressed at FDR <0.05 between the two groups, representing 27% of all genes tested. These results suggest major transcriptional differences between PF1 and PF2 in both datasets.

The 39 signature genes of EPN_PFA and EPN_PFB subgroups were then investigated. This confirmed that PF1 in both FFPE and FF cohorts overlapped with 15/20 (75%) of the EPN_PFA signature genes. PF2 in both FFPE and FF cohorts overlapped with 19/19 (100%) of the EPN_PFB signature genes ($p<0.001$, hypergeometric test) (Table 5-5). Supervised hierarchical clustering, using these 39 genes, recapitulated the PF gene expression subgroup assignment in both datasets.

Signature Gene (PF Group)	**FF Cohort**		**FFPE Cohort**	
	LFC (PF2 vs PF1)	**FDR**	**LFC (PF2 vs PF1)**	**FDR**
CYP1B1 (PFA)	-2.12	<0.001	-2.88	<0.001
SERPINA5 (PFA)	-1.87	<0.001	-1.78	0.011
TGFBI (PFA)	-1.64	<0.001	-2.21	<0.001
COL4A1 (PFA)	-1.11	0.002	-1.49	<0.001
COL4A2 (PFA)	-0.98	0.005	-1.55	<0.001
VEGFA (PFA)	-1.49	<0.001	-2.56	<0.001
HILPDA (PFA)	-1.07	0.002	-1.29	<0.001
PRSS23 (PFA)	-1.43	<0.001	+0.60	0.163
TIMP1 (PFA)	-2.66	<0.001	-0.72	0.139
NNMT (PFA)	-2.80	<0.001	-2.28	<0.001
C1S (PFA)	-1.91	<0.001	-1.43	<0.001
CFB (PFA)	-1.22	0.013	-2.32	<0.001
SERPINA3 (PFA)	-1.46	0.002	-1.65	0.007
TAGLN (PFA)	-2.21	<0.001	-0.94	0.008
CXCL2 (PFA)	-1.69	<0.001	-1.76	<0.001
OLFML1 (PFA)	-0.85	0.074	-2.23	<0.001
CHODL (PFA)	+0.45	0.419	-0.89	0.171
NSG1 (PFA)	-0.29	0.520	-0.37	0.518
BMP5 (PFA)	+0.43	0.532	-1.29	0.107
LAMA2 (PFA)	-0.30	0.513	-1.34	0.001
IQCA1 (PFB)	+1.67	<0.001	+1.88	<0.001
CCDC170 (PFB)	+1.58	<0.001	+2.02	<0.001
FHOD3 (PFB)	+1.10	<0.001	+1.59	<0.001
ATP4B (PFB)	+3.00	<0.001	+3.35	<0.001
CDS1 (PFB)	+1.72	<0.001	+2.23	<0.001

SHANK2 (PFB)	+2.33	<0.001	+2.47	<0.001
DNAI2 (PFB)	+3.25	<0.001	+3.51	<0.001
SPEF1 (PFB)	+1.84	<0.001	+2.00	<0.001
DNAH6 (PFB)	+1.76	<0.001	+1.64	<0.001
SPAG8 (PFB)	+2.24	<0.001	+1.94	<0.001
AGBL2 (PFB)	+2.04	<0.001	+2.29	<0.001
ZBBX (PFB)	+2.08	<0.001	+2.13	<0.001
ADGB (PFB)	+2.39	<0.001	+2.52	<0.001
CFAP69 (PFB)	+1.61	<0.001	+1.71	<0.001
DNAI1 (PFB)	+2.07	<0.001	+1.77	<0.001
CCDC81 (PFB)	+1.52	<0.001	+1.66	<0.001
RIBC2 (PFB)	+1.57	<0.001	+1.33	<0.001
TUBA4B (PFB)	+3.10	<0.001	+3.81	<0.001
SH3GL3 (PFB)	+0.95	0.005	+1.70	<0.001

Table 5-5: Signature genes of the PFA and PFB molecular subgroups as defined by Witt et. al., compared to differentially expressed genes in groups PF1 and PF2 in the FFPE and FF cohorts. LFC: Log2 Fold Change. FDR: False Discovery Rate.

To further investigate the overlap of PF1 and PF2 with the expression patterns of EPN_PFA and EPN_PFB, more comprehensive lists of differentially expressed genes associated with the nine ependymoma subgroups were obtained (Pajtler et al., 2015). The Pajtler dataset contained 715 EPN_PFA associated genes, 696 of which could be evaluated in both FFPE and FF cohorts; and 1267 EPN_PFB associated genes, 1221 of which could be evaluated in both FFPE and FF cohorts. The overlap between the Pajtler genes and FFPE and FF cohorts were evaluated using the hypergeometric test.

In the FFPE cohort, 396 (56.8%) EPN_PFA genes were significantly upregulated in the PF1 compared to the PF2 group ($p<0.001$), and 633 (90.9%) genes exhibited the same direction of fold change as seen in EPN_PFA ($p<0.001$). 706 (57.8%) EPN_PFB genes were significantly upregulated in the PF2 compared to the PF1 group ($p<0.001$), and 1060 (86.8%) exhibited the same direction of fold change as seen in EPN_PFB ($p<0.001$) (Figure 5-21).

In the FF cohort, 320 (46.0%) EPN_PFA genes were significantly upregulated in the PF1 compared to the PF2 group ($p<0.001$), and 544 (78.2%) exhibited the same direction of fold change as seen in EPN_PFA ($p<0.001$). 657 (53.8%) EPN_PFB genes were significantly upregulated in the PF2 compared to PF1 group ($p<0.001$), and 1059 (86.7%) exhibited the same direction of fold change as seen in EPN_PFB ($p<0.001$) (Figure 5-21).

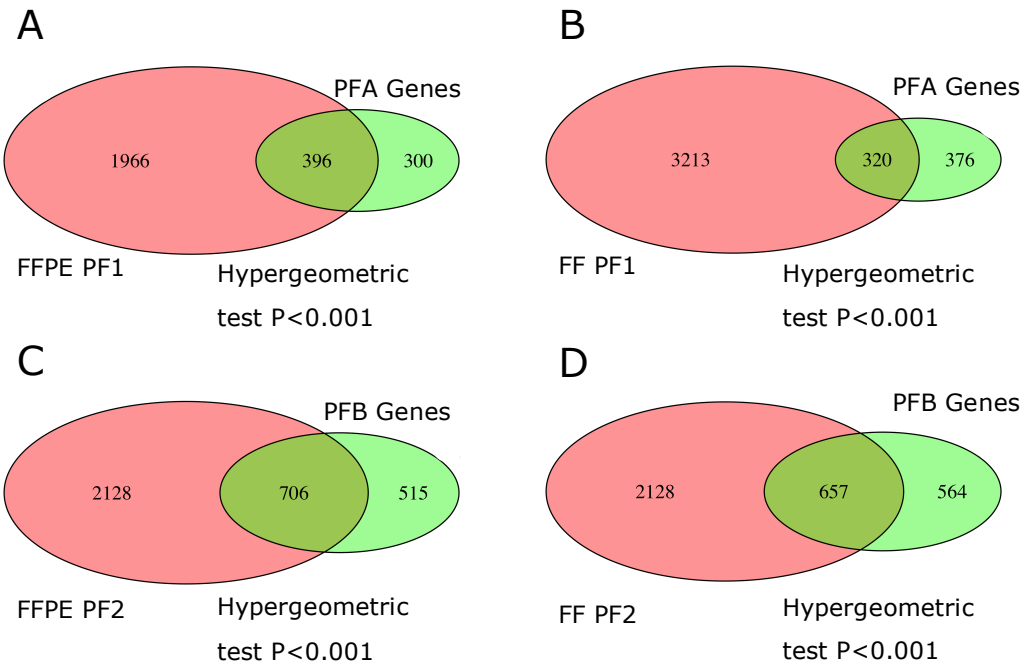


Figure 5-21: Summary of the overlap between PF1 and PF2 groups with PFA and PFB gene expression profiles (Pajtler et. al. 2015). PF1 and PFA overlap in (A) the FFPE dataset and (B) the FF dataset. PF2 and PFB overlap in (C) the FFPE dataset and (D) the FF dataset. All overlaps were highly significant as determined by the hypergeometric test, $p<0.001$. Datasets generated by this study appear in red, the Pajtler dataset genes appear in green.

To further validate the two PF subgroups, gene level comparisons were made between the FFPE and FF cohorts. The PF1 compared to the PF2 subgroup had 1062 shared genes significantly upregulated in both FFPE and FF datasets ($p<0.001$). The PF2 compared to the PF1 subgroup had 1585 shared genes significantly upregulated in both FFPE and FF datasets ($p<0.001$) (Figure 5-22).

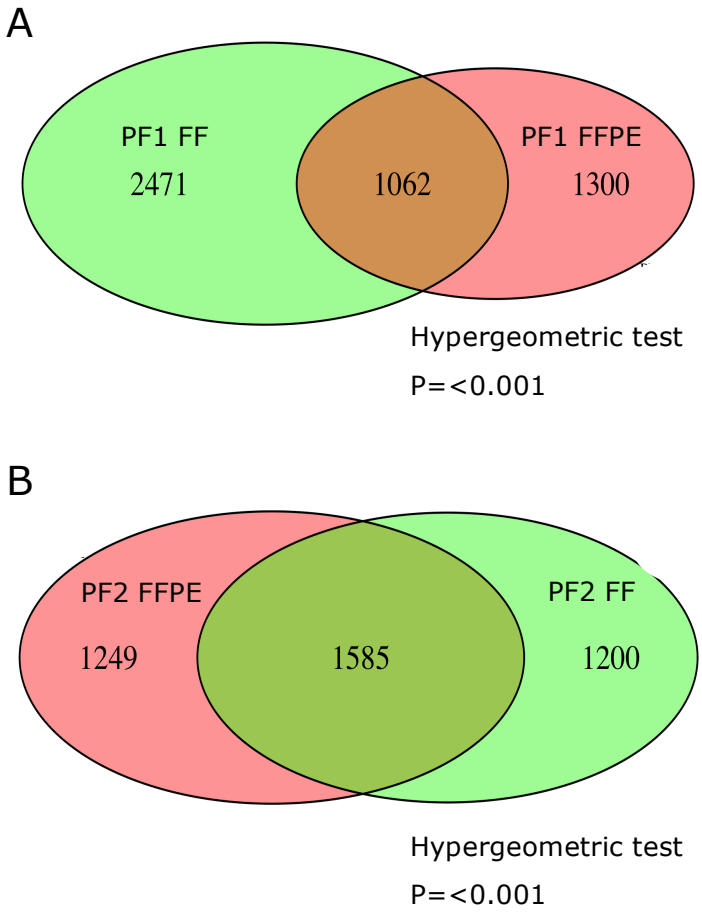


Figure 5-22: Venn diagrams illustrating the overlap between significantly differentially expressed genes in the two posterior fossa subgroups in the FFPE and FF datasets. (A) The PF1 (EPN_PFA like) subgroup demonstrated 1062 overlapping differentially expressed genes between FFPE and FF datasets plus 2471 that were exclusive to the FF samples and 1300 that were exclusive to the FFPE samples. (B) The PF2 subgroup demonstrated 1585 overlapping differentially expressed genes between the FFPE and FF datasets plus 1200 that were exclusive to the FF samples and 1249 that were exclusive to the FFPE samples.

Ontology of PF1 and PF2 subgroups

Gene ontology analyses were performed on group PF1 and PF2 genes using DAVID. Analyses was performed separately on the FFPE and FF datasets and shared terms were identified.

52 statistically significant gene ontologies, in the category of biological process, were identified for group PF1; including the immune and inflammatory response, organisation of the extracellular matrix, and cell adhesion (Table 5-6). Group PF2 was associated with 15 statistically significant ontologies which were less diverse, being mainly related to cilia and microtubule function (Table 5-7).

GO Term: PF1	FF FDR	FFPE FDR
GO:0006952~defense response	<0.001	0.003
GO:0006954~inflammatory response	<0.001	<0.001
GO:0006955~immune response	<0.001	0.020
GO:0007267~cell-cell signaling	<0.001	0.004
GO:0009611~response to wounding	<0.001	<0.001
GO:0007610~behavior	<0.001	0.003
GO:0055082~cellular chemical homeostasis	<0.001	0.041
GO:0055066~di-, tri-valent inorganic cation homeostasis	<0.001	0.044
GO:0006873~cellular ion homeostasis	<0.001	0.049
GO:0048545~response to steroid hormone stimulus	<0.001	0.015
GO:0042127~regulation of cell proliferation	<0.001	<0.001
GO:0022610~biological adhesion	<0.001	<0.001
GO:0007155~cell adhesion	<0.001	<0.001
GO:0007626~locomotory behavior	<0.001	0.003
GO:0019932~second-messenger-mediated signaling	<0.001	<0.001
GO:0032101~regulation of response to external stimulus	<0.001	0.007
GO:0030198~extracellular matrix organization	<0.001	<0.001
GO:0043062~extracellular structure organization	<0.001	<0.001
GO:0044057~regulation of system process	<0.001	<0.001
GO:0003013~circulatory system process	<0.001	0.002

Table 5-6: List of 20 most significant biological process gene ontology terms upregulated in group PF1 compared to PF2 in both FFPE and FF datasets.

GO Term: PF2	FF FDR	FFPE FDR
GO:0000003~reproduction	<0.001	0.002
GO:0001539~ciliary or flagellar motility	<0.001	<0.001
GO:0022414~reproductive process	<0.001	0.002
GO:0007018~microtubule-based movement	<0.001	<0.001
GO:0019953~sexual reproduction	<0.001	0.003
GO:0007276~gamete generation	<0.001	0.009
GO:0007017~microtubule-based process	<0.001	0.006
GO:0032501~multicellular organismal process	<0.001	<0.001
GO:0032504~multicellular organism reproduction	<0.001	0.016
GO:0048609~reproductive process in a multicellular organism	<0.001	0.016
GO:0006811~ion transport	<0.001	<0.001
GO:0006813~potassium ion transport	0.002	0.005
GO:0015672~monovalent inorganic cation transport	0.006	0.004
GO:0030001~metal ion transport	0.034	<0.001
GO:0006812~cation transport	0.049	0.009

Table 5-7: List of all 15 significant biological process gene ontology terms upregulated in group PF2 compared to PF1 in FFPE and FF datasets.

5.3.17 Summary of FFPE and FF clustering

- Significant levels of overlap were confirmed between FFPE and FF cohorts for both PF subgroups (PF1 and PF2);
- PF1 was a predominantly posterior fossa group consisting of up to 100% EPN_PFA tumours. There was significant overlap with EPN_PFA gene expression patterns and ontologies;
- PF2 was a predominantly posterior fossa group consisting of up to 80% EPN_PFA tumours, but also a number of non-EPN_PFA entities, including EPN_YAP and EPN_PFB tumours. There was significant overlap with EPN_PFB gene expression patterns and ontologies;
- ST was a predominantly supratentorial group containing all of the EPN_RELA tumours, and tumours of older children.

5.4 Discussion and conclusions

The primary aims of this chapter were to establish:

(1) Is RNA-seq, from FFPE tissue, feasible on a large scale?

(2) How does the quality of the data compare to FF samples?

(3) Is the data of adequate quality to include in investigating other research questions?

(4) What are the potential pitfalls of this approach and can recommendations be made to advise future research?

This chapter has demonstrated that paediatric brain tumour samples, fixed in formalin and embedded in paraffin for up to 30 years, can be profiled by RNA sequencing to produce robust data that reflect known findings. Analysis of this large cohort of 106 samples provided adequate data to be able to identify poorly performing samples. The key challenge in working with this type of material was the risk of bacterial contamination of the RNA, particularly when extracting from scrolls instead of tissue cores.

Parameters used for previously published FFPE RNA sequencing studies were heterogeneous (Table 5-1), with varying read lengths; read depths; age distributions; and a mixture of single- and paired-end sequencing. The findings of this chapter therefore contribute to the technical knowledge of FFPE RNA sequencing and describe the largest cohort ever presented using brain material.

5.4.1 Is RNA-seq from FFPE feasible on a large scale?

Many of the previous FFPE RNA-seq studies have been limited in size, and therefore unable to answer questions about its scalability. The median size of previous studies was just five samples, with only three presenting more than 20 cases (Table 5-1). This study represents the second largest RNA sequencing study from human FFPE tissue found in the literature search and increases the overall number of samples from 316 to 422.

19 (18%) FFPE samples clustered into the QC Fail group. Loss of one fifth of samples is substantial, and could create difficulties for research investigating large cohorts. This is an issue which may not have been detected in smaller studies and is therefore an important finding. Despite sample loss, the study resulted in the successful sequencing of over 80 FFPE libraries, demonstrating that a large-scale approach is feasible.

The library preparation and sequencing on a large scale was a significant undertaking. It was critical that these stages of the research were performed by a service provider with experience and expertise in this type of project.

5.4.2 How does FFPE data quality compare to FF data quality?

The gold standard for assessing the quality of FFPE RNA-seq is to compare with FF samples prepared in the same way. Previous attempts to do this have been limited by differing methods of library preparation, resulting in exaggerated differences between the two sample types. In one study, where FF libraries underwent the poly-A tail selection method of RNA enrichment, and the FFPE samples underwent ribodepletion, there were over 11,000 differentially expressed genes, many of which were related to non-coding sequences (Jovanović et al., 2017). In fact, there was only a single study in the literature review where library preparation had been the same for both FF and FFPE samples (Hedegaard et al. 2014). The number of differentially expressed genes ranged from 2000-7000, which is more consistent with the 1188 genes identified in this chapter. One advantage of the present approach, therefore, is that both FF and FFPE samples underwent whole transcriptome RNA-seq library preparation following ribodepletion.

RIN values were measured for all FF samples to ensure that they were above seven. RIN values were not measured for the FFPE samples as they are known to be low in this type of material. A recent study of nucleic acid extraction from 108 FFPE samples demonstrated that 99% had a RIN score below three (Yakovleva et

al., 2017). Therefore, knowledge of the exact values would not have changed the approach to sequencing.

The input requirements for the cDNA synthesis for the FFPE RNA-Seq was 600-1000 ng. This is in line with previous publications using whole transcriptome RNA-seq on FFPE (Esteve-Codina et al., 2017; French et al., 2017; Guo et al., 2016; Haile et al., 2017; Just et al., 2016; Li et al., 2014; Norton et al., 2013; Sinicropi et al., 2012; Vukmirovic et al., 2017). This was much higher than the requirement for FF tissue (100 ng). The nature of this higher requirement meant that 20% of samples were excluded on the basis of RNA yield. There is hope that this requirement may reduce in the future as biotechnology companies produce alternative library preparation kits and sequencing approaches. An example of this is Illumina's RNA access kit, which claims to only need 20 ng of input RNA from FFPE material (Illumina, USA). However, this kit is focussed on capturing regions of coding RNA, meaning that it cannot assess the whole transcriptome.

This study identified an association between the concentration of the input RNA and sequencing outcomes; samples with higher concentrations appeared to perform better when judged by read alignment statistics. Higher quality RNA may be more highly concentrated, or RNA at very low concentrations may be measured less accurately by the spectrophotometer. No association was found between measures of RNA purity (260/280 and 260/230 values) and alignment quality, despite many of the 260/230 values being suboptimal.

A discrepancy was identified in the number of reads generated by each sample type. Whilst both sequencing runs generated a median value of reads close to the target (50 million per sample), the FFPE cohort had a much wider range of values than the FF cohort. This was believed to be due to the more degraded nature of the RNA from the FFPE blocks, making it a more challenging material to sequence (personal communication, L Klitten, Exiqon). The increased variability in the FFPE cohort makes planning future experiments challenging, because it cannot be guaranteed that the target number of reads will be obtained for any specific sample.

The FF samples had marginally more reads removed through trimming than the FFPE samples (4.7% versus 3%). Whilst this was statistically significant, the magnitude was small and is likely to have a negligible effect on downstream processes. Only one study of whole transcriptome RNA-seq reported on the level

of read trimming in FFPE compared to FF samples, and found that in two of their three datasets there was more trimming in the FFPE reads, and in the third dataset there was more trimming in the FF reads (Hedegaard et al. 2014). It is probable that the level of read trimming is affected by the RNA extraction processes and the parameters of each sequencing run, and not necessarily directly by the nature of FF versus FFPE tissue. Despite the small difference in medians, the distribution between the two groups was very different. The FF samples had a small range of reads removed, whereas in the FFPE dataset, seven percent of samples lost over 10% of reads, again highlighting the greater variability of the FFPE cohort.

There was no difference in the levels of ribosomal reads between the two cohorts. Overall, 3-4% of reads were removed during the abundant sequences filtering, which is consistent with some studies (Hedegaard et al., 2014; Zhao et al., 2014) but higher than others (Haile et al., 2017; Jovanović et al., 2017; Li et al., 2014; Sinicropi et al., 2012) and is likely a product of the specific protocol used for ribodepletion. Both had some libraries with greater than 10% rRNA content and, interestingly, these were consecutively processed samples. It is likely that for these samples that the ribodepletion efficiency fell for technical reasons. Overall, ribodepletion seemed to be equally effective in both FF and FFPE tissue.

Whilst the FFPE cohort was associated with a significantly lower rate of alignment to the human genome than the FF cohort (83.9% versus 72.1%), the level of alignment was still consistent with previously published data (Haile et al. 2017; Jovanović et al. 2017; Esteve-Codina et al. 2017; Vukmirovic et al. 2017; Guo et al. 2016; Just et al. 2016; Li et al. 2014; Hedegaard et al. 2014; Morton et al. 2014; Norton et al. 2013; Xiao et al. 2013; Sinicropi et al. 2012). Some studies reported rates of alignment up to 95% (Esteve-Codina et al., 2017; Haile et al., 2017; Jovanović et al., 2017), however, these used small datasets. The only other large study generated an average read alignment of 69.45% (Sinicropi et al., 2012). This lower rate of alignment is consistent with the values described for the current study, leading to the conclusion that an alignment rate above 90% may not be realistic across a larger, more diverse sample set.

There was a strong negative correlation between reads aligning to the human and bacterial genomes (r=-0.95), suggesting that the limiting factor in human genome alignment was contamination with bacteria, rather than RNA degradation. This was emphasised by the finding that when all FFPE samples with

less than 10% bacterial reads were analysed, median human genome alignment was 78.3%; in keeping with levels seen in FF tissue. There is no evidence to suggest that this has been identified in other published datasets, even in those with low levels of read alignment, where this was generally attributed to poor RNA quality (Guo et al., 2016; Morton et al., 2014; Vukmirovic et al., 2017; Xiao et al., 2013).

FastQC analysis (Andrews, 2010) was performed to investigate the possible causes for differences between FF and FFPE samples. Surprisingly, few studies provided a detailed report of their FastQC or equivalent analysis. The only exceptions being three studies which provided information on GC content (Esteve-Codina et al., 2017; Graw et al., 2015; Guo et al., 2016). Unfortunately, all of these studies were small with a combined total of 14 samples. The key differences between FF and FFPE samples were found to be the Phred quality scores, GC content, 'N' content, duplication levels and levels of adapter read-through.

The distribution of Phred scores was different between the two cohorts, with FFPE samples exhibiting some Phred scores under 15. However, both cohorts demonstrated acceptable scores overall, confirming that the technical quality of the sequencing was good.

When assessing GC distribution, both cohorts had a smooth peak at 35-42%. This was consistent with the GC content of the human genome (39.3%), although below that of protein coding sequences (48.9%) (Guo et al., 2016). The inclusion of numerous non-coding sequences in whole transcriptome RNA-seq may explain why the values are more consistent with the human genome content rather than just the protein coding sequences. It is highly probable that the 35-42% peak represented human RNA.

The two cohorts had differing additional GC peaks. The FFPE cohort demonstrated a further 'spike' at 54%. Interestingly, Graw and colleagues also reported a similar peak at 56% in their FFPE dataset of six samples. They attributed this to increased levels of intronic sequences (Graw et al., 2015). However, in the present study this peak was strongly associated with those samples that had high alignment to the bacterial genome. This GC peak is therefore more likely to represent bacterial, rather than intronic, sequences. The FF samples demonstrated several 'shoulders' at 47-63% which were associated with rRNA

contamination. These samples had been randomised to the library preparation process consecutively, suggesting a partial failure of the ribodepletion step. This was therefore not specific to sample type.

This study used a read length of 100 bases, which was longer than most reported research (Table 5-1). As FFPE material is known to be more degraded, it was hypothesised that the fragments being sequenced would be shorter than the FF equivalents. This was supported by a significantly shorter insert length, increased 'N' content and increased adapter read-through towards the end of each read in the FFPE cohort. With shorter sequences, by the time the sequencer reaches higher base positons in the read, it is reading either into the adapter sequence or beyond the end of the fragment. This results in either high levels of adapter content or insertion of 'Ns' in the output. This is unlikely to have any impact on downstream analysis, as adapter and 'N' sequences are removed by trimming. Only two previous FFPE studies used a read length beyond 75 bases (Hedegaard et al., 2014; Morton et al., 2014) but neither reported on the adapter or 'N' content at higher base positions. It may be better to use a shorter read length to prevent this 'wasted' sequencing.

Consistent with a previous study (Esteve-Codina et al., 2017), FFPE demonstrated less library complexity than FF samples. Samples with the highest duplication levels in the FFPE cohort were also those that had very high levels of bacterial reads. This reduction in complexity therefore resulted from repeated sequencing of high levels of bacterial rRNA. These reads were removed during the filtering steps, but again represented 'wasted' sequencing.

To quantify gene expression, only reads mapped to exons were counted. There was a large difference between the FFPE and FF cohorts, with exon mapping of 23.2% and 43.7% respectively. The numbers for both datasets were lower than for poly-A enriched libraries, as ribodepletion does not remove unspliced and nascent RNAs, which may subsequently align to introns. It has been reported that brain tissue is enriched for nascent RNAs and it would therefore be anticipated that a greater proportion of reads align to introns, irrespective of library preparation methods (Ameur et al., 2011). The fact that the FFPE samples had even lower exonic fragments than FF supports the findings of previous FFPE, non-brain, studies where exonic fragments have been reported to be as low as 10-13% of total mapped reads (Hedegaard et al., 2014; Sinicropi et al., 2012). The reasons for this are unclear. One way to overcome this problem would be to

increase the number of reads sequenced, to counter this large loss of reads at the exon mapping step, thus increasing the absolute number of reads aligning to exons.

FFPE samples had significantly fewer genes identified than FF. This is likely due to the fact that there were fewer reads, and a lesser proportion of reads mapping to exons in the FFPE cohort. No evidence was found to suggest that this was due to a lesser quality of FFPE RNA. The number of genes identified per FFPE sample was consistent with that reported by some FFPE studies (Haile et al., 2017; Morton et al., 2014) but much higher than others (Li et al., 2014; Sinicropi et al., 2012; Vukmirovic et al., 2017). However, it was evident in the FFPE cohort that the variation was high (1651 genes for the worst performing sample to 31270 for the best) and this must be taken into consideration when assessing the quality of an FFPE RNA sequencing run.

For both FFPE and FF cohorts the level of correlation between biological replicates was high, even prior to normalisation, with median correlation coefficients of 0.85 and 0.88 respectively. Differences between FFPE and FF cohorts were apparent in the variability of the FFPE; following normalisation, the range of coefficients for the FFPE cohort was 0.57-1.00 compared to 0.91-1.00 in the FF cohort. This again emphasises the need for removal of poor quality samples.

The level of correlation between paired FFPE and FF technical replicates was high; the mean correlation coefficient was 0.85, increasing to 0.98 with normalisation. This suggests that despite some technical differences between the two cohorts, the FFPE and FF sequencing results were highly similar for matched pairs. This is encouraging in terms of the technical reliability of the FFPE RNA-seq. It would be surprising if technical replicates reached 100% similarity, given that samples would have been obtained from different areas of the same tumour. Little is known about the level of intra-tumour heterogeneity in ependymoma, but at the microscopic level there is variation across tumour samples.

5.4.2.1 Challenges specific to FFPE samples

This study was unique in obtaining FFPE tissue samples from a mixture of scrolls and cores. This provided an opportunity to compare the impact of these approaches on alignment to the human genome. The difference in tissue sampling appeared to be the largest factor in determining the level of bacterial contamination of the samples, with cores demonstrating a significantly higher median number of reads aligning to the human genome compared with scrolls

(77% versus 42%, $p<0.001$). A search of the literature was unable to find any other study that had attempted to investigate this question. Therefore, this is both a critical and novel finding, leading to the recommendation that samples sequenced from FFPE should be taken from cores where possible. Should use of scrolls be unavoidable, they should not be taken from the block surface.

Based on the theory that older blocks would contain more degraded RNA, it was hypothesised that the age of the block may have an impact on sequencing outcomes. The findings of this study were not consistent with this hypothesis. In fact, some of the best performing runs actually resulted from the oldest blocks. The oldest sample in the analysis had been in storage for almost 28 years, yet still had a human genome alignment rate of 77.8% and identified 26864 genes. Blocks in storage for more than 20 years did not have a significantly different alignment rate to those in storage for less than 20 years ($p=0.79$). This finding contradicts the work of other authors who have suggested that older blocks result in poorer sequencing runs (Hester et al., 2016; Jovanović et al., 2017).

In comparison to previous publications, this study investigated older blocks (eight blocks in excess of 20 years, median 11.7 years). The age of the blocks was unreported in five, less than 10 years in nine and 10-20 years in two studies. One study sequenced a single sample of 94 years storage time, however no other human study included samples exceeding 20 years in age (Xiao et al., 2013). Consequently, a possible explanation for the lack of age related differences in this research, is that nucleic acid deterioration occurs early on in the storage process and, therefore, samples beyond a certain age do not show measurable differences in quality. Given that the effect of storage time on nucleic acids has been the subject of some debate it was surprising that 6 of the 19 relevant investigations did not clearly report the age of the blocks used (Table 5-1).

5.4.3 Is the data of adequate quality to use in answering other research questions?

The finding that RNA sequencing of FFPE is technically possible on a large scale is positive for researchers who wish to perform molecular profiling on historical cohorts of patients with rare disease. However, despite demonstrating the technical feasibility of this approach, it still needed to be shown that the data was adequate to analyse a clinical cohort. One of the challenges was that a lack of research in this area has meant that there are no clear quality control parameters guiding which samples should be removed from downstream analysis. Therefore, the FFPE was directly compared with the FF cohort and previously published work

to check its performance in a clinical analysis.

Unsupervised hierarchical clustering of the FFPE cohort generated distinct subgroups relating to tumour location, which were supported by DNA methylation class predictions. Crucially, clustering identified an additional group of poor quality samples, with very low human genomic reads as a result of enrichment of bacterial sequences. It was of interest that this poor-quality group was also significantly less likely to produce DNA methylation class predictions, suggesting there may be global problems with the samples, preventing molecular profiling.

In both cohorts, the gene expression and ontology analyses of the two PF groups (PF1 and PF2) were highly consistent with previously described EPN_PFA and EPN_PFB tumours (Pajtler et al. 2015, Hoffman et al. 2014a; Pajtler et al. 2015; Witt et al. 2011). Interestingly, most of the DNA methylation class predictions in both PF1 and PF2 groups were for EPN_PFA. In fact, only two EPN_PFB specimens were identified across both datasets. Consequently, the PF clustering was suggestive of two EPN_PFA molecular subgroups, supporting the analysis of the DNA methylation profiles in Chapter 4. This data indicated that EPN_PFA and EPN_PFB expression patterns are not always suggestive of EPN_PFA and EPN_PFB tumours, but may be seen in subtypes of EPN_PFA. This needs to be corroborated by further research, but has the implication that using gene expression profiling alone may result in misleading interpretations of ependymoma clustering.

In both cohorts, the PF2 group consisted of tumours with a number of DNA methylation profiles, the majority of which were EPN_PFA. However, the PF1 group consisted almost exclusively of DNA methylation defined EPN_PFA tumours. This suggested that one subgroup of EPN_PFA (PF1) has very distinct gene expression patterns, whilst the other (PF2) may have patterns similar to other ependymoma types. This may also have arisen as a result of the relatively low numbers of non-EPN_PFA tumours being unable to form a separate cluster. It also explains why a number of ST tumours, particularly EPN_YAP, clustered into the PF2 group.

In view of the ability of the FFPE RNA-seq to reflect previously described clustering, gene expression and ontology patterns, and to correlate strongly with the patterns in the FF cohort, it was decided that it was appropriate to continue to use these results in the analysis of matched primary and recurrent ependymoma in the remainder of the thesis.

5.4.4 What recommendations can be given to future researchers?

The recommendations that should be applied to future FFPE RNA sequencing studies are:

(1) Minimising bacterial contamination:

a. Specimens should be taken from cores of blocks rather than scrolls. If scrolls are necessary, they should not be from the surface of the block but the first few sections should be discarded;
b. Consideration should be given to performing PCR to identify bacterial ribosomal sequences in the extracted RNA prior to sequencing. Removal of contaminated samples at this stage may minimise costs of wasted sequencing.

(2) Dealing with shorter RNA fragments:

a. RIN measurements may not be helpful for FFPE RNA-seq in view of known high levels of degradation. However, it is unnecessary to sequence with long read lengths due to the high number of uninformative calls towards the end of reads. 25-50 base pairs may be more appropriate than 100 but further research is needed to confirm the exact number.

(3) Number of reads:

a. A higher read target (in excess of 50 million) should be generated for each sample in view of the variability of FFPE samples. This would ensure a higher minimum number of reads and increase the absolute number of reads mapping to exons.

(4) Identifying poorly performing samples:

a. An unsupervised clustering approach to quality control may be appropriate for investigators using large sample sets. For those with smaller numbers further work is needed to identify quality control cut-offs for excluding samples. In this study samples with human genome aligning reads of <20% clustered into the QC fail group.

(5) Planning for sample loss:

a. Up to 20% of samples may be lost due to quality inadequacies. This needs to be taken into account when planning projects;
b. On the basis of this study, sample age should not be a barrier to inclusion.

5.4.5 Summary

The decision to undertake RNA-seq of FFPE samples was made in an attempt to improve the power of the primary and recurrent ependymoma study, and represented a novel approach to addressing this problem.

Few studies have published FFPE RNA-seq data and of those that have, only two included more than 25 samples, making the understanding of the utility of this approach in larger datasets relatively unknown. Additionally, no other studies have yet been conducted using paediatric brain tumour samples.

This chapter has demonstrated that NGS of RNA from large sets of archival tissue specimens is feasible and can produce results which replicate known patterns. Challenges of FFPE RNA sequencing were: obtaining adequate RNA yields; increased variability in data; shorter fragment lengths; and contamination with bacterial reads. These problems could be overcome in the future with the potential of new RNA extraction kits, modifying read parameters, and clustering approaches to identify poorly performing samples.

Data from the FFPE cohort reflected the findings of the FF cohort and previously published studies with regards to molecular subgrouping. It has also provided supporting evidence for the presence of more than one EPN_PFA molecular subgroup, which has previously been proposed by an international collaboration which included DNA methylation profiles from this study.

The FFPE samples that performed well in this analysis were taken forward to the analysis of matched primary and recurrent ependymoma described in the remainder of the thesis.

6 RNA Sequencing of Matched Primary and Recurrent Ependymoma Pairs

6.1 Introduction

Despite much research into the underlying biology of primary paediatric ependymoma, with the exception of two studies, little is known about tumour biology at recurrence (Peyre et al., 2010; Hoffman et al., 2014a). Given that there are no specific treatments for recurrence after radiotherapy (Peyre et al., 2010), a better understanding of the biological behaviour of these tumours is warranted to provide further insights into possible therapeutic interventions. By using paired samples, a clearer picture can be obtained of changes at recurrence by minimising the impact of inter-individual variation on the study design (Peyre et al., 2010).

Previous authors (Hoffman et al., 2014a; Peyre et al., 2010) have investigated matched primary and recurrent ependymomas, but neither used RNA sequencing. Using gene expression arrays, Hoffman and colleagues identified immune system changes as being important in the posterior fossa molecular subgroups; whilst Peyre identified increases in kinetochore proteins alongside downregulation of metallothioneins at recurrence (Chapter 1.4.4). Peyre also identified the downregulation of a number of immune related genes. On the basis of these overlapping findings, further consideration of the immune system in recurrent ependymoma was warranted.

Recent research on cancer and the immune response has defined ways to investigate immune activity in tumours, based on gene expression data from whole tumour samples.

A report by a group at the University of Innsbruck described the creation of the Cancer Immunome Atlas (TCIA), based on patient samples derived from the Cancer Genome Atlas (TCGA) (Charoentong et al., 2017). This report introduced the concept of the immunophenoscore, based on RNA sequencing of whole tissue samples from tumour specimens. The immunophenoscore is a marker of tumour immunogenicity and was developed from a panel of immune related genes, divided into subcategories, based on function:

(1) Effector cells;

(2) Immunosuppressive cells;
(3) Major histocompatibility (MHC) molecules;
(4) Immunomodulators.

Increased cancer mutational load was associated with a higher immunophenoscore (Charoentong et al., 2017). The immunophenoscore was also found to be a predictor of response to immune checkpoint blockade in patients with melanoma.

Rooney and colleagues investigated properties associated with local cytolytic (CYT) activity (Rooney et al., 2015). CYT is a measure of T-cell effector activity is considered important because it is the final step in tumour cell killing. CYT was scored based on expression levels of Perforin (*PRF1*) and Granzyme A (*GZMA*), which assessed the level of CD8+ T-cell effector activity within tumour. (Rooney et al., 2015).

Another suggested mechanism by which tumours can up- and downregulate their immune profile is the expression or suppression of CTAs (Almeida et al., 2009). These are genes with restricted expression in normal tissue, being expressed only in the human germ line and numerous cancers (Simpson et al., 2005). They are immunogenic and have been identified as potential therapeutic targets for cancer vaccines; phase I and II trials have been conducted (Krishnadas et al., 2015). Whilst there are some reports of studies of specific CTAs in mixed cohorts of glioma (Grizzi et al., 2006; He et al., 2014; Li et al., 2017) there was no evidence of published data on CTA expression in cohorts of recurrent paediatric ependymoma.

Chapter Aims

This chapter aimed to:

- Compare gene expression at primary diagnosis compared to matched recurrence across FF and FFPE datasets based on:
    - All tumours combined;
    - Location;
    - Molecular designation.
- Consider the impact of therapy on differential expression at recurrence;
- Assess the role of the immune system in primary and recurrent ependymoma using predefined scoring systems (immunophenoscore and cytolytic activity score) plus consideration of CTAs.

6.2 Materials and methods

The RNA-seq data of FFPE and FF cohorts validated in Chapter 5 were used for the analysis of matched primary and recurrent ependymoma pairs.

6.2.1 Differential expression analysis

Raw gene expression data was analysed using EdgeR for differential expression analyses as described in Chapter 5.2.2. All analyses in this chapter were paired tests.

6.2.2 The hypergeometric test

As described in Chapter 5.2.4, the hypergeometric test was performed to compare the significance of the overlap of differentially expressed genes, between FF and FFPE cohorts. It was based on the number of matching genes, with an unadjusted p-value of <0.05, in each cohort.

6.2.3 RNA-seq meta-analysis

RNA-seq meta-analysis was performed using the R package 'metaRNASeq' (Rau et al., 2014). This generated lists of shared differentially expressed genes across FFPE and FF datasets, and created 'signature sets' of differentially expressed genes for recurrence. Not only did this approach allow p-values to be combined, using Fisher's combined probability test (Fisher, 1925; Fisher and Mosteller, 1948), but it also allowed the application of an FDR correction to resultant gene sets.

6.2.4 Gene enrichment analysis

Details of the methodology for gene enrichment analyses are provided in section 5.2.5. As the GOrilla analyses tended to produce very long lists of terms, 'Reduce and Visualise Gene ontology' (REVIGO) (Supek et al., 2011) was used in an attempt to develop a better understanding of these lists. This software removed redundant terms from long lists of ontologies and then displayed the remaining terms in semantic scatterplots based on their similarity to each other.

Enriched gene lists were entered into the REVIGO web based platform (revigo.irb.hr) after extracting them from GOrilla (Eden et al., 2009), and scatterplots were modified in the R statistical environment for visualisation.

6.2.5 Normalisation and transformation of counted reads

Transcript per million (TPM) normalised data was used for immunophenoscore, CYT and CTA analyses as they were unable to manage raw data (Charoentong et al., 2017; Rooney et al., 2015). Samples were TPM normalised by correcting for

the number of reads (number of counts per million mapped reads) and gene length (number of counts per kilobase). TPM was calculated within the R statistical environment using the GenomicFeatures and R base packages (Lawrence et al., 2013; R Core Team, 2014) (Appendix 3). TPM values were then transformed by a log2 conversion to account for the high variance of lowly sampled genes (Love et al., 2014).

6.2.6 Generation of ependymoma immunophenoscores

Immunophenoscores were calculated from log2 normalised TPM values. The results were generated in the R statistical environment using the published immunophenoscore script (Charoentong et. al., 2017). Results were also generated for the four individual components of the immunophenoscore: antigen presentation (MHC); effector immune cells (EC); suppressor immune cells (SC); and checkpoint inhibitors or stimulators (CP).

6.2.7 Calculation of immune cytolytic activity (CYT)

CYT was measured by calculating the geometric mean of the TPM for Perforin (*PRF*1) and Granzyme A (*GZMA*), for each sample, as outlined in a previous study (Rooney et al., 2015).

The formula used was:

$$\text{Immune Cytolic Activity} = \sqrt{(PRF1 * GZMA)}$$

6.2.8 Cancer-Testis antigen analysis

A list of verified CTAs was obtained from the CTDatabase (cta.Incc.br). CTAs with low expression across both FFPE and FF datasets (TPM less than one) were removed. The remaining CTAs were compared to the differentially expressed gene lists for primary and recurrent samples to identify CTAs that changed at recurrence.

6.2.9 Cell culture and cell lines

BJ hTERT + SV40 Large T+ (BJLE) (Weinberg et al., 1999) were used in this study, which were kindly provided by Professor Robert Weinberg (Professor for Cancer Research at the Massachusetts Institute of Technology).

Cells were cultured in a medium consisting of Knockout DMEM (Thermo Fisher Scientific, USA), 14.5% Fetal Bovine Serum (Thermo Fisher Scientific, USA),

16.5% Medium 199 (Thermo Fisher Scientific, USA), 1.76 mM L-Glutamine (Sigma, UK) and 0.88% penicillin-streptomycin (Sigma, UK).

To recover cells, cryovials were thawed quickly, in a 37°C water bath, before being transferred to a T25 flask (Eppendorf AG, Hamburg, Germany, Catalogue No. 0030710118) in 5 ml of culture medium. Flasks were placed in a standard 5% CO_2-air incubator (Panasonic, UK) at 37ºC. After 24 hours, the flasks were inspected under a microscope (Olympus, UK) to ensure that cells had adhered, before the culture medium was replaced.

Cells were grown in T75 flasks (Eppendorf AG, Hamburg, Germany, Catalogue No. 0030711122) and passaged when they reached approximately 70% confluence (48-72 hours). At each passage, cells were washed with 10 ml Hank's Balanced Salt Solution (HBSS) (Thermo Fisher Scientific, UK) before incubation at 37ºC with 2ml of 1X Trypsin-EDTA (Sigma, UK) for five minutes, until the cells had detached. 10 ml of media were added to inhibit the trypsin and the resultant solution was split into new flasks in a ratio of 1:2 to 1:5. The flasks were then placed in the incubator for cells to adhere and grow.

6.2.10 cDNA synthesis for qPCR

For the purposes of the validation experiments, all of the cDNA was synthesised simultaneously to minimise the impact of varying conversion efficiencies. It was not possible to use the same cDNA created for the sequencing libraries.

RNA was treated with DNase to reduce genomic contamination. For each sample, 1 μg of RNA was incubated with 2 μl of DNase (Promega, USA) and 2 μl of DNase buffer (Promega, USA), then made up to a total volume of 20 μl with double distilled water. The mixture was incubated at 37°C for 30 minutes before the addition of 2 μl stop buffer and incubation at 65°C for 15 minutes.

Next, half of the RNA was transferred to a second vial. One vial was labelled RT (reverse transcriptase) and the other NRT (no reverse transcriptase). Two master mixes were created (RT mastermix and NRT mastermix) (Table 6-1). 10 μl of RT and NRT master mixes were added to each RT and NRT vial respectively. Samples were then incubated on a thermal cycler (Techne, UK) for ten minutes at 25°C, one hour at 42°C and five minutes at 70°C. 40 μl of nuclease free water was added to each sample before storage at -20ºC.

Reagent	Amount per sample	Manufacturer
5x Revertaid buffer	4 μl	Thermo Scientific, UK
RNAse inhibitor	0.5 μl	
dNTP mix	2 μl	
Oligo dTs	1 μl	Eurofins genomics, Germany
Random primers	50 ng	Thermo Scientific, UK
Double distilled water	1.5 μl	-
Reverse transcriptase	1 μl*	Thermo Scientific, UK

Table 6-1: Table of reagents for the RT and NRT mastermixes. *For the NRT mastermix the reverse transcriptase was replaced with an additional one microlitre of water.

6.2.11 qPCR primer design

Transcripts for selected genes were identified using Ensembl (http://www.ensembl.org). Primers were designed based on these transcripts using the NCBI Primer Blast software (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primer Blast settings were modified to: identify PCR products of 50-200 base pairs; include introns (to prevent amplification of genomic DNA); and allow amplification of splice variants.

Primer Blast results were analysed to identify primers of 20-25 base pairs with balanced GC composition, and no guanine at the five-prime end. The selected primers were pasted into NCBI Nucleotide BLAST (Basic Local Alignment Search Tool) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and compared to human sequences to ensure specificity to the gene of interest.

Primers were supplied in lypophilised format by Eurofins Genomics (Ebersberg, Germany) and diluted to a concentration of 100 pmol/μl with nuclease free water to make a stock solution.

The genes of interest were C-X-C motif chemokine ligand 12 (*CXCL12*), 2'-5' Oligoadenylate synthetase 1 (*OAS1*), 2'-5' Oligoadenylate synthetase 2 (*OAS2*), 2'-5' Oligoadenylate synthetase 3 (*OAS3*), Bone marrow stromal cell antigen 2 (*BST2*) and Radical S-adenosyl methionine domain contain 2 (*RSAD2*). Primers were also generated for glyceraldehyde-3-phopshate dehydrogenase (*GAPDH*) as a normalising gene (Table 6-2).

Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')
OAS1	CGAGGTAGCTCCTACCCTGT	TCTCCCCGGCGATTTAACTG
OAS2	CAGGAACCCGAACAGTTCCC	GGACAAGGGTACCATCGGAG
OAS3	CGCGGAAGGAGTTCGTAGAG	AAGCAGTCGAGGAAGATGACAA
BST2	GCGTCCTGAAGCTTATGGTTT	TTCAGGATGTGGAGGCCCA
RSAD2	GCAAAGTAGAGTTGCGGCTG	CCACGGCCAATAAGGACATTG
CXCL12	TGAGCTACAGATGCCCATGC	CTTCAGCCGGGCTACAATCT
MX1	ATCTGGAGTGAAGAACGCCG	CAGCTGGATCAGCTTTTGCG
GAPDH	TCTGGCCCCCTCTGCTGATGC	GGTGGCAGTGATGGCATGGAC

Table 6-2: Summary of primer sequences used for the qPCR validation of the RNA sequencing data.

6.2.12 Polymerase chain reactions (PCR)

PCR involves the amplification of few copies of cDNA to a detectable level, through the generation of a complementary strand to a DNA template. The reaction consists of three main steps:

(1) Denaturation: Breaking apart the complementary cDNA strands;

(2) Annealing: The attachment of forward and reverse primers to denatured cDNA;

(3) Extension: Synthesis of the remainder of the complementary strand by extending the primer sequences under the activity of DNA polymerases.

The steps are repeated 40 times with a doubling of the amount of cDNA each time.

The contents of each reaction are outlined in Table 6-3. For each PCR experiment, a mastermix was made up by calculating the volume for the number of reactions required and adding 10%. 24 μl of mastermix was then added to each well of an unskirted 96-well plate (Bio-Rad, UK) before the addition of either cDNA template, no template control or negative control (RNAse free water).

Reagent	Amount per well (μl)	Manufacturer
SYBR Green	12.5	Bio-Rad, UK
Double distilled water	2.5	-
Forward Primer	2.5	Eurofins Genomics, Germany
Reverse Primer	6.5	

Table 6-3: Contents of each PCR reaction.

PCR reactions were performed on a C1000 thermal cycler within the CFX96 Real-Time System (Bio-Rad, UK). The thermal cycler protocol involved:

- Initial activation: five minutes at 95°C;

- 39 cycles of:
    - Denaturation: 30 seconds at 95°C;
    - Annealing: one minute at the temperature determined during primer optimisation;
    - Extension: one minute at 72°C.
- Final extension: ten minutes at 72°C.

For each reaction, melt curves were generated to ensure that only one product was amplified. A read of the fluorescence level was taken after each denaturation, annealing, and extension cycle; to determine when the level of fluorescence rose above background (the C(t) value). C(t) values were used for calculation of primer efficiencies and relative expression levels.

6.2.13 Primer optimisation and efficiency calculations

Primers were optimised by temperature gradient end-point PCR and electrophoresis, on a 2% (w/v) agarose gel (50 ml TAE, 1 g agar, 1 μl ethidium bromide). The brightest bands identified the optimal temperatures. For all primers 58°C was adequate.

Primer efficiencies were calculated using the BJLE cells as input cDNA. A serial 1:2 dilution was made from the neat cDNA and five different known concentrations were run on the same 96-well plate. Each reaction was performed in triplicate and C(t) values were taken as the mean. The Bio-Rad CFX software was used to calculate the primer efficiencies using a regression equation based on the mean C(t) values on the y-axis and the log concentration of the input cDNA concentration on the x-axis:

$$y = mx + c$$

The value of m was the gradient of the slope and was used to calculate the efficiency of the primer pair with a standardised equation (Pfaffl, 2001):

$$E = (10^{(-1/\text{Gradient})})$$
$$\text{Efficiency} = E - 1$$

Target efficiencies were 90-110% for each primer pair. All but one primer pairs (*MX1, efficiency 154%*) achieved this.

6.2.14 qPCR comparison of genes in primary and matched recurrences

One 96 well PCR plate was used per gene validation. On each the BJLE cell line was used as the control sample and four primary and matched recurrent tumours were investigated. Each sample was run in triplicate and mean C(t) values were calculated. Negative and NRT controls were also run for each sample.

Relative gene expression was calculated based on the Pfaffl equation (Pfaffl, 2001):

$$\text{Relative Ratio} = E_{target}^{\Delta C(t)(control-sample)} / E_{reference}^{\Delta C(t)(control-sample)}$$

Where E was the efficiency of the primer pair, control was the mean C(t) value of the BJLE cells, sample was the mean C(t) value of the ependymoma specimen in question, target was the gene of interest and reference was the reference gene (*GAPDH*).

The relative ratio was the expression level of the gene of interest relative to the reference gene, and was used to compare samples to calculate the relative fold change of expressed transcripts. Fold change at recurrence was calculated by:

$$\text{Relative Ratio}_{\text{ Recurrent Sample}} \text{ / Relative Ratio}_{\text{ Primary Sample}}$$

The fold changes were then compared with those produced by RNA-seq to confirm that the genes had a similar direction of fold change at recurrence compared to primary disease.

6.3 Results

Descriptions of the overall FFPE and FF cohorts, as compared to the clinical recurrent cohort, are found in Chapter 5, Table 5-2 and Table 5-3.

For each location and molecular subgroup based comparison between primary and first recurrence, Table 6-4 and Table 6-5 outline the number of samples included, differentially expressed genes and enriched ontologies. Unless otherwise specified, the results quoted for ontology analyses were derived from GOrilla.

6.3.1 Recurrence across all tumour types

In order to obtain a global overview of the differences between primary and first recurrence across the entire dataset, a differential expression analysis was

conducted. This was irrespective of tumour location, molecular subtype or therapeutic intervention.

The FFPE and FF cohorts included 25 and 27 matched pairs respectively. The two sets were largely independent with overlap between two pairs (25A and B, 26A and B). After filtering, 14315 genes were tested in both cohorts.

In the FFPE cohort 652 genes were differentially expressed at first recurrence at $p<0.05$. None met the FDR<0.05 threshold. In the FF cohort 794 genes were differentially expressed at first recurrence with $p<0.05$. Two genes remained after FDR correction (*RASL10A* and *ASIC4*). Across both cohorts 65 overlapping genes were differentially expressed, in the same direction of fold change. This was highly statistically significant ($p<0.001$, hypergeometric test).

When combining the two cohorts in RNA-seq meta-analysis, three genes were downregulated at first recurrence with FDR<0.05. These were Elastin microfibril interfacer 3 (*EMILIN3*) (Fold change -3.4, FDR<0.001), Acid sensing ion channel subunit family member 4 (*ASIC4*) (Fold change -2.7, FDR=0.001) and Lipoma HMGIC fusion partner-like 3 (*LHFPL3*) (Fold change -3.8, FDR=0.025).

GOrilla was used to assess whether there was enrichment of comparable gene ontologies between FFPE and FF cohorts (Eden et al., 2009).

Across both cohorts, 22 overlapping terms were upregulated at first recurrence ($p<0.001$, hypergeometric test) (Appendix 5). After removal of redundant terms with REVIGO (Supek et al., 2011), the remaining terms were related to multicellular organismal catabolism, oxygen transport, regulation of response to external stimulus, extracellular matrix organisation, regulation of lymphocyte proliferation and the immune response (Figure 6-1).

Across both cohorts, 122 overlapping terms were downregulated at first recurrence ($p<0.001$, hypergeometric) (Appendix 5). Terms relating to biological adhesion, cell-cell adhesion, cell communication and signalling pathways were identified. No immune related ontologies were downregulated (Figure 6-2).

Comparison		Pairs included		Genes tested	Genes with p-value <0.05		Genes with FDR<0.05		Overlapping genes (p<0.05)	Hypergeometric p-value (FF vs FFPE)	No. Genes in meta-analysis
		FFPE	FF		FFPE	FF	FFPE	FF			
All cases		25	27	14315	652	794	0	2	65	<0.001	3
Location	ST	8	-	15330	560	-	0	-	-	-	-
	PF	16	20	14485	475	1217	3	5	40	0.452	8
Molecular groups	EPN_RELA	5	-	13742	411	-	0	-	-	-	-
	PF1	9	13	14108	569	1759	3	323	102	<0.001	113
	PF2	3	11	14289	499	1219	39	22	55	0.029	47

Table 6-4: Summary of differential expression analyses of primary compared to first recurrent tumours for location and molecular subgroups. Genes tested include number of genes tested after filtering to remove lowly expressed genes. Hypergeometric tests were performed comparing FF and FFPE datasets to establish whether the number of significant genes appearing in both cohorts are likely to have occurred by chance ($p>0.05$) or represent significant overlap between the two cohorts ($p<0.05$).

Comparison		Downregulated terms at first recurrence with FDR<0.05			Hypergeometric p-value (FF vs FFPE)	Upregulated terms at first recurrence with FDR<0.05			Hypergeometric p-value (FF vs FFPE)
		FFPE	FF	Overlap		FFPE	FF	Overlap	
All cases		171	327	122	<0.001	32	456	22	<0.001
Location	ST	30	-	-	-	459	-	-	-
	PF	193	168	9	<0.001	191	529	0	1.000
Molecular groups	EPN_RELA	0	-	-	-	178	-	-	-
	PF1	348	496	179	<0.001	6	138	1	0.056
	PF2	306	119	2	0.723	195	684	51	<0.001

Table 6-5: Summary of GOrilla gene set enrichment analysis of primary compared to first recurrent tumours for location and molecular subgroups. 14412 biological process gene ontology terms were tested for each comparison. Hypergeometric tests were performed comparing FF and FFPE datasets to establish whether the number of significant ontologies appearing in both cohorts are likely to have occurred by chance ($p>0.05$) or represent significant overlap between the two cohorts ($p<0.05$).

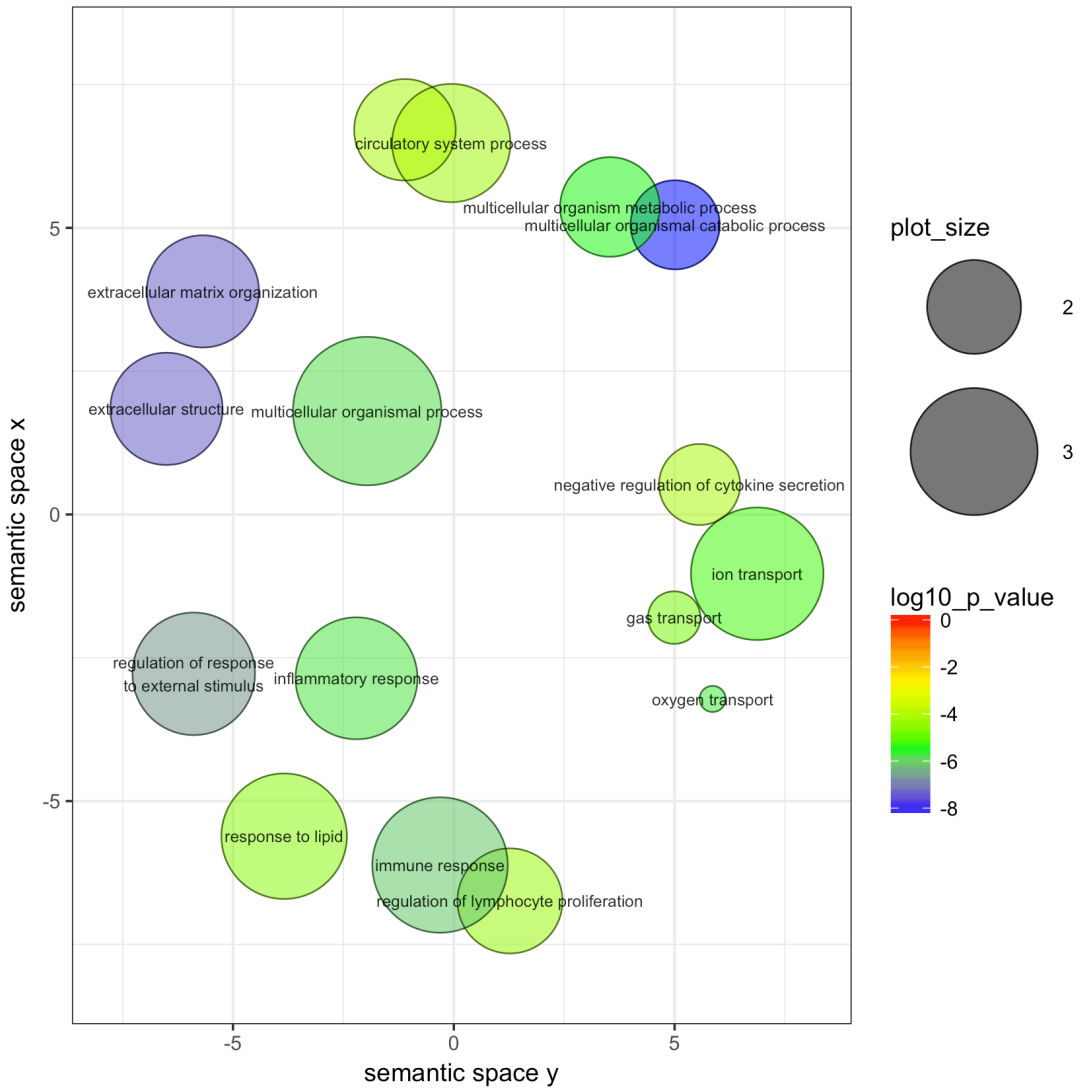


Figure 6-1: A scatterplot of non-redundant significantly upregulated gene ontology terms at first recurrence across the entire primary and recurrence dataset. The colour of the circles represents the p-value and the size of the circle represents the specificity of the term (more general terms appear as larger circles).

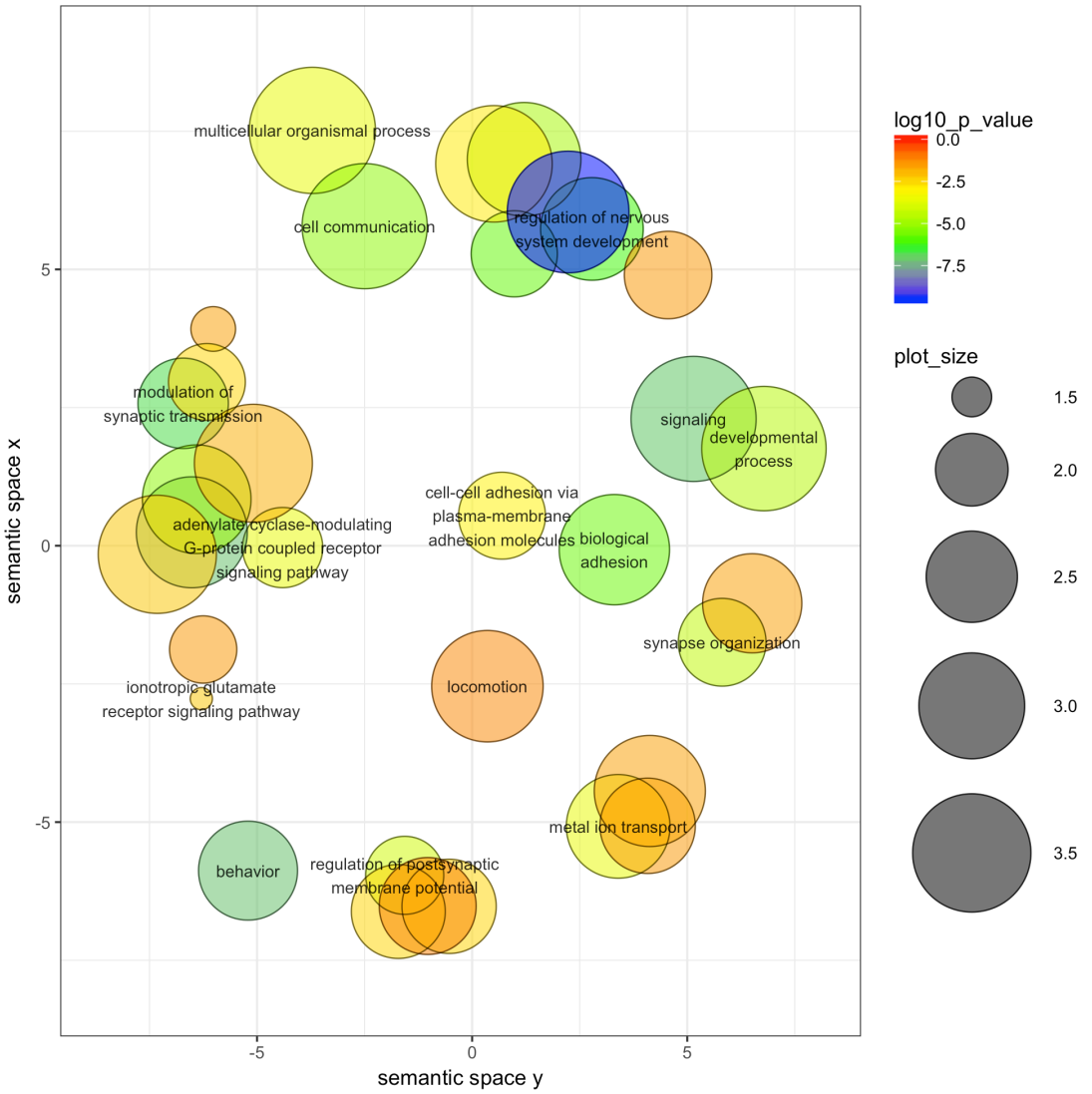


Figure 6-2: A scatterplot of none-redundant significantly downregulated gene ontology terms at first recurrence across the entire primary and recurrence dataset. The colour of the circles represents the p value and the size of the circle represents the specificity of the term (more general terms appear as larger circles).

When considering upregulated ontology terms that did not overlap between the two cohorts, the FFPE cohort generated fewer terms than the FF cohort. Of the 10 gene ontology terms exclusively upregulated at recurrence in the FFPE dataset, three (30%) were related to the immune response (GO:0002250 – Adaptive immune response, GO:0046641 – Positive regulation of alpha-beta T-cell proliferation, GO:0050701 – Interleukin-1 secretion). Of the 434 terms exclusively upregulated at recurrence in the FF dataset, 145 (33%) were related to the immune response.

When considering downregulated ontology terms that did not overlap between the two cohorts, the FFPE cohort generated fewer terms than the FF. Of the 49 terms exclusively downregulated in the FFPE dataset, none were related to the immune response. Of the 205 terms exclusively downregulated at recurrence in the FF dataset, only one was related to the immune response (GO:0009611 – Response to wounding).

6.3.2 Recurrence in tumours clustering in supratentorial groups

There were insufficient numbers of primary and matched recurrences clustering into the FF ST group for analysis. Therefore, ST location and DNA methylation subgroup analyses were based on the FFPE cohort. All eight matched primary and first recurrent ST ependymomas were included; five of which were in the EPN_RELA DNA methylation subgroup, one in the EPN_YAP subgroup and two with HGNET class predictions.

After filtering, 15330 genes were tested. 560 genes were differentially expressed at first recurrence with $p<0.05$ and none with FDR<0.05.

GOrilla analysis identified 459 terms upregulated at first recurrence, mainly related to the innate and adaptive immune responses and the production of cytokines (Table 6-6) (Appendix 5). REVIGO further classified these terms, of which three subheadings were particularly prominent; 'Regulation of cytokine production' encompassed 64 terms, 'immune response' encompassed 63 and 'response to external biotic stimulus' encompassed 20 terms (Figure 6-3).

Under the 'regulation of cytokine production' subheading, the interleukins were heavily represented, with ontologies identified for regulation of interleukin 1, 2, 6, 8, 10 and 12 and positive regulation of interleukin 1, 2, 6 and 12. The GSEA analysis supported these findings, identifying positive regulation of interleukin 4 production as a significant term (Figure 6-4).

Under the 'immune response' subheading there were terms related to T-cell and B-cell activation and taxis. Interestingly, there were two terms related to NF-kB signalling; GO:0051092 – 'Positive regulation of NF-kappaB transcription factor activity' and GO:0042346 – 'Positive regulation of NF-kappaB import into nucleus'. This is relevant given that five out of the eight profiled tumours were from the DNA methylation class prediction EPN_RELA, a subgroup associated with aberrant NF-kB signalling (Parker et al., 2014). The 'immune response' subheading also included a number of terms related to the type I interferon (IFN)

response including GO:0060337 – 'Type I interferon signalling pathway', GO:0032481 – 'Positive regulation of type I interferon production' and GO:0032479 – 'Regulation of type I interferon production'.

Under the 'response to external biotic stimulus' subheadings there were 20 terms, all directly related to inflammatory and immune responses.

GO Term	Description	P-value	FDR value
GO:0006955	immune response	<0.001	<0.001
GO:0002376	immune system process	<0.001	<0.001
GO:0006952	defense response	<0.001	<0.001
GO:0002682	regulation of immune system process	<0.001	<0.001
GO:0050776	regulation of immune response	<0.001	<0.001
GO:0002684	positive regulation of immune system process	<0.001	<0.001
GO:0070665	positive regulation of leukocyte proliferation	<0.001	<0.001
GO:0043207	response to external biotic stimulus	<0.001	<0.001
GO:0001817	regulation of cytokine production	<0.001	<0.001
GO:0042102	positive regulation of T cell proliferation	<0.001	<0.001

Table 6-6: Top ten most significantly enriched ontology terms at first recurrence in the ST FFPE cohort.

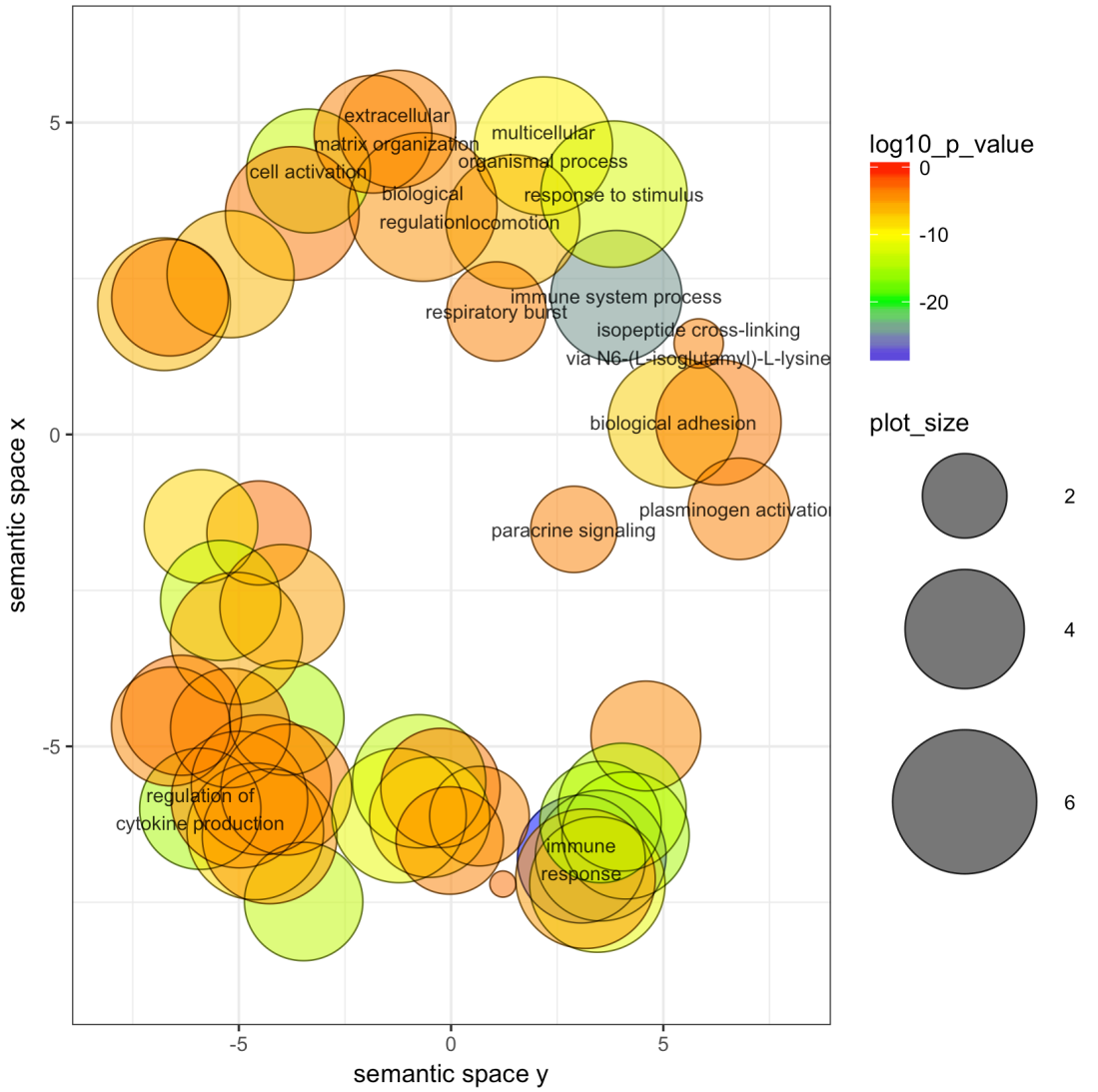


Figure 6-3: REVIGO plot of the ontologies significantly enriched at first recurrence in the supratentorial cohort. Redundant terms are removed. Plot size represents specificity of the term and colour represents the p value. Note the large number of circles representing terms overlapping with immune response and regulation of cytokine production categories, representing 63 and 64 sub terms respectively.

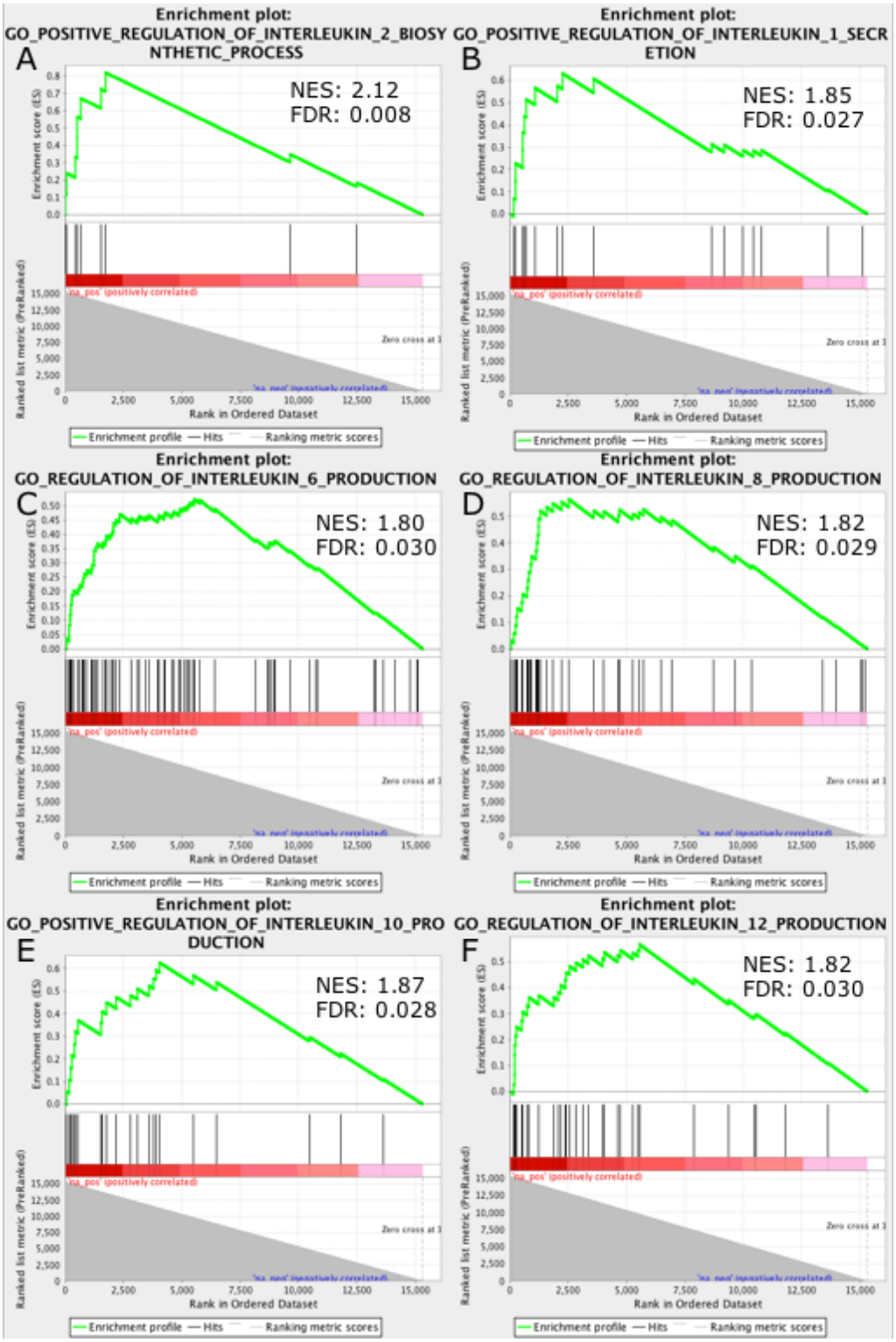


Figure 6-4: A-F: Enrichment plots of interleukin related terms significantly enriched at first recurrence in the FFPE supratentorial dataset. FDR: False discovery rate significance level.

GOrilla identified 30 significantly downregulated terms at first recurrence (Appendix 5). None of these were related to the immune response. Nine terms (30%) were associated with the glycolytic process and energy metabolism (Table 6-7). When the process was repeated with GSEA, no significant ontologies were identified.

Gene Ontology Terms	P-Value	FDR
GO:0006735 - NADH regeneration	<0.001	0.006
GO:0061718 – Glucose catabolic process to pyruvate	<0.001	0.005
GO:0061621 – Canonical glycolysis	<0.001	0.005
GO:0061620 – Glycolytic process through glucose-6-phosphate	<0.001	0.006
GO:0061615 – Glycolytic process through fructose-6-phosphate	<0.001	0.005
GO:0006007 – Glucose catabolic process	<0.001	0.015
GO:0006096 – Glycolytic process	<0.001	0.018
GO:0006757 – ATP generation from ADP	<0.001	0.024
GO:0019320 – Hexose catabolic process	<0.001	0.044

Table 6-7: List of terms downregulated in ST ependymoma at relapse and associated with energy metabolism or glycolysis.

6.3.3 Recurrence in the EPN_RELA molecular group

EPN_RELA tumours formed the largest molecular subgroup within the ST location and were re-analysed independently to identify whether they exhibited features differentiating them from the remainder of the cohort.

Five FFPE EPN_RELA samples were analysed; after filtering, 13742 genes were tested. This generated 411 genes differentially expressed at $p<0.05$ at first recurrence but none at FDR<0.05.

138 of the terms identified as upregulated in the ST location cohort were also identified as upregulated in the EPN_RELA specific analysis ($p<0.001$, hypergeometric test) (Appendix 5). Over 50% of these were related to the immune or inflammatory responses, including terms related to interleukin upregulation and cytokine production (Appendix 5). No terms downregulated at first recurrence in the overall ST cohort reached statistical significance in the EPN_RELA only cohort.

40 terms were upregulated at first recurrence in the EPN_RELA cohort but not in the overall ST cohort. Only one of these terms was related to the immune response, 'GO:00771347: Cellular response to interleukin-1'. The remaining

terms were heterogeneous but a number were related to transport, including transport of lipids, organic anions, carboxylic acids, phospholipids and organic acids (Table 6-8).

GO Term	Description	P-value	FDR value
GO:0043087	regulation of GTPase activity	<0.001	<0.001
GO:0009653	anatomical structure morphogenesis	<0.001	0.002
GO:0006869	lipid transport	<0.001	0.003
GO:0048646	anatomical structure formation involved in morphogenesis	<0.001	0.005
GO:0043085	positive regulation of catalytic activity	<0.001	0.005
GO:0003073	regulation of systemic arterial blood pressure	<0.001	0.005
GO:0051056	regulation of small GTPase mediated signal transduction	<0.001	0.006
GO:0015711	organic anion transport	<0.001	0.007
GO:0014910	regulation of smooth muscle cell migration	<0.001	0.009
GO:0006820	anion transport	<0.001	0.009
GO:0032502	developmental process	<0.001	0.011
GO:0035023	regulation of Rho protein signal transduction	<0.001	0.011
GO:0098609	cell-cell adhesion	<0.001	0.012
GO:0006811	ion transport	<0.001	0.013
GO:0046942	carboxylic acid transport	<0.001	0.019
GO:0015849	organic acid transport	<0.001	0.019
GO:0015914	phospholipid transport	<0.001	0.019
GO:0035850	epithelial cell differentiation involved in kidney development	<0.001	0.020
GO:0007154	cell communication	<0.001	0.020
GO:0042908	xenobiotic transport	<0.001	0.023

Table 6-8: Top 20 terms upregulated at first recurrence in the FFPE EPN_RELA tumours but not the ST tumours.

6.3.4 Recurrence in posterior fossa tumours

The FFPE and FF cohorts included 16 and 20 matched primary and first recurrences respectively, in which the primary tumours had both PF location and clustered into a PF gene expression group. DNA methylation class predictions were available for 15 (94%) FFPE and 13 (65%) FF primary samples. All predictions for both groups were EPN_PFA, apart from one FFPE prediction which was EPN_PFB.

In order to interpret any differences in expression changes at recurrence between FFPE and FF PF cohorts, clinical parameters were compared between the two groups. There was no difference in extent of initial resection or tumour grade (63% GTR in the FFPE groups Vs 62% GTR in the FF group, p=1.000, 50% WHO

Grade II in the FFPE group versus 38% WHO Grade II in the FF group, p=0.718). In the FFPE cohort, 10 (63%) children had not received radiotherapy between primary diagnosis and first recurrence, whereas in the FF cohort 4 (24%) had not received radiotherapy (p=0.037).

After filtering, 14485 genes were tested in both FFPE and FF cohorts. In the FFPE cohort, 475 genes were differentially expressed at first recurrence at $p<0.05$ and three genes remained after FDR correction. In the FF cohort, 1217 genes were differentially expressed at first recurrence at $p<0.05$ and five genes remained after FDR correction. Across both cohorts, 40 overlapping genes were differentially expressed with the same fold change direction. This was not statistically significant (p=0.452, hypergeometric test), suggesting that there was not a strong overlap in change at first recurrence between these two cohorts.

RNA-Seq meta-analysis of the FFPE and FF datasets identified three genes that were significantly upregulated (*KCNE1*, *LOC388820*, *COL15A1),* and five genes that were significantly downregulated, at first recurrence (*PDGFRA*, *ASIC4*, *EMILIN3*, *LHFPL3* and *NKD1)* (Table 6-9).

Genes upregulated at first posterior fossa recurrence		
Gene	**Fold Change**	**FDR Value**
Potassium Voltage-gated Channel Subfamily E Regulatory Subunit 1 (*KCNE1*)	+2.42	0.007
Putative Uncharacterized Protein LOC388820 *(LOC388820)*	+3.30	0.008
Collagen Type XV Alpha 1 Chain (*COL15A1*)	+2.43	0.044
Genes downregulated at first posterior fossa recurrence		
Gene	**Fold Change**	**FDR Value**
Platelet Derived Growth Factor Receptor Alpha (*PDGFRA*)	-2.55	0.005
Acid Sensing Ion Channel Subunit Family Member 4 (*ASIC4*)	-2.68	0.007
Elastin Microfibril Interfacer 3 (*EMILIN3*)	-3.27	0.008
Lipoma HMGIC Fusion Partner-Like 3 Protein (*LHFPL3*)	-3.25	0.008
Naked Cuticle Homolog 1 (*NKD1*)	-1.48	0.044

Table 6-9: List of genes significantly up and down regulated at first recurrence in all PF tumours based on meta-analysis of FFPE and FF datasets.

In the GOrilla analysis, between FFPE and FF cohorts, there were no shared gene ontology terms upregulated at first recurrence. There were nine, non-specific, shared terms downregulated at first recurrence. These were:

- GO:0048519 – Negative regulation of biological process;
- GO:0030154 – Cell differentiation;
- GO:0032502 – Developmental process;
- GO:0048731 – System development;
- GO:0048856 – Anatomical structure development;
- GO:0048869 – Cellular developmental process;
- GO:0009887 – Animal organ morphogenesis;
- GO:0007399 – Nervous system development;
- GO:0045664 – Regulation of neuron differentiation.

It became apparent that FFPE and FF datasets exhibited opposing patterns of gene expression at first recurrence. 98 of the 193 upregulated ontologies in the FFPE dataset overlapped with 98 of the 529 downregulated ontologies in the FF dataset. This was highly statistically significant ($p<0.001$, hypergeometric test) (Appendix 5). These terms were almost exclusively related to the immune and inflammatory responses.

Immune ontologies were also seen in the terms that did not overlap, but were related to different mechanisms; the adaptive immune response was upregulated at first recurrence in the FF cohort, with 50 related terms, including those pertaining to both T-cell stimulation and B-cell activation. The innate immune response was downregulated at first recurrence in the FFPE cohort with the identification of no B- or T-cell specific terms. This pattern was recapitulated during the GSEA analysis (Figure 6-5).

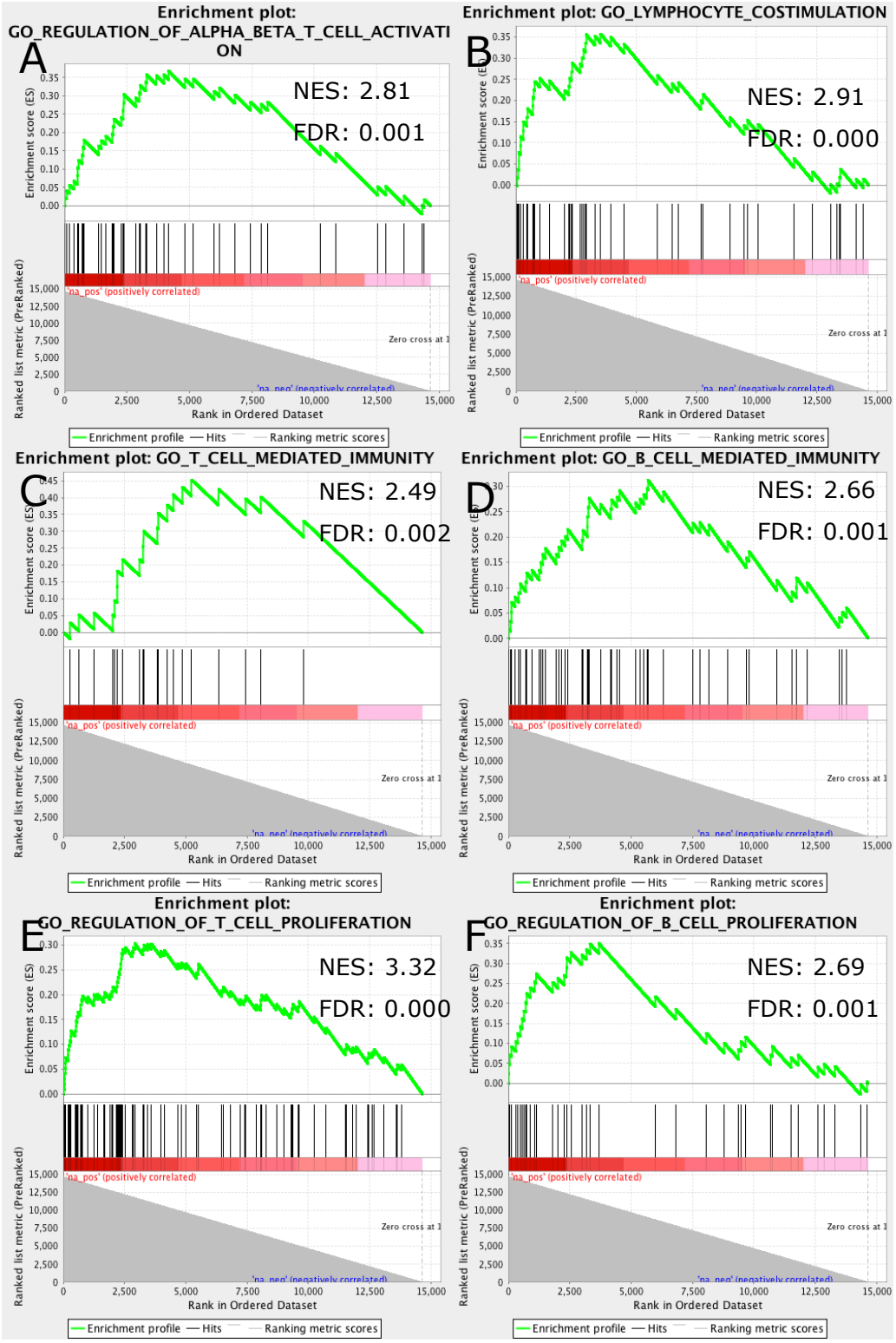


Figure 6-5: A-F: Significantly enriched lymphocyte and adaptive immunity related terms upregulated at first recurrence in the FF posterior fossa dataset and not upregulated in the response in the FFPE posterior fossa dataset. FDR: False discovery rate significance. NES: Normalised enrichment score.

GOrilla and GSEA also identified a particular immune ontology of interest, in which the FFPE and FF datasets diverged; 'GO:0034340 - Response to Type I interferon'. In the FF dataset, the term was significantly upregulated in the recurrent tumours, with 41 out of a possible 50 genes being associated with the ontology (ES 4.40, FDR=0.000), whereas in the FFPE dataset the term was downregulated in the recurrent tumours, with 32 out of a possible 50 genes being associated (ES -3.36, FDR=0.000) (Figure 6-6). 26 genes were in the core list for the Response to type I IFN ontology in both datasets, demonstrating opposing directions of fold change (Table 6-10).

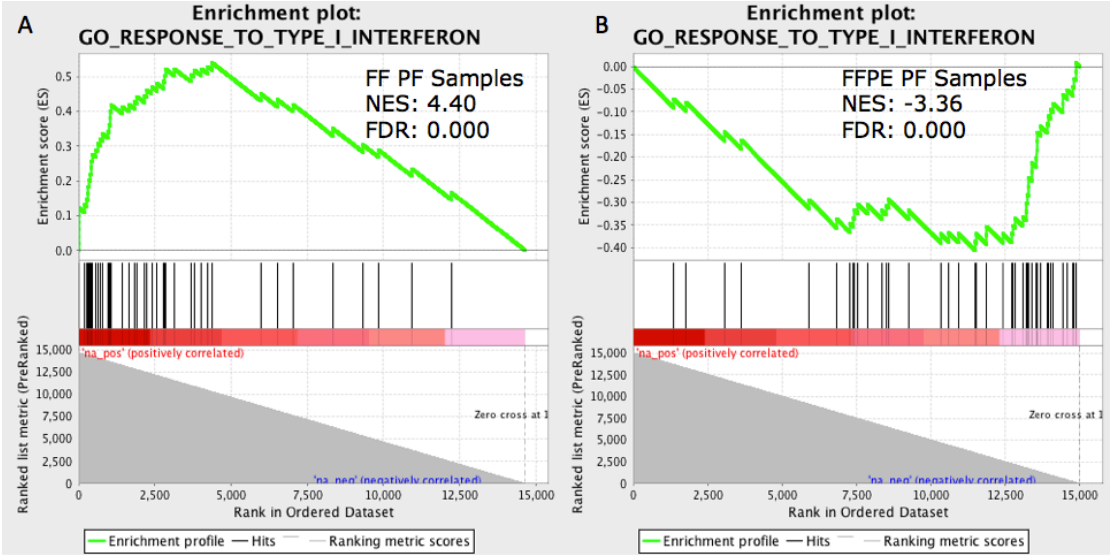


Figure 6-6: Enrichment plots for GO:0034340 - Response to type I IFN at first recurrence compared to primary presentation in (A) the FF posterior fossa cohort and (B) the FFPE posterior fossa cohort. FDR: False discovery rate significance level, NES: Normalised enrichment score.

Guanylate binding protein 2 (*GBP2*)
2’-5’-oligoadenylate synthetase 2 (*OAS2*)
2’-5’-oligoadenylate synthetase 3 (*OAS3*)
SAM and HD domain containing deoxynucleoside triphosphate triphosphohydrolase 1 (*SAMHD1*)
Major histocompatibility complex, class I, C (*HLA-C*)
Interferon stimulated exonuclease (*ISG20*)
Signal transducer and activator of transcription 2 (*STAT2*)
Major histocompatibility complex, Class I, B (*HLA-B*)
Interferon regulatory factor 8 (*IRF8*)
Interferon alpha inducible protein (*IFI27*)
Interferon induced protein with tetratricopeptide repeats 2 (*IFIT2*)
Chromosome 19 open reading frame 66 (*C19orf66*)
Signal transducer and activator of transcription 1 (*STAT1*)
XIAP associated factor 1 (*XAF1*)
Tripartite motif containing 56 (*TRIM56*)
2’-5’-oligoadenylate synthetase 1 (*OAS1*)
Major histocompatibility complex, class I, F (*HLA-F*)
Major histocompatibility complex, class I, E (*HLA-E*)
MX dynamin like GTPase 1 (*MX1*)
ISG15 ubiquitin-like modifier (*ISG15*)
SP100 nuclear antigen (*SP100*)
Interferon induced protein with tetratricopeptide repeats 3 (*IFIT3*)
Interferon regulatory factor 2 (*IRF2*)
Bone marrow stromal cell antigen 2 (*BST2*)
Interferon regulatory factor 1 (*IRF1*)
Interferon induced transmembrane protein 1 (*IFITM1*)

Table 6-10: Core list of genes for term GO:0034340 - Response to type I IFN expressed in opposing directions in FFPE (downregulated at recurrence) and FF (upregulated at recurrence) posterior fossa location datasets when analysed by GSEA.

When looking at the composition of the FFPE and FF posterior fossa cohorts by molecular subgroup; the FFPE posterior fossa group contained 82% PF1 and 18% PF2 primary tumours, compared to the FF posterior fossa group which contained 45% PF1 and 55% PF2 primary tumours (p=0.011, chi-square test). It was therefore hypothesised that the divergent pattern of immune enrichment described in the FFPE and FF posterior fossa location cohorts was due to a different distribution of molecular subgroups. Therefore, recurrence patterns within PF1 and PF2 molecular subgroups were investigated.

6.3.5 Recurrence in the PF1 subgroup

The FFPE and FF PF1 groups contained 9 and 13 primary and recurrent pairs respectively. Other than receipt of radiotherapy, there were no differences between the two cohorts. In the FFPE dataset only 4 of 13 (30%) patients received radiotherapy between initial diagnosis and first recurrence, whilst in the FF dataset 7 of 9 (78%) patients received radiotherapy (p=0.030).

After filtering, 14108 genes were tested across both datasets. In the FFPE cohort, 569 genes were differentially expressed at first recurrence with p<0.05 and three at FDR<0.05. In the FF cohort, 1759 genes were differentially expressed at first recurrence with p<0.05 and 323 at FDR<0.05. A total of 102 statistically significant genes with matching fold change direction appeared in both datasets. This represented a statistically significant association between FFPE and FF datasets (p<0.001, hypergeometric test).

Genes significantly downregulated in both datasets included a number of cytokines and their receptors: Interleukin 1 receptor type 2 (*IL1R2*); C-X-C motif chemokine ligand 1 (*CXCL1*); C-X-C motif chemokine ligand 8 (*CXCL8*); C-C motif chemokine ligand 2 (*CCL2*); C-X-C motif chemokine ligand 3 (*CXCL3*); interleukin 1 receptor type 1 (*IL1R1*); and interleukin 11 receptor subunit alpha (*IL11RA*).

RNA-seq meta-analysis of the FF and FFPE datasets identified 133 differentially expressed genes, 37 of which were upregulated (Table 6-11) and 96 which were downregulated at first recurrence (Table 6-12).

Gene Symbol	FDR	Gene Symbol	FDR	Gene Symbol	FDR
DLEC1	<0.001	*PLCXD3*	0.004	*CTNNA2*	0.029
SLCO3A1	<0.001	*DCHS2*	0.007	*C14orf180*	0.030
PER3	<0.001	*CXCL12*	0.007	*TMPRSS7*	0.033
FGF1	<0.001	*NR1D2*	0.007	*PAMR1*	0.034
CLSTN2	0.001	*LINC01354*	0.008	*PNMAL1*	0.035
KIAA1217	0.001	*LOC153684*	0.008	*CD200*	0.036
MYO5C	0.001	*ANKRD45*	0.010	*EPHX1*	0.036
AGT	0.001	*USH1C*	0.014	*RASGRF2*	0.036
CDS1	0.001	*EFHB*	0.015	*MATN2*	0.039
NWD1	0.002	*CNGA3*	0.016	*GRID1*	0.040
LGR6	0.003	*CACNA1D*	0.017	*FHOD3*	0.043
C5orf64	0.003	*NAV3*	0.017		
COL15A1	0.003	*IFIT1*	0.024		

Table 6-11: Upregulated genes in PF1 recurrence from RNA-seq meta-analysis of FFPE and FF datasets.

Gene Symbol	FDR	Gene Symbol	FDR	Gene Symbol	FDR
LHFPL3	<0.001	*HMCN1*	0.008	*SCNN1B*	0.030
EMILIN3	<0.001	*PLEKHG4*	0.008	*HAPLN4*	0.030
C1QL4	<0.001	*MDFI*	0.010	*NAMPT*	0.030
GRK5	<0.001	*CHI3L2*	0.010	*PPAT*	0.032
MEG3	<0.001	*ASIC4*	0.010	*ANKRD36BP2*	0.033
MME	<0.001	*OLFM2*	0.010	*CDC42EP3*	0.033
CXCL1	<0.001	*C1QL1*	0.010	*TMEM158*	0.033
TGFBI	<0.001	*WNT7A*	0.011	*PRPH*	0.033
BMP2	0.001	*ARC*	0.012	*IRX2*	0.034
SLC26A7	0.001	*CACNA1I*	0.012	*AQP9*	0.035
SLC6A11	0.001	*SEMA5B*	0.012	*CA10*	0.035
ITGA5	0.001	*IL1R2*	0.013	*FGF7*	0.035
HSD11B2	0.001	*CDCP1*	0.013	*FKBP10*	0.035
MEG8	0.001	*CRTAC1*	0.013	*CHPF*	0.036
SERPINA3	0.002	*OBSCN*	0.014	*FERMT1*	0.036
FNDC4	0.002	*LGR5*	0.014	*IQGAP2*	0.036
IL1RL1	0.002	*IL1R1*	0.015	*TIPARP*	0.036
MET	0.002	*SCNN1G*	0.015	*GLIPR2*	0.036
NRK	0.002	*HSPB1*	0.015	*MEIS3*	0.036
CDH15	0.002	*CSPG4*	0.016	*RARA*	0.037
WEE1	0.003	*TSPAN18*	0.016	*TNFRSF12A*	0.037
STAC	0.003	*ANGPTL4*	0.016	*PTX3*	0.037
GLRX	0.004	*ADGRG1*	0.020	*ALDH1L1*	0.040
NFIL3	0.004	*GRIK3*	0.021	*DCHS1*	0.041
PLIN2	0.004	*TRIM47*	0.022	*CD163*	0.043
EMP1	0.005	*CXCL6*	0.024	*LDHA*	0.043
NNMT	0.006	*H19*	0.024	*FAM20C*	0.043
PRRG3	0.007	*COL6A2*	0.024	*PEG10*	0.045
CADPS	0.007	*C1R*	0.027	*ERBB3*	0.045
STC2	0.008	*SERPINH1*	0.027	*PHLDA1*	0.049
DIO3	0.008	*CACNG4*	0.029	*SOCS3*	0.049
FAM46B	0.008	*TMEM97*	0.030	*GABRA5*	0.050

Table 6-12: Downregulated signature genes of PF1 recurrence generated from RNA-seq meta-analysis combining FFPE and FF datasets.

FFPE and FF cohorts generated 179 significantly downregulated overlapping ontologies at first recurrence ($p<0.001$, hypergeometric test) (Appendix 5). This included 52 ontologies related to the immune and inflammatory responses, including neutrophil chemotaxis and cytokine mediated signalling pathways (Table 6-13, Figure 6-7).

GO Term	Description
GO:0006954	inflammatory response
GO:0002526	acute inflammatory response
GO:0006952	defense response
GO:0006935	chemotaxis
GO:0042330	taxis
GO:0043408	regulation of MAPK cascade
GO:0060326	cell chemotaxis
GO:0043410	positive regulation of MAPK cascade
GO:0050900	leukocyte migration
GO:0006953	acute-phase response
GO:0030595	leukocyte chemotaxis
GO:0050920	regulation of chemotaxis
GO:0050727	regulation of inflammatory response
GO:0030334	regulation of cell migration
GO:0002687	positive regulation of leukocyte migration
GO:0002690	positive regulation of leukocyte chemotaxis
GO:0071621	granulocyte chemotaxis
GO:0050921	positive regulation of chemotaxis
GO:0030593	neutrophil chemotaxis
GO:0002685	regulation of leukocyte migration
GO:0009617	response to bacterium
GO:2000145	regulation of cell motility
GO:0002688	regulation of leukocyte chemotaxis
GO:0006959	humoral immune response
GO:0097529	myeloid leukocyte migration
GO:0097530	granulocyte migration
GO:1990266	neutrophil migration
GO:0016477	cell migration
GO:0034097	response to cytokine
GO:0002237	response to molecule of bacterial origin
GO:0042742	defense response to bacterium
GO:0043406	positive regulation of MAP kinase activity
GO:0043207	response to external biotic stimulus
GO:0019221	cytokine-mediated signaling pathway
GO:0006955	immune response
GO:0048870	cell motility
GO:0050729	positive regulation of inflammatory response
GO:0070098	chemokine-mediated signaling pathway
GO:0009607	response to biotic stimulus
GO:0001819	positive regulation of cytokine production

Table 6-13: Top 40 significantly downregulated gene ontology terms, related to immune and inflammatory responses, at first relapse in both PF1 gene expression groups. Terms in this table are shared between the FFPE and FF datasets. Full list of all terms alongside significance values can be found in Appendix 5.

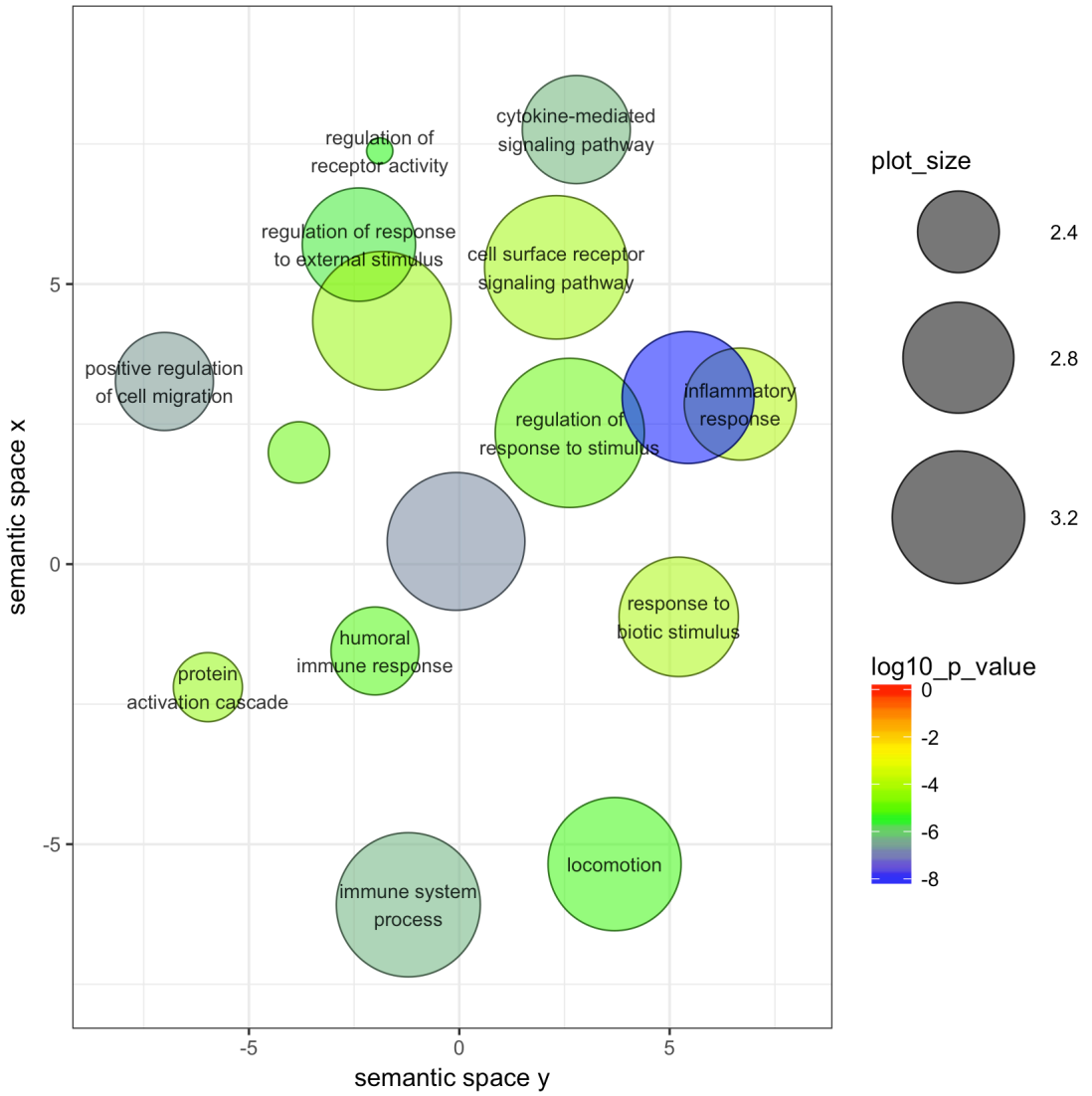


Figure 6-7: A scatterplot of none-redundant significantly downregulated gene ontology terms at first recurrence in the PF1 subgroup, generated from genes that were statistically significant in both the FFPE and FF cohorts. The colour of the circles represents the p-value and the size of the circle represents the specificity of the term (more general terms appear as larger circles).

The GSEA analysis supported these findings with downregulation of the eight following overlapping terms at recurrence: GO:0006954 - Inflammatory response; GO:0030155 – regulation of cell adhesion; GO:0001817 – regulation of cytokine production; GO:0043410 – Positive regulation of MAPK cascade; GO:0050727 – Regulation of inflammatory response; GO:0043408 – Regulation of MAPK cascade; GO:0040017 – Positive regulation of locomotion; GO:0045596 – Negative regulation of cell differentiation.

The FFPE cohort demonstrated fewer ontology terms upregulated at first recurrence than the FF cohort (6 and 138 respectively). Only one significantly upregulated ontology overlapped in both datasets ('GO:0006813: Potassium ion transport') (p=0.056, hypergeometric test).

Key upregulated ontologies exclusive to the FF dataset were related to:
(1) the IFN I pathway and defense response to virus;
(2) cell-cell signalling and adhesion;
(3) ion transport.
GSEA also identified GO:0051607 – 'Defense response to virus' with FDR 0.013.

A number of individual genes, differentially expressed at the FDR level in the FF dataset, also featured in the Type I IFN pathway related ontologies. These were *OAS1* (LFC 1.88, FDR 0.002), *OAS2* (LFC 1.34, FDR 0.013), *RSAD2* (LFC 1.44, FDR 0.028), *MX1* (LFC 1.23, FDR 0.050), *IFI27* (LFC 1.78, FDR 0.018).

The equivalent type I IFN pathway related terms were not significantly upregulated in the FFPE dataset.

6.3.5.1 PF1 relapse profile and radiotherapy

Patients in the FFPE PF1 subgroup were less likely to have been treated with radiotherapy between primary and first recurrence than in the FF PF1 subgroup (p=0.030). It was hypothesised that the upregulated immune response, in particular the type I IFN pathway, seen in the FF dataset may be associated with treatment with radiotherapy. To investigate this further, the FFPE cohort was divided into those treated with radiotherapy and those who were not.

In the non-irradiated FFPE group, there were no significantly upregulated terms at first recurrence. In the four patients who received radiotherapy there were 227 terms significantly upregulated at FDR<0.05 at first recurrence. 32 (14%) of these terms were related to the immune response (Table 6-14) and 8 (3.5%) were related to cell death (Table 6-15). The type I IFN pathway was not upregulated in this analysis, therefore only partially supporting the hypothesis. Consistent with the FF cohort, in both irradiated and non-irradiated FFPE groups, the downregulated ontology terms reflected a decreased inflammatory and immune response.

GO Term	Description	FDR
GO:0045055	regulated exocytosis	<0.001
GO:0043299	leukocyte degranulation	<0.001
GO:0002275	myeloid cell activation involved in immune response	<0.001
GO:0043312	neutrophil degranulation	<0.001
GO:0002283	neutrophil activation involved in immune response	<0.001
GO:0008284	positive regulation of cell proliferation	<0.001
GO:0036230	granulocyte activation	<0.001
GO:0002366	leukocyte activation involved in immune response	<0.001
GO:0001775	cell activation	<0.001
GO:0002274	myeloid leukocyte activation	<0.001
GO:0002263	cell activation involved in immune response	<0.001
GO:0042119	neutrophil activation	<0.001
GO:0006887	exocytosis	<0.001
GO:0045321	leukocyte activation	<0.001
GO:0002376	immune system process	<0.001
GO:0030334	regulation of cell migration	<0.001
GO:0002252	immune effector process	0.006
GO:0016477	cell migration	0.020
GO:1902105	regulation of leukocyte differentiation	0.022
GO:0040011	locomotion	0.027
GO:0048870	cell motility	0.028
GO:0050863	regulation of T cell activation	0.032
GO:0030335	positive regulation of cell migration	0.032
GO:0071496	cellular response to external stimulus	0.032
GO:0034097	response to cytokine	0.035
GO:0050900	leukocyte migration	0.038
GO:0006950	response to stress	0.038
GO:0006954	inflammatory response	0.040
GO:0009611	response to wounding	0.046
GO:1903037	regulation of leukocyte cell-cell adhesion	0.046
GO:0071345	cellular response to cytokine stimulus	0.047
GO:0002694	regulation of leukocyte activation	0.048

Table 6-14: Significantly enriched immune related terms at first recurrence in irradiated FFPE PF1 tumours (n=4).

GO Term	Description	FDR
GO:0010941	regulation of cell death	<0.001
GO:0060548	negative regulation of cell death	<0.001
GO:0043069	negative regulation of programmed cell death	<0.001
GO:0043067	regulation of programmed cell death	<0.001
GO:0042981	regulation of apoptotic process	<0.001
GO:0012501	programmed cell death	0.009
GO:0008219	cell death	0.012
GO:0043068	positive regulation of programmed cell death	0.046

Table 6-15: Significantly enriched cell death related terms at first recurrence in irradiated FFPE PF1 tumours (n=4).

6.3.5.2 RT-qPCR validation of upregulated type I IFN related genes in the FF PF1 dataset

Whilst radiotherapy treated tumours were associated with upregulation of immune ontologies in both FFPE and FF cohorts, the upregulation of the type I IFN pathway was only identified in the FF, and not the FFPE, dataset. Therefore, in order to provide a technical validation of this finding in the FF dataset, real time quantitative PCR was performed for a subset of significantly differentially expressed type I IFN related genes.

RNA was only available to perform this for four cases. An a priori decision was made not to perform statistical analysis on this sample set in view of low statistical power, but to confirm that the direction of fold change matched for the qPCR and RNA-seq samples. The genes tested were *CXCL12, OAS3, BST2, RSAD2, OAS2, OAS1* and *MX1*.

Every gene tested expressed a direction of fold change consistent with that seen in the RNA-seq dataset (upregulated at first recurrence). The values calculated for log2 fold change for each gene for the qPCR analysis were: *CXCL12* +1.57, *OAS3* +2.71, *BST2* +1.08, *RSAD2* +4.10, *OAS2* +3.11, *OAS1* +3.32, *MX1* +2.70. For the RNA-seq data the log2 fold changes were: *CXCL12* +3.22, *OAS3* +0.94, *BST2* +1.44, *RSAD2* +2.33, *OAS2* +1.86, *OAS1* +3.78, *MX1* +2.69. The lower point of the error bar, based on standard error of the mean, still remained positive for all samples. For most genes, there was a greater than two-fold increase in gene expression at first recurrence, measured by both RNA-seq and RT-qPCR (Figure 6-8); providing technical confirmation of the FF cohort RNA-seq findings.

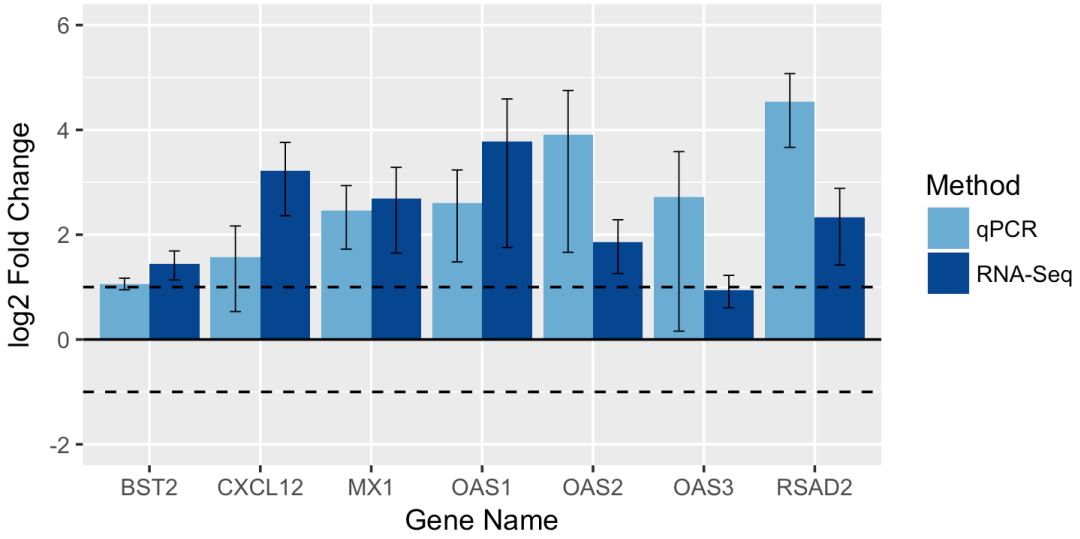


Figure 6-8: Bar plots demonstrating log2 of the fold change for genes chosen for the technical validation in the posterior fossa group PF1 of the FF dataset (n=4). All genes demonstrated upregulation at first recurrence and this is represented by a fold change greater than zero. The dashed lines represent a fold change of two (positive and negative). The solid line represents zero-fold change. Error bars are based on standard error of the biological replicates.

6.3.6 Recurrence in the PF2 subgroup

The FFPE and FF PF2 subgroups contained 3 and 11 matched pairs respectively. Tumours with ST location were excluded.

Due to small sample size in the FFPE dataset, clinical variables were compared using a Fisher's exact test. No statistically significant differences were demonstrated for patient demographics or treatment approaches between the FFPE and FF datasets. Nine of the 14 patients in this group received radiotherapy between primary and first recurrence (2/3 FFPE patients, 7/11 FF patients, p=1.000).

After filtering, 14289 genes were tested across both datasets. In the FFPE dataset 499 genes were differentially expressed at first recurrence with $p<0.05$ and 22 with FDR<0.05. In the FF dataset 1219 genes were differentially expressed at first recurrence with $p<0.05$ and 39 with FDR<0.05. There were 55 overlapping significant genes with the same direction of fold change between FF and FFPE cohorts (p=0.029, hypergeometric test).

RNA-Seq meta-analysis identified 47 genes below the FDR<0.05 cut off, 7 downregulated (Table 6-16) and 40 upregulated (Table 6-17).

Gene Symbol	FDR
PCDH15	0.001
RNY1	0.003
ACSS1	0.012
DSCAM	0.031
SLC16A4	0.032
NKAIN4	0.035
LHFPL3	0.036

Table 6-16: Downregulated genes generated from RNA-Seq meta-analysis of FF and FFPE datasets in PF2 group relapses.

Gene Symbol	FDR	Gene Symbol	FDR	Gene Symbol	FDR
MOBP	<0.001	*LTBP2*	0.003	*PLVAP*	0.035
POSTN	<0.001	*CORO2A*	0.010	*ANK3*	0.036
UNC5C	<0.001	*SPTB*	0.010	*COL1A1*	0.036
HHIP	<0.001	*SVIL*	0.010	*DOCK5*	0.036
LUM	<0.001	*HBA1*	0.013	*ENPP2*	0.036
ALDH1A1	<0.001	*ARSJ*	0.013	*GCNT1*	0.036
THBS1	<0.001	*SLC7A14*	0.022	*MBP*	0.036
CLDN11	0.001	*SLC6A6*	0.023	*SELL*	0.036
LYZ	0.001	*LAMB1*	0.027	*UBASH3B*	0.036
MICAL2	0.002	*RYR2*	0.028	*RASGEF1B*	0.039
COL1A2	0.003	*ESM1*	0.030	*SLAMF8*	0.049
HBB	0.003	*HMOX1*	0.033	*KIAA1324L*	0.049
ST18	0.003	*C7*	0.033		
CADPS2	0.003	*GPNMB*	0.035		

Table 6-17: Upregulated genes generated from RNA-Seq meta-analysis of FF and FFPE datasets in PF2 group relapses.

In FFPE and FF PF2 datasets, 51 overlapping gene ontology terms were upregulated at first recurrence (Figure 6-9) (Appendix 5). Five were related to cell adhesion and 18 were related to either the immune or inflammatory response, cytokine release or chemotaxis. Four terms were related to changes in the extracellular matrix and its organisation. 18 immune related terms were not visible in the REVIGO plot as they were classified under the subheading 'immune system process'; they are presented separately (Table 6-18).

GO Term	Description
GO:0006955	immune response
GO:0006952	defense response
GO:0001819	positive regulation of cytokine production
GO:0006935	chemotaxis
GO:0042330	taxis
GO:0050900	leukocyte migration
GO:0001817	regulation of cytokine production
GO:0006954	inflammatory response
GO:0002376	immune system process
GO:2000145	regulation of cell motility
GO:0045087	innate immune response
GO:0001775	cell activation
GO:0040011	locomotion
GO:0060326	cell chemotaxis
GO:0030334	regulation of cell migration
GO:0040012	regulation of locomotion
GO:0050707	regulation of cytokine secretion
GO:0002697	regulation of immune effector process

Table 6-18: List of upregulated gene ontology terms related to immune function significantly enriched (FDR<0.05) in FFPE and FF PF2 datasets.

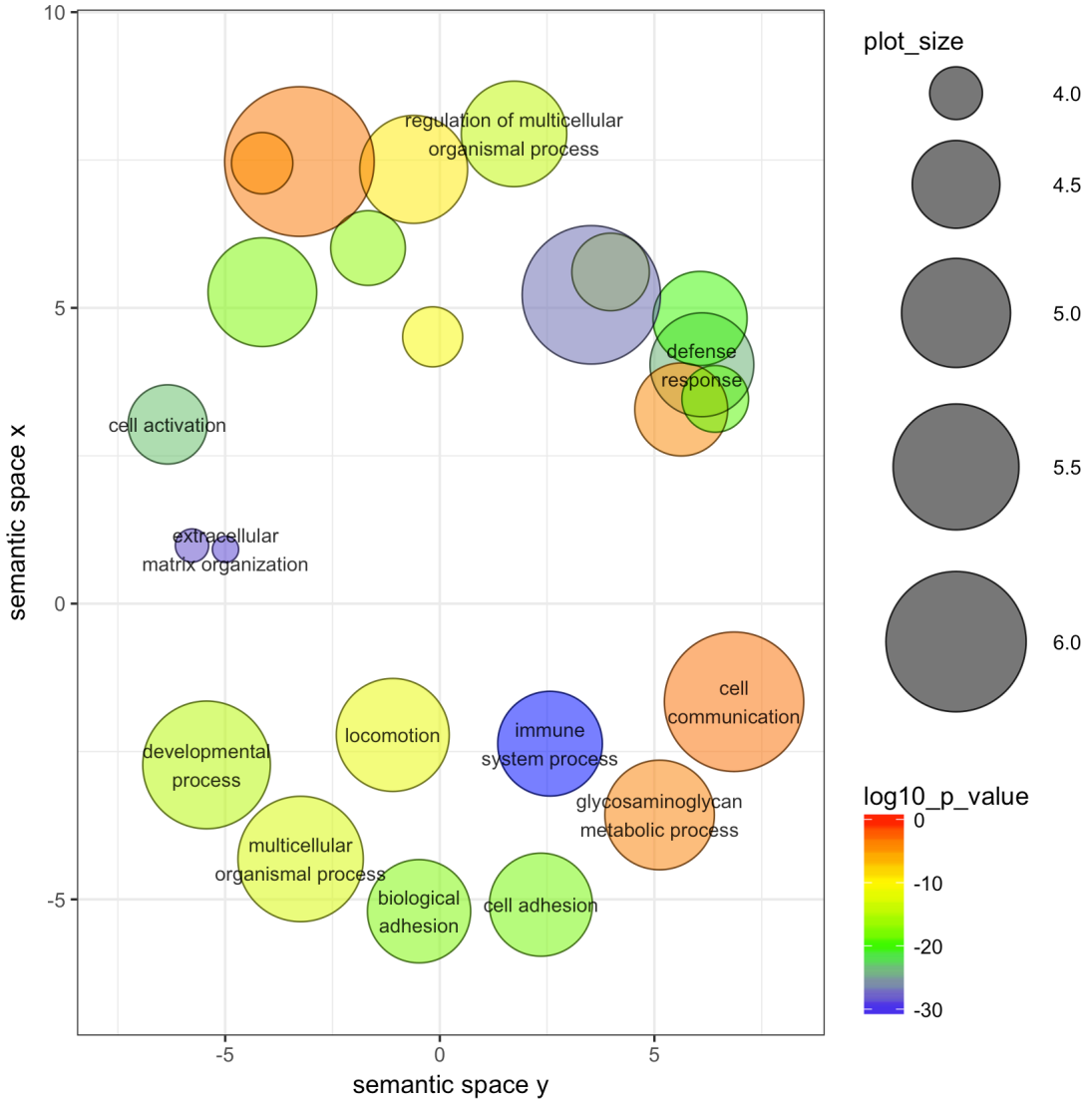


Figure 6-9: A scatterplot of non-redundant significantly upregulated gene ontology terms in FFPE and FF PF2 tumours at first recurrence generated from GOrilla enrichment analysis. The colour of the circles represents the p-value and the size of the circles represents the specificity of the term.

There were two overlapping downregulated terms: GO:0044282 – 'Small molecule catabolic process' and GO:0032787 – 'Monocarboxylic acid metabolic process'.

The FFPE cohort had fewer upregulated immune related terms than the FF cohort. The additional FF terms were particularly related to lymphocyte activation including GO:0050870 – 'Positive regulation of T cell activation' and GO:0050864 – 'Regulation of B cell activation'. Of interest for potential therapy was the upregulation of several ontologies related to programmed cell death, for example GO:0043067 – 'Regulation of programmed cell death'. The FF dataset also

demonstrated significant upregulation at first recurrence of multiple IFN related pathways, including GO:0034341 – 'Response to interferon gamma' and GO:0060337 – 'Type I interferon signalling' which was represented by the genes *IFITM1, HLA-A, RSAD2, HLA-C, HLA-B, OAS3, HLA-F, MYD88, IFI35, ISG20, IFITM2, GBP2, EGR1, XAF1, BST2, IRF1, RNASEL, MX2, MX1, OAS1, OAS2, IFNAR2, SP100* and *IFI6* (Figure 6-10).

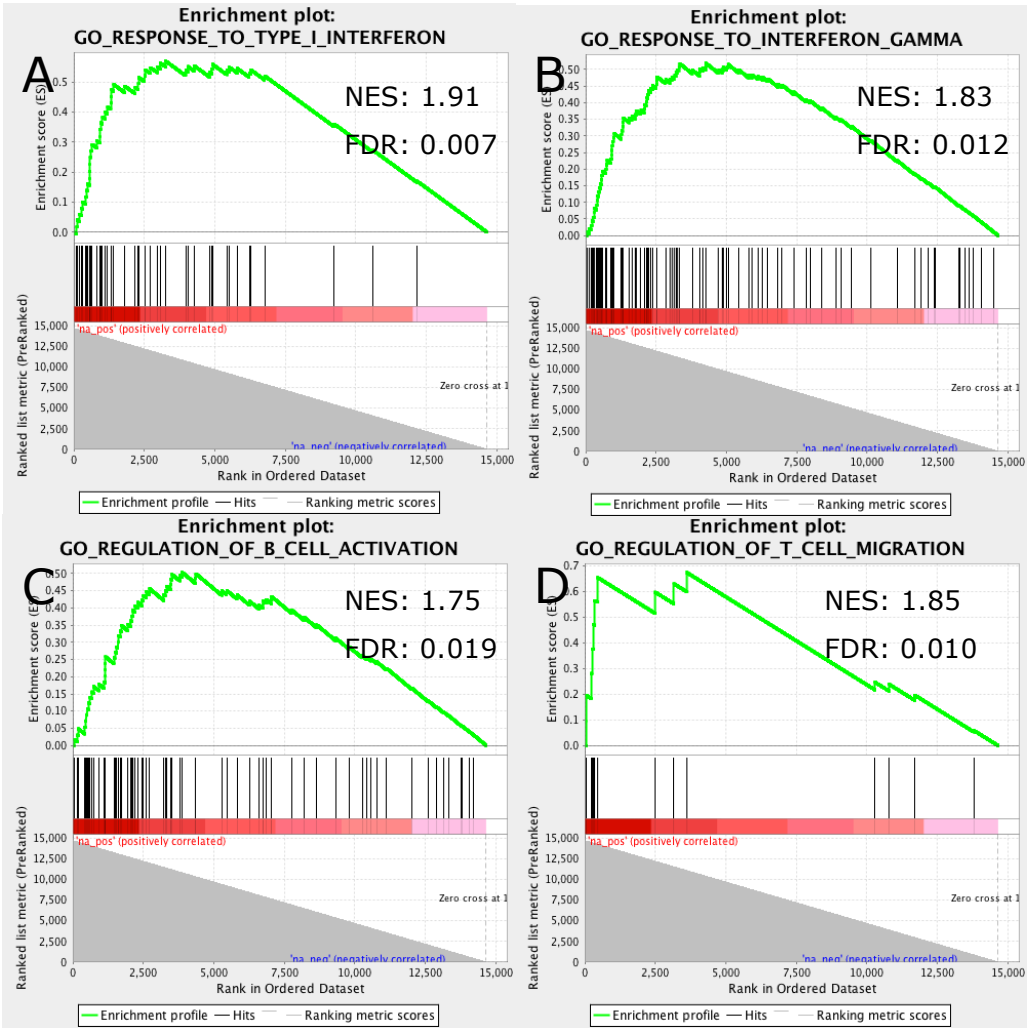


Figure 6-10: Significantly enriched IFN (A, B) and adaptive immunity (C, D) related terms at first recurrence in the FF PF2 dataset. FDR: False discovery rate significance. NES: Normalised enrichment score.

6.3.7 Immunophenoscores

The differential expression and enrichment analyses for the molecular subgroups provided evidence of changes to the immune response at first recurrence in paediatric ependymoma. In order to investigate for evidence of a functional change in gene expression, and therefore to provide an indication of any change in the susceptibility to checkpoint blockade therapy, the immunophenoscore of

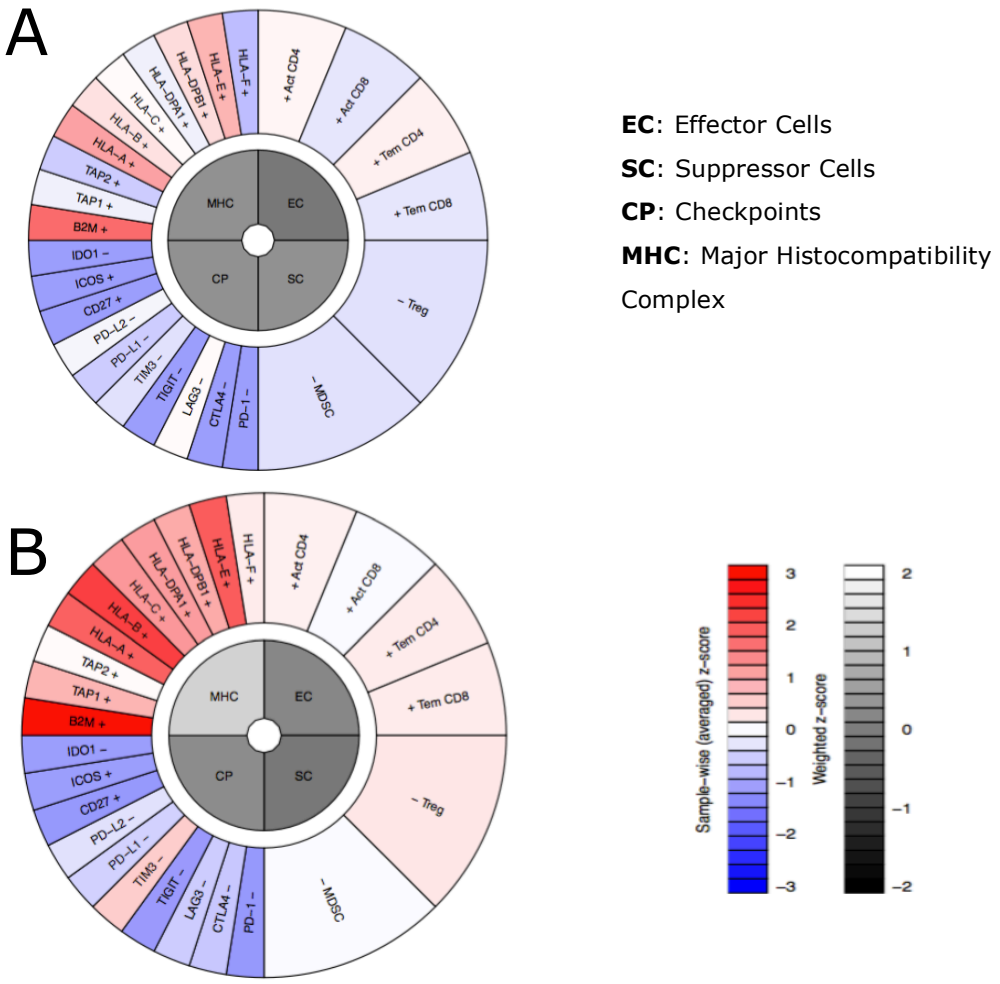
each tumour was assessed (Charoentong et al., 2017). As immunophenoscores are based on reference datasets and normalised, once generated they are comparable between datasets. Therefore, for this analysis, FFPE and FF cohorts were combined.

The immunophenoscore cohort included all 151 tumour samples, consisting of 67 primary tumours, 60 first recurrences, 12 second recurrences, 4 third recurrences, 5 fourth recurrences and 1 fifth, sixth and seventh recurrence. Of those with DNA methylation profiles, 86 were EPN_PFA, 13 EPN_RELA, 7 EPN_YAP, 3 EPN_PFB, 3 EPN_MPE and 11 non-ependymoma designations. DNA methylation profiles were unavailable for 28 of the samples. There were 53 matched primary and recurrent pairs.

The mean immunophenoscore was 5.31 (range 3-8) for all samples and 5.28 for primaries alone. There was no difference between the primary and recurrent tumours as two separate groups, so they were then analysed together ($p=0.662$). There were significant differences between scores for tumours in different locations; ST mean score 4.70 versus PF mean score 5.53 ($p<0.001$). Additionally, EPN_RELA had a significantly lower score than EPN_PFA (4.78 versus 5.48, $p=0.015$).

Categories contributing to the immunophenoscore (MHC, effector cells, suppressor cells, checkpoints and immunomodulators) were then compared between locations. In the MHC category, a significant difference was identified between the PF and ST cohort in the MHC category, indicating differing antigen presenting potential (PF mean 1.35, versus ST mean 1.12, $p=0.002$). There were no differences in effector cells, suppressor cells or checkpoint genes (Figure 6-11).

When comparing the two PF gene expression subgroups there were no significant differences between the overall immunophenoscores (PF1 mean 5.60 versus PF2 mean 5.45), however when comparing the categories, the PF1 tumours were enriched for MHC molecules and effector cells ($p=0.040$ and $p<0.001$ respectively), whilst the PF2 tumours were enriched for suppressor cells ($p=0.001$). There was no significant difference between the two PF subgroups for checkpoint scores ($p=0.080$). When including just the primary tumour specimens from PF subgroups, only effector cells maintained significance; with PF1 being more enriched than PF2 tumours ($p=0.041$).



EC: Effector Cells

SC: Suppressor Cells

CP: Checkpoints

MHC: Major Histocompatibility Complex

Figure 6-11: Immunophenoscores demonstrating (A) an ST sample (Score 3) and (B) a PF sample (Score 7). The main difference is identified in the Major Histocompatibility Complex (MHC) score, which is enriched in the PF sample. Both samples had low levels of checkpoint molecule expression.

When comparing all paired primary and first recurrent tumours there was no significant difference in overall immunophenoscore. However, effector cells significantly increased at recurrence (p=0.032). When subdividing this analysis into PF1, PF2 and ST tumours, no significant differences were identified between paired primary and first recurrence in either the overall immunophenoscore or in any of the four categories.

For all tumours, the levels of checkpoint associated genes were low, including checkpoint blockade targets *PD1* and *CTLA-4*. In many comparisons, the level of expression was so low that the gene did not pass the filters to be included in the

analysis. The median number of transcripts per million for *PD1* was 0. The median number of transcripts per million for *CTLA-4* was 0.39.

6.3.8 Levels of immune checkpoint gene expression in ependymoma

The identification of low levels of checkpoint genes in the immunophenoscores stimulated further analysis of a wider spectrum of checkpoint genes in ependymoma. Based on immunoinhibitory markers associated with T-cell anergy, discussed in a recent review (Catakovic et al., 2017), the gene expression data from both FF and FFPE datasets were re-reviewed. Data was separated into receptors (the immune checkpoints expressed on immune cells), consisting of: *PD1, CTLA-4, LAG3, TIM3, TIGIT, BTLA, CD96* and *CD244;* and ligands (the antigens expressed by cancer cells and presented to immune cells by antigen presenting cells), consisting of: *CD80, PDL1, PDL2, GAL9, CD112, CD155, HVEM, CD48* and *CD2* (Figure 6-12). Levels of receptors, apart from *TIM3,* were extremely low, with median values below one count per million. The ligand molecules were expressed at higher levels than the receptors, but only two genes were above the median expression level for all genes in the dataset. The only receptor with matched ligand that was expressed at reliably detectable levels was *TIM3* (receptor)/*GAL9* (ligand) (Figure 6-12).

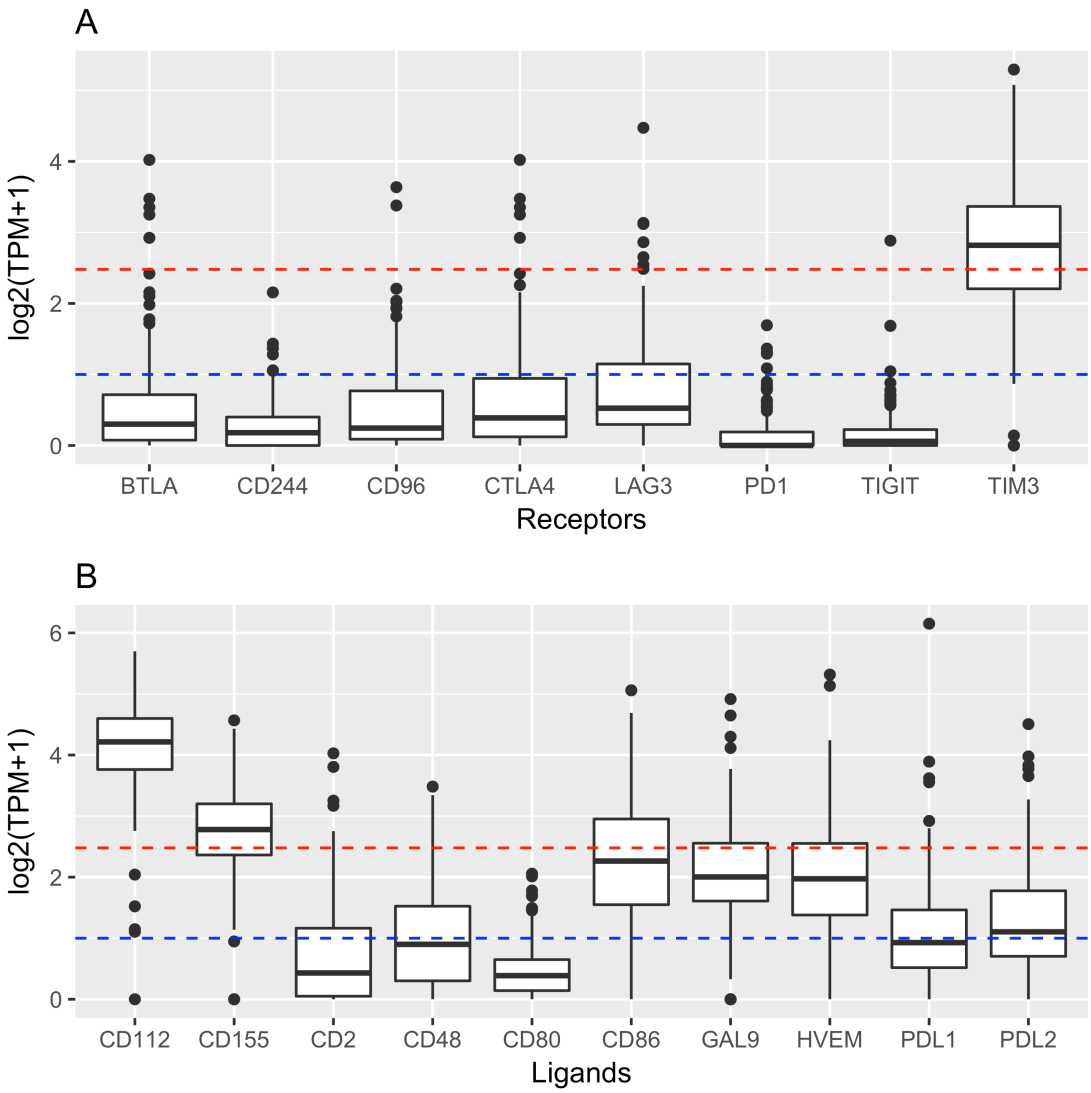


Figure 6-12: Bar plots of levels of immune checkpoint receptors and ligands across all samples in both ependymoma datasets (n=151). (A) Receptor molecules expressed by immune cells. (B) Ligands/antigens expressed by tumour cells. Red hashed line: Median gene expression level for all genes with median expression above 0 TPM in the dataset (n=26038). Blue hashed line: One count per million mapped reads; the level considered to represent expression of the gene.

In light of a recent study indicating that EPN_RELA tumours may be enriched for *PDL1* (Witt et al., 2018), further analysis was conducted to see whether there were differences in immune checkpoint ligand and receptor expression between tumours in PF versus ST locations or for tumours with EPN_PFA versus EPN_RELA DNA methylation classifications. In view of the almost absent nature of EPN_RELA class predictions in the FF samples this analysis was based on the FFPE tumour set only. 12 EPN_RELA tumours and 45 EPN_PFA tumours were included.

One receptor, *TIM3*, was significantly enriched in PF compared to the ST tumours and in EPN_PFA compared to EPN_RELA tumours (Fold change 1.8, p<0.001 and fold change 1.7, p=0.036 respectively). One ligand, *CD112*, was significantly enriched in PF compared to the ST tumours and in EPN_PFA compared to EPN_RELA tumours (Fold change 1.5, p=0.001 and fold change 1.6, p=0.008 respectively). Notably, there were no other significant differences based on location or DNA methylation group. In particular, no differences were seen for *PD1* (p=0.090 and p=0.775 respectively), *PDL1* (p=0.460 and p=0.352 respectively) or *PDL2* (p=0.934 and p=0.194 respectively).

6.3.9 Cytolytic activity (CYT)

A CYT score was generated for each of the 151 tumour samples. The median score was 0.40 TPM (range 0-3.25). There was no significant difference in the level of cytolytic activity between molecular subtypes, all primary and recurrent pairs, or primary and recurrent pairs stratified by molecular subgroup (PF1, PF2, ST).

6.3.10 CT antigens

In assessing CTA expression across all tumours, PF1 and PF2 molecular subgroups were noted to exhibit different profiles. The PF2 tumours significantly overexpressed a greater number of CTAs compared to PF1 (22/276 versus 8/276, p=0.013). In order to establish whether the differentially expressed CTAs identified in PF1 and PF2 were associated with evidence of being able to generate an immune response, the CT-database was consulted. 18/30 of the CTAs identified had been associated with an immune response in at least one cancer type (Table 6-19) (Almeida et al., 2009).

In the PF1 group, two CTAs were significantly upregulated at recurrence in both FF and FFPE datasets; *ANKRD45* and *CTNNA2*. Neither of these genes have ever been experimentally associated with the ability to generate an immune response (Almeida et al., 2009). In the PF2 groups no CTAs were upregulated at recurrence. Additionally, no CTAs underwent significant changes at recurrence in the ST group.

CT-Antigen	Expression	PF Group	Immune Response
SPAG8	Testis	2	Humoral
SPEF2	Testis-Brain	2	No
TSGA10	Testis	2	Humoral
ARMC3	Testis	2	Humoral
TMEM108	Testis-Brain	2	No
ANKRD45	Testis	2	No
RGS22	Unknown	2	No
SPA17	Testis-Brain	2	Cellular
CCDC33	Testis	2	No
SPAG17	Unknown	2	Humoral
PTPN20	Unknown	2	No
SPAG1	Unknown	2	Humoral
ODF2	Unknown	2	Humoral
CCDC110	Testis	2	Humoral, Cellular
SPAG6	Testis	2	Humoral
CTNNA2	Testis-Brain	2	No
CCNA1	Unknown	2	Cellular
CEP290	Testis	2	Humoral
ZNF165	Testis	2	Humoral
KIF20B	Testis	1	Cellular, Induced
IGSF11	Testis-Brain	2	Cellular
TTK	Testis	1	Cellular, Induced
MAEL	Testis	2	No
AKAP3	Testis	2	No
CEP55	Testis	1	Cellular
KIF2C	Unknown	1	Cellular
OIP5	Testis	1	No
NUF2	Testis	1	Cellular
ATAD2	Unknown	1	No
PBK	Testis	1	No

Table 6-19: Table of CTAs expressed in Posterior Fossa Ependymoma. Expression refers to whether the expression is restricted to testis, testis and brain or unknown. PF group (PF1 or PF2) indicates which group the gene is significantly enriched in when comparing the two groups. Expression and immune response data derived from (Almeida et al., 2009).

6.4 Discussion

6.4.1 Aims

This chapter aimed to:

- Compare gene expression at primary diagnosis compared to matched recurrence across FF and FFPE datasets based on:
    - All tumours combined;
    - Location;
    - Molecular classification.
- Consider the impact of therapy on differential expression at recurrence;
- Assess the role of the immune system in primary and recurrent ependymoma using predefined scoring systems (immunophenoscore and cytolytic activity score) plus consideration of CTA expression.

6.4.2 Merits of the study design

By using RNA sequencing, more biological replicates than previous authors, and a paired study design; this study attempted to provide an in-depth analysis of biological functioning in recurrent paediatric ependymoma. RNA sequencing profiles low abundance transcripts with greater accuracy than microarray studies (Wang et al., 2014), potentially making this investigation more detailed than previous research. However, some argue that the most effective way to increase the power of a study is to use more biological replicates (Yuwen Liu et al., 2014), which was achieved by including FFPE samples. Paired analysis not only increases statistical power, but also minimises variation between individuals to give a clearer indication of genuine biological change (Peyre et al., 2010). This was particularly important when cohort sizes were limited by availability of specimens.

A further advantage was the ability to correlate gene expression profiling with DNA methylation profiles, thus providing a way to link the findings to the developing 'molecular era' of paediatric brain tumour research (Mack and Taylor, 2017).

One of the challenges presented by this design was the generation of extremely large datasets. Whilst the analysis aimed to focus on the changes that were most consistent across multiple tumour subtypes and tissue cohorts, there are still likely to be more conclusions that can be drawn from further analysis of this dataset.

6.4.3 Summary of key findings

Analysis of molecular subgroups proved more informative than analysis of the combined dataset. PF1 tumours were associated with downregulation of the innate immune and inflammatory responses at first recurrence, whilst PF2 and ST tumours developed adaptive immune responses. All groups exhibited changes in the type I IFN pathway in at least one dataset. There was some evidence that radiotherapy may be associated with upregulated immune responses at recurrence. Immunophenoscores varied between PF and ST tumours, but cytolytic activity and CTA expression were low throughout (Table 6-20).

	PF1 tumours	**PF2 tumours**	**ST tumours**
FFPE Samples	13	3	8 (5 EPN_RELA)
FF Samples	9	11	0
Upregulated ontologies	Radiotherapy – Type I IFN pathway including IFN related DNA damage signature (IRDS) genes. No radiotherapy – None	Adaptive immune (T and B cell functions). Extracellular matrix and adhesion. IFN pathways.	Adaptive immune response including T and B cells related functions. Type I IFN pathway.
Downregulated ontologies	Innate immune and inflammatory. Chemokines and chemotaxis.	None	Glycolytic and metabolic.
Key (immune) mediators	Up (in radiotherapy treated FF tumours): *OAS1, OAS2, OAS3, BST2, RSAD2, MX1, CXCL12, IFI44, IFI44L, CX3CR1*. Down in all: *IL1R1, IL1R2, IL11RA, CXCL1, CXCL3, CXCL8, CCL2.*	Up: *OAS1, OAS2, OAS3, BST2, RSAD2, MX1, CXCL12*.	Up: Interleukins 1,2,4,6,8,10,12. NF-kB pathway.
Immune Response	Innate	Adaptive	Adaptive
Immunophenoscore	Highest MHC enriched EC enriched	Intermediate SC enriched	Lowest MHC depleted
Cytolytic Activity	Low	Low	Low
CTA Expression	*CTNNA2, ANKRD45*	Nil	Nil

Table 6-20: Summary of the key up and down regulated ontologies and mediators mediators associated with change at recurrence in the different molecular subgroups investigated with paired differential expression analyses.

6.4.4 Overall and location based analyses

Differential expression analysis of all tumour pairs, irrespective of location or molecular classification, identified few shared differentially expressed genes at first recurrence (*EMILIN3*, *ASIC4* and *LHFPL3*). *EMILIN3* is an extracellular matrix molecule which plays a role in skin development in mammals (Corallo et al., 2017) and has been associated with notochord development in zebrafish (Corallo et al., 2013). *ASIC4* encodes for an acid sensing ion-channel involved in synaptic signalling, nociception and mechanoperception (Brown et al., 2015). *LHFPL3* encodes for a tetraspan transmembrane protein associated with lipoma and deafness (Brown et al., 2015). Individually these genes provided little insight into functional changes at relapse; recurrence is likely to result from the interaction of multiple functions or pathways rather than single genes. However, the identification of a change in *EMILIN3* is consistent with changes in extracellular matrix ontologies, which were identified as being altered at first recurrence across all tumour types.

Interestingly, when examining shared gene ontologies, immune and inflammatory responses were upregulated at first recurrence across all tumours. It was difficult to discern further details about the nature of these responses, as these terms were later identified to be associated with the various molecular subgroups. This global overview was therefore unlikely to be helpful in devising future research and treatment strategies.

Tumours with ST location were associated with the upregulation of a number of immune related ontologies and downregulation of genes associated with metabolic pathways. Unfortunately, due to lack of FF ST tumours, it was not possible to examine a comparison cohort for the FFPE ST tumours. It was therefore unclear whether these results were representative of all ST ependymomas, especially given that this group contained a number of molecular diagnoses. The genes differentially expressed were largely distinct from the only other study which analysed relapse in ST tumours (Peyre et al., 2010). However, the study by Peyre and colleagues provided supplementary data which identified 'immune response' and 'immunological disease' as significantly enriched ontologies.

Three terms related to the Type I IFN response were upregulated in the ST GOrilla analysis; GO:0032481 – 'Positive regulation of type I interferon production' (FDR=0.008), GO:0032479 – 'Regulation of type I interferon

production' (FDR=0.012) and GO:0060337 – 'Type I interferon signalling pathway' (FDR=0.012). In addition to this, GO:0032727 – 'Positive regulation of interferon alpha production' was upregulated in the GSEA analysis (FDR=0.037). IFN alpha forms part of the type I IFN signalling pathway and consequently the results from these two separate analyses were consistent.

Multiple interleukin related ontologies were upregulated at first recurrence in the FFPE ST cohort. Interleukins are cytokines involved in modulating the immune system. They are able to act upon T- and B-cells, encouraging their development and contributing to both adaptive immune responses and inflammation (Brocker et al., 2010). They have been shown to be associated with the activation of NF-kB and MAPK pathways in GBM, with a consequent increase in tumour growth and proliferation (Yeung et al., 2013). Therefore, it was important to note that a number of NF-kB related ontologies were also upregulated at ST tumour recurrence. Further areas of investigation include whether targeting downstream pathways of interleukin and NF-kB signalling may have a role in therapy. This is particularly important in light of the association between ST ependymomas with the *C11orf95-RelA* fusion gene and NF-kB signalling (Parker et al., 2014; Pietsch et al., 2014).

Ontologies associated with energy metabolism, particularly glycolysis, were downregulated at first recurrence in the ST tumours. It has long been suggested that cancer cells are able to modulate their metabolic activity to provide a survival advantage in energy restricted surroundings (Warburg, 1956). In fact, 'Reprogramming Energy Metabolism' has been added as an emerging 'Hallmark of Cancer' alongside 'Evading immune destruction' (Hanahan and Weinberg, 2011). Given the small sample size, further validation is needed in an independent dataset regarding metabolic changes at recurrence.

The FFPE and FF posterior fossa location cohorts produced disparate results. The FFPE posterior fossa dataset was associated with a significant downregulation of immune ontologies and the FF posterior fossa dataset was associated with upregulation of immune and inflammatory ontologies. As there was a difference in the proportion receiving radiotherapy between these two groups (patients were more likely to have received radiotherapy in the FF cohort, p=0.037), it would be tempting to conclude that radiotherapy was the factor responsible. Indeed, this would be a biologically plausible explanation; radiotherapy has been associated with the ability to stimulate an immune response by exposing tumour antigens to

antigen presenting cells (Park et al., 2014). However, when the molecular composition of the PF location cohort was analysed in more depth, it became apparent that there was a difference between FF and FFPE cohorts. Whilst both datasets contained predominantly EPN_PFA tumours by DNA methylation profiling (FF 100%, FFPE 93%), the FF dataset had a greater proportion of tumours in the PF2 gene expression subgroup (FF 55%, FFPE 18%, p=0.041). PF1 and PF2 subgroups also demonstrated contrasting immune responses at recurrence. It is, therefore, probable that the difference in expression patterns for the FFPE and FF posterior fossa location cohorts was due to the different composition of molecular diagnoses. This finding supports two conclusions. Firstly, the use of tumour location alone to predict clinical behaviour, including recurrence pattern, is inadequate. Secondly, not all EPN_PFA tumours, predicted by DNA methylation classification, exhibit the same biological behaviour at primary presentation and first recurrence.

6.4.5 Molecular subgroup analyses

This is believed to be the first study to profile paired primary and recurrent tumours with known EPN_RELA DNA methylation class predictions. However, the analysis was based on a limited dataset of just five tumours, so findings must be interpreted with caution and validated in other studies. EPN_RELA tumours demonstrated similar changes in gene ontology to those seen in the PF2 tumours; namely the upregulation of T- and B-cell related functions, indicating the presence of an adaptive immune response. This may suggest shared mechanisms between these tumour classifications at recurrence. It is not clear whether this adaptive response is pro- or anti-tumour and further investigation is required.

Unsurprisingly, given that EPN_RELA made up 5/8 (63%) of the ST tumours, there was a significant overlap with the ontologies expressed in the overall ST location cohort, particularly related to immune functioning. However, the association with terms related to NF-kB signalling was not seen in this cohort, which was unexpected given the link between NF-kB and EPN_RELA tumours in the published literature (Parker et al., 2014; Pietsch et al., 2014).

A number of downregulated terms exclusive to EPN_RELA tumours were related to transport of organic acids, phospholipids and lipids. These may link to changes associated with metabolic pathways seen at recurrence in the ST location cohort, but again need further confirmation.

Downregulation of immune and inflammatory ontologies at recurrence were seen in PF1 tumours in both FFPE and FF cohorts, irrespective of treatment with radiotherapy. This indicates that therapeutic intervention may have minimal impact on tumour biology. The multicentre origin of samples in these cohorts suggests that this finding can be generalised. Whilst this immune and inflammatory downregulation has been described in a small, single centre, cohort of seven patients (Hoffman et al., 2014a), this is the first time that this phenomenon has been confirmed by an alternative technique (RNA-seq). This approach also added more detailed information about the nature of the downregulated ontologies, including specific changes in cell taxis and cytokine release.

Cytokines are proteins with a role in cell signalling. It was evident from the gene expression analysis that there were differences in cytokine distribution between primary and matched recurrence within molecular subgroups.

CCL2, *CXCL1*, *CXCL3*, *CXCL6* and *CXCL8* were chemokines representative of the PF1 subgroup downregulated in GSEA at first recurrence. Additionally, *CXCL1* and *CXCL6* were both differentially expressed at the FDR<0.05 level in the RNA-seq meta-analysis of the PF1 subgroups. These are all inflammatory chemokines and are associated with innate immunity and have chemoattractant properties for neutrophils and monocytes (Esche et al., 2005). Therefore, their decrease at first recurrence is consistent with a fall in the innate immune response.

Chemokine release is stimulated by Pattern Recognition Receptors (PRRs) which are able to detect specific Pathogen Associated Molecular Patterns (PAMPs) and Damage Associated Molecular Patterns (DAMPs). The presence of PAMPs in the ependymoma samples in this study, was supported by changes in gene ontology terms such as GO:0009617 – 'Response to bacterium'. PAMPs can take a variety of formats, including bacterial products and hypomethylated DNA containing CpG motifs (Esche et al., 2005). This is interesting given that, in addition to reports of large areas of hypermethylated CpG islands (Mack et al., 2014), EPN_PFA has also been reported to be globally hypomethylated (Bayliss et al., 2016). A hypothesis is that hypomethylated DNA may act as a PAMP in ependymoma. Numerous cancer associated molecules can serve as DAMPs including: heat shock proteins (HSPs); high-mobility group box-1 protein (HMGB1); adenosine triphosphate (ATP); and extracellular matrix danger molecules such as the S100 proteins, hyaluronan, heparan sulfate proteins and fibronectin (Liu and Zeng,

2012; Schaefer, 2010). One explanation for the fall in chemokine response is that PAMPs and DAMPs are less evident in PF1 tumours at recurrence. A protein level analysis would be important to investigate this theory.

A further hypothesis for the downregulated immune and inflammatory response, in the PF1 cohort is that the tumours have undergone an immunoselection process resulting in tumour immune escape (Zitvogel et al., 2006). Mechanisms of immune escape can vary depending on the level of T-cell infiltration into the tumour. Tumours with high levels of infiltration can suppress antigen expression and upregulate immune checkpoint markers; tumours with low levels of infiltration can alter the tumour microenvironment to prevent further T-cell recruitment (Spranger, 2016). Therefore, it would be important to establish the level of T-cell involvement in ependymoma, in order to interpret the potential of immune escape further. This is addressed in Chapter 7.

There was an association between radiotherapy and upregulation of the adaptive immune response at first recurrence in the PF1 groups. The FF dataset, in which patients were more likely to have been treated with radiotherapy, exhibited an upregulation of the type I IFN pathway. The FFPE PF1 tumours treated with radiotherapy showed evidence of an adaptive immune response, however, specific terms related to type I IFN were not seen. Possible reasons for this include: type I IFN being genuinely unaffected in the FFPE cohort; the FFPE sample set being underpowered to detect a difference; or an artefactual result in the FF cohort. A technical validation of this finding was therefore performed in the FF samples. Although statistical analysis was precluded due to low sample numbers, this technical validation was supportive of the type I IFN changes given the fold change increases seen in matched RNA-seq and RT-qPCR results. The retrospective nature of this study meant that a causative role should not be assigned to radiotherapy. However, the findings in the PF1 groups add weight to the hypothesis that irradiation following primary disease contributes to the development of an adaptive immune response, which may be associated with type I IFN. Radiotherapy is known to augment the type I IFN response in other cancers (Woo et al., 2015; Zitvogel et al., 2015) so this merits further consideration.

PF2 tumours demonstrated an increase in terms related to IFN signalling and adaptive immunity at first recurrence. In particular there was enrichment of T- and B-cell activation related terms, suggestive of an immunogenic tumour

response. As recurrence still occurred, it is possible that this represented either an incomplete anti-tumour response or a driver of progression. As almost all the PF2 patients were treated with radiotherapy, it was not possible to assess whether radiotherapy may have contributed to this response.

The type I IFN pathway was seen, to some extent, in ST, PF1 and PF2 tumours, suggesting that this may be a shared mechanism at recurrence. The type I IFN response can act as a 'bridge' between an innate immune response and the development of adaptive immunity by priming CD8+ T-cells. Additionally, there is evidence that modulation of the type I IFN pathway is a potential therapeutic approach in cancer (Medrano et al., 2017). To know whether this would be an option in ependymoma, a better understanding of its functional role is required.

In the FF PF1 subgroup the type I IFN response was seen, but with no associated change in B- and T-cell ontologies. This may suggest that, whilst IFN was released, possibly in response to radiotherapy or chemotherapy, the adaptive immune effector cells were either not present or were unable to respond. A lack of T-cell responsiveness has previously been suggested in EPN_PFA ependymoma (Hoffman et al., 2014a; Griesinger et al., 2015).

Upregulation of type I IFN has also been associated with the ability to modulate the host immune response and augment the efficacy of chemotherapy and radiotherapy (Bracci et al., 2017). Whilst a type I IFN response was seen in the PF2, ST and some radiotherapy treated PF1 ependymomas, the fact that all of these tumours still recurred suggests that this response was not fully protective. It may be that the type I IFN response and development of adaptive immunity alone was insufficient to prevent recurrence or may actually have contributed to further tumour progression or therapeutic resistance mechanisms (Dunn et al., 2005). Interestingly, the clinical data analysed in Chapter 3 suggested that treatment with radiotherapy was associated with delayed time to progression but not OS. One possible explanation for this is a therapy induced immunological response which is eventually overwhelmed by the tumour.

There is emerging evidence that, in some situations, the IFN response may be immunosuppressive (Bracci et al., 2017; Medrano et al., 2017; Minn, 2015). A recent study in GBM indicated that constitutive type I IFN signalling may contribute to immune escape mechanisms (Silginer et al., 2017). Additionally, constitutive type I IFN signalling has been associated with the development of

resistance to chemotherapy and radiotherapy (Bracci et al., 2017). In a breast cancer study, a set of seven IFN stimulated genes associated with resistance were described; the 'Interferon Related DNA-Damage Signature' (IRDS) (*STAT1*, *MX1*, *ISG15*, *OAS1*, *IFIT1*, *IFIT3* and *IFI44)* (Weichselbaum et al., 2008). On review of the PF1 FF dataset, all seven IRDS genes were upregulated at first recurrence, six of them reaching statistical significance, with *STAT1* having an increased fold change but not reaching significance. These genes were also upregulated at first recurrence in the PF2 and ST datasets; however, fewer reached statistical significance. The role of type I IFN in ependymoma, therefore, remains undefined and further molecular subgroup specific research is required.

Whilst numerous terms relating to the immune and inflammatory responses, chemotaxis and cell signalling at recurrence were identified, a proportion of terms were also related to changes in the extracellular matrix (ECM). ECM components are able to signal through pattern recognition receptors to induce innate immune responses (Jiang et al., 2005). The ECM can bind cellular adhesion molecules and change the ability of leucocytes to freely migrate through a tissue, thus playing a role in immune cell recruitment and immune evasion (Morwood and Nicholson, 2006). The ECM can also play a role in polarisation of T-cells towards a Th1 or Th2 phenotype, thereby affecting immune function (Morwood and Nicholson, 2006). Modulators within the ECM are also able to contribute to the downregulation of inflammation, for example by exerting control over the complement cascade by binding to, and inactivating, elements of the pathway (Groeneveld et al., 2005). Therefore, it is interesting that elements of *C1q*, a recognition molecule for the complement cascade, were downregulated at first recurrence in the PF1 ependymomas.

Other molecules of interest which occurred in both ECM gene ontology sets for the PF1 tumours included *CD44, ITGA5, COL6A2, TGFBI, FERMT1, MMP16, THBS1, ABI3BP, ADAM12, SERPINE1, COL11A1* and *COL27A1*. Downregulation of CD44 molecules has been associated with a decrease in neutrophilic infiltrate in a kidney model (Rouschop, 2005). Therefore, a hypothesis for the decreased innate immune response in PF1 tumours is that ECM modifications contribute to immune evasion.

The follow up of the gene expression data was not primarily focussed on ECM changes, but they are mentioned to highlight a potential role for the ECM in recurrent ependymoma.

6.4.6 Ependymoma immunophenoscores and cytolytic activity

High immunophenoscores are associated with increased immunogenicity and neoantigen load (Charoentong et al., 2017). Tumours with high scores have also been associated with improved responses to *CTLA-4* and *PD1* checkpoint blockade. Clinical benefit to *CTLA-4* blockade was only seen in tumours with an immunophenoscore greater than six, and to *PD1* blockade with a score greater than five (Charoentong et al., 2017). The current study identified extremely low levels of expression of these potential therapeutic targets, suggesting that *CTLA-4* or *PD1* blockade may not be an appropriate strategy for ependymoma. This finding is consistent with another report on ependymoma, examining *PD1* and *PDL1* at the protein level, which identified very low levels of expression (Dumont et al., 2017).

After investigating the checkpoint blockade component of the immunophenoscore, all checkpoint molecules, identified in a review of markers associated with T-cell anergy, were examined (Catakovic et al., 2017). Consistent with the *PD1* and *CTLA-4* findings, almost all of the checkpoint receptors were present at levels too low to be considered expressed. The only exception to this was *TIM3* and its ligand *GAL9* which were expressed at a median of 2.8 log2(TPM+1) and 2.00 log2(TPM+1). *TIM3* has been suggested as a potential alternative target for checkpoint inhibition therapy (Anderson, 2014; Cheng and Ruan, 2015).

Whilst the results of this gene expression analysis suggested that levels of checkpoint molecules in ependymoma may be too low to target with checkpoint blockade, further research is required to confirm the *TIM3*/*GAL9* findings. This should include IHC staining, to confirm protein expression levels, and investigation into the functional role of *TIM3* in suppression of the immune response.

A recent study has suggested that levels of *PDL1* are increased in EPN_RELA tumours compared to other subtypes, making EPN_RELA a potential target for checkpoint blockade (Witt et al., 2018). This was not the experience in this study. In fact, the only checkpoint molecules that showed differences between locations or DNA methylation groups were upregulated in the EPN_PFA tumours (*TIM3* and *CD112*). There could be a number of reasons for this difference, including the specific selection of a cohort of children who recurred as opposed to a cohort of patients with mixed outcomes. However, given this conflicting finding, further

research is warranted. The results from this study did not support the conclusion that one subtype of disease may be more amenable to checkpoint blockade than another.

Although changes in immune genes were present between primary and matched recurrences, the immunophenoscore failed to identify any differences. This may relate to the fact that the immunophenoscore measures specific parameters of the immune response which were not reflected in the primary and recurrent pairs.

An alternative explanation for the lack of difference in immunophenoscore is the possibility that the changes seen in the paired gene expression analyses are a result of immunosuppressive actions of the tumour itself. Work by one research group has suggested that in EPN_PFA tumours, T-cells are anergic and therefore unable to function effectively (Hoffman et al., 2014a; Griesinger et al., 2015). Genes representative of anergic T-cells have been described in an in-vivo mouse model (Zheng et al., 2013). Interestingly, a number of these (*LAG-3*, *CRABP2, NRGN, SEMA7A)* overlapped with enriched genes in the PF1 groups. However, this result was not exclusive to PF1 tumours (EPN_PFA like) as the FFPE PF2 dataset was also enriched for a number of other genes associated with T-cell anergy (*CRTAM* and *NRN1)*. This could either represent an artefact of the unsupervised clustering process or evidence of T-cell exhaustion in both tumour subtypes. There was no significant change in gene markers of T-cell anergy at recurrence in either of the PF datasets.

The immunophenoscore data provided a novel insight into the ST tumours, which appeared to be even less immunogenic than the PF tumours. ST tumours were depleted for the MHC apparatus category, suggesting that one way in which the ST tumours had lower immunogenicity may be in downregulation of their ability to present tumour antigen. These results need to be taken into consideration when planning immunotherapy studies in these patients.

Cytolytic activity (CYT) was found to be very low. This suggests that, across all molecular subtypes and recurrence statuses, cytotoxic T-cells (CTL) and Natural Killer (NK) cells are not undergoing activation to conduct active cell killing. Low CYT has been associated with a lack of neoantigen expression and poorer clinical outcomes, and has been used as a proxy marker for immunogenicity

(Charoentong et al., 2017). The findings are therefore consistent with those for the immunophenoscore.

When comparing the CYT scores for the current ependymoma cohort with CYT scores generated for other cancer types, it was apparent that ependymoma had scores consistent with the least immunogenic tumours. The levels for ependymoma were close to the levels described in 499 glioma samples, obtained from the cancer genome atlas. The levels were also below those described for GBM, which was the tumour with the second lowest CYT scores of any cancer type (Rooney et al., 2015).

The level of ependymoma immunogenicity was assessed using two different, but complementary and previously validated, approaches. This demonstrated that, despite immune gene expression changes at recurrence, there was little change in the already low levels of tumour immunogenicity. On the basis of these findings, interventions targeting immune modulation alone, particularly checkpoint blockade, are unlikely to be successful at either primary presentation or recurrence. Whilst changes in the type I IFN pathway were seen, the end point for any anti-tumour immune effect of this pathway would be the CD8+ mediated destruction of tumour cells. In this dataset, there was no discernible effect on cytolytic activity at the gene expression level to suggest efficacy of the type I IFN response.

6.4.7 Tumour antigens: neoantigens and CT-antigens

One way in which the immune system recognises cancer is via identification of tumour neoantigens. Higher immunophenoscores are associated with increased neoantigen burden (Charoentong et al., 2017). The nature of RNA sequencing meant that it was not possible to perform a direct assessment of the level of tumour neoantigens in these ependymoma samples. RNA-seq can identify single nucleotide polymorphisms (SNPs) that may form the basis of neoantigens. However, this analysis would only provide details about a very biased (coding) region of the genome.

It could be inferred that the relatively low immunophenoscores and lack of significant changes in the score at recurrence suggest that the neoantigen burden is low. This was supported by the lack of detection of recurrent mutations in paediatric ependymomas and the very low mutational burden identified in whole genome and whole exome sequencing of a cohort of 47 PF ependymomas (Mack et al., 2014). However, in order to conclusively evidence any lack of change in

neoantigen burden at first recurrence, a study to perform whole genome sequencing of matched primary and recurrent paediatric ependymoma pairs is required.

CTAs are immunogenic proteins that exhibit a restricted pattern of expression and can be re-expressed in cancer (Almeida et al., 2009). Whilst there was a different pattern of expression between the PF1 and PF2 subgroups, there was very little change in CTA expression at first recurrence for any molecular subgroup. It therefore seems unlikely that this is responsible for the change in immune profiles at recurrence.

6.4.8 Potential implications for future therapies

There has been a great deal of focus on the adaptive immune response in cancer, for example by the use of checkpoint blockade and chimeric antigen receptor (CAR) T-cell therapy. A greater understanding of the innate immune response to cancer is also developing and may lead to other innovative therapeutic approaches (Liu and Zeng, 2012; Woo et al., 2015). Whilst type I IFN signalling appears to be implicated in some way in each of the molecular subgroups, a better understanding of its role and impact on tumour development is needed, given its identification as a bridge between the innate and adaptive immune system (Woo et al., 2015). A recent study began to elucidate the role of type I IFN signalling in glioblastoma (Silginer et al., 2017), but such a specific approach is also required in primary and recurrent ependymoma. Without a better knowledge of how these pathways behave in this particular tumour type, clinical approaches to modulate IFN I signalling, for example by use of Toll Like Receptor ligands or oncolytic viruses (Woo et al., 2015) in patients could be ineffective, and in the worst case, dangerous, by stimulating an unexpected response. Of particular concern for this approach would be the significant increase in IRDS genes in the FF PF1 dataset. Alternatively, a better understanding of IFN type I signalling in ependymoma may suggest that this pathway would be a good target for therapeutic interventions at recurrence.

6.4.9 Conclusions

Through the use of two independent tumour cohorts, this chapter has demonstrated that analysis based on tumour location is unreliable in determining molecular outcomes. The analysis of groups based on gene expression and DNA methylation patterns was more informative. PF1 ependymomas were associated with a downregulation of innate immune and inflammatory responses, associated with possible immune escape. PF2 and ST tumours were associated with the

upregulation of the adaptive immune response but the impact of this was unclear given that these tumours still recurred. Irrespective of the ontological changes at recurrence, immunophenoscores and cytolytic activity were low for all groups, suggesting that ependymoma has low immunogenicity and may not be responsive to immune checkpoint blockade therapies. The type I IFN pathway may represent a potential therapeutic target in recurrent paediatric ependymoma, however further research is required to determine its role.

7 A Study of Infiltrating Immune Cell Populations in Primary and Recurrent Ependymoma

7.1 Introduction

The spatial distribution and levels of immune cells within tumours may be important in aiding the understanding of tumour biology and outcomes. Researchers have described different tumour phenotypes in other cancer types: T-cell inflamed; immune excluded; and immune desert tumours (Chen and Mellman, 2017; Herbst et al., 2014; Kim and Chen, 2016) (Figure 7-1). T-cell inflamed tumours are associated with substantial T-cell infiltration of the tumour parenchyma; immune excluded tumours are associated with the accumulation of cells in the stroma but not the parenchyma; and immune desert tumours have minimal evidence of T-cells in the tumour environment overall. It has been suggested that these phenotypes are associated with various host characteristics, including age and environment, as well as tumour characteristics such as tumour type, chemokine profile and neoantigen burden (Chen and Mellman, 2017). The phenotypes have also been associated with different outcomes, including survival and response to immunotherapy. Different levels of immune cell infiltration may also be associated with different patterns of immune escape, contributing to tumour recurrence (Spranger, 2016). Knowledge of the level of immune and inflammatory cell infiltration and spatial distribution may help to ascertain whether ependymoma fits any of these phenotypes. This may lead to more informed approaches to new therapies and a better understanding of clinical outcomes.

Cell types relevant to a basic understanding of tumour immunity include: CD4+ and CD8+ cells relating to T-cell infiltration; CD20+ cells relating to B-cell infiltration; and CD45+ cells related to infiltration with inflammatory cells. CD4+, CD8+ and CD20+ cells represent the adaptive immune response, whilst CD45+ cells represent the role of the inflammatory and innate immune responses. Details of the function of these cells were reviewed in section 1.5.2.

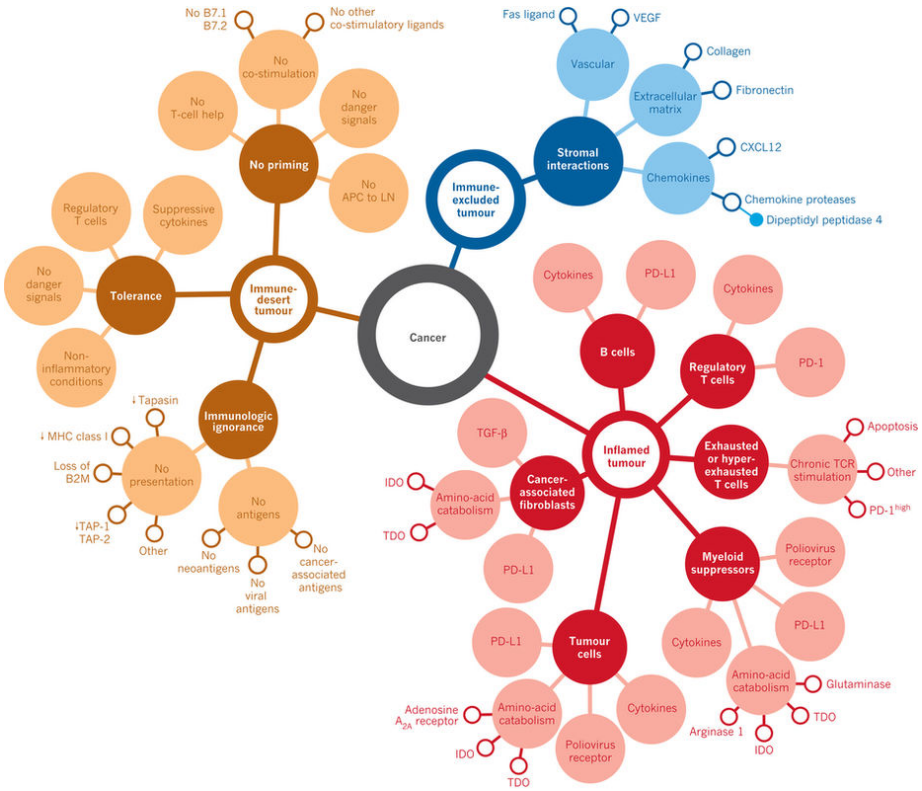


Figure 7-1: Immune phenotypes in cancer. Reproduced with permission from Chen and Mellman, 2017.

Three studies have investigated immune infiltration in ependymoma. One contained nine primary tumours that recurred and ten non-recurrent primary tumours, from a mixture of intracranial locations (Donson et al., 2009). The second contained 13 PF tumours with matched recurrences (Hoffman et al., 2014a). A third, recent study, included a greater number of samples but examined the levels of *PDL1* infiltration in ependymoma specimens, postulating a DNA methylation group based difference (Witt et al., 2018). The samples included in these studies were not independent of one another. The first study found a significant association between CD4+ cell infiltration and recurrence in univariate analysis, but did not comment on the spatial distribution of the immune cells. The second found that tumours in the group equivalent to PF2 had slightly increased levels of CD4+ cells at first recurrence. Associations were not found with CD8+ cells in either study. The third study suggested that EPN_RELA tumours, compared to other molecular subtypes, were enriched for *PDL1* and that more CD4+ and CD8+ cell infiltration was seen in EPN_RELA compared to the PF molecular diagnoses. The study did not specifically look at differences between primary and recurrent disease.

This chapter aimed to describe ependymoma immune cell infiltration in a larger cohort than previous works. Given the RNA sequencing findings in Chapter 6, it was also important to establish whether immune and inflammatory changes were seen at the protein, in addition to the gene, level.

Whilst the samples included in the RNA sequencing analysis were selected to minimise non-tumour cells, they were not laser dissected or subject to cell sorting. Consequently, the gene expression changes may not have been isolated to tumour parenchyma. It was not possible to discern the spatial distribution of immune cells within the tumours from the gene expression data, but this can be established using immunohistochemistry (IHC).

The specific aims of this chapter were to:

(1) Investigate the spatial distribution of immune and inflammatory cells (CD4+, CD8+, CD20+ and CD45+) in all ependymoma types at primary and recurrence using IHC;

(2) Investigate whether tumour location and grade were associated with altered quantities and distribution of immune and inflammatory cells;

(3) Identify whether there were changes in the levels of immune and inflammatory markers within the parenchymal areas at recurrence, compared with paired primary disease;

(4) Assess whether infiltration with immune or inflammatory cells was associated with TTP or OS, in univariate and multivariate analyses, in recurrent paediatric ependymoma.

7.2 Materials and methods

Samples from all tumour locations at both primary and recurrence were used to investigate tumour infiltration with CD8+, CD4+, CD20+ and CD45+ cells. The methods were designed in discussion with the authors of the two aforementioned studies in order to co-ordinate future collaborations in this area, however, the samples were new (Donson et al., 2009; Hoffman et al., 2014a).

7.2.1 Tissue sections

A cohort of 59 primary tumours, 55 with matched recurrences, was established (Appendix 1). The optimal approach for IHC projects of this size is to use tissue microarrays consisting of multiple cores of tumour on one microscope slide. This approach was attempted with the cluster of differentiation (CD) marker antibodies, but immune cells were present in insufficient quantities for reliable

identification. Consequently, IHC was performed across whole tumour sections to gain a more representative view. To ameliorate run to run variability, paired primary and recurrent tumours were stained within the same IHC run.

7.2.2 Immunohistochemistry

The details of the antibodies used to identify cellular infiltrates within the specimens are included in Table 7-1. The IHC reagents were from the EnVision Detection System, Peroxidase/DAB, Rabbit/Mouse (DAKO, UK).

Marker	Clone	Manufacturer	Concentration	Function
CD45	2B11+PD7/26	DAKO (M0701)	1/100	Common leucocyte antigen. Inflammatory marker.
CD20	L26	DAKO (M0755)	1/200	Humoral immunity. Adaptive immunity.
CD8	C8/144B	DAKO (M7103)	1/100	Cell killing. Adaptive immunity.
CD4	SP35	Cell Marque (104R-18)	Pre-diluted	T-helper cells. Adaptive immunity.

Table 7-1: List of antibodies used in the IHC analysis along with functional properties of the cells they represent.

5 μm tissue sections were placed in a rack and rehydrated through xylene (15 minutes), 100% ethanol (10 minutes) and 95% ethanol (10 minutes) before being washed with tap water. Antigen retrieval involved placing the rack in a pre-heated steamer, in a sodium citrate buffer (Table 7-2), at pH six for 40 minutes.

Reagent	Amount
Water	5 L
Sodium Citrate Monohydrate	10.5 g
2M Hydrochloric Acid	65 ml

Table 7-2: Reagents for the sodium citrate buffer used in antigen retrieval.

Following antigen retrieval, slides were cooled in the steamer for 10 minutes before a two minute wash in Phosphate Buffered Saline (PBS).

Slides were wiped with a tissue before a hydrophobic pen was used to draw around each section to contain the liquid used in subsequent steps. Peroxidase

blocking solution was then added for five minutes before a further five minute PBS wash. After blotting with a tissue, diluted primary antibody was added in a sufficient quantity to cover the tissue section before being incubated for one hour at room temperature in a humidified chamber. Slides were washed again in PBS for five minutes before further blotting with a tissue and the addition of secondary antibody. Sections were incubated for a further 30 minutes at room temperature before another five minute PBS wash. 3,3'-diaminobenzidine (DAB) solution (1 part DAB to 49 parts substrate buffer) was added to the slides and incubated for five minutes before being washed off with tap water. Slides were then de-hydrated through 95% ethanol, 100% ethanol and finally xylene before mounting with DPX (Sigma-Aldrich, UK).

Optimal primary antibody concentrations were determined by testing a range of dilutions on positive control tissue, to identify the concentration giving the most specific staining and least background. Positive controls were selected by consulting the Human Protein Atlas database (www.proteinatlas.org). The opinion of a neuropathologist (Dr Simon Paine) was also sought.

7.2.3 Scoring tumour sections

Photographs of 20 high power fields (HPFs), at 400x magnification, were selected at random from each section for CD4, CD8 and CD20 (Olympus, UK). 10 HPFs were selected for CD45. Attention was given to ensuring fields were taken from areas of tumour parenchyma, as opposed to surrounding stroma. For all markers, 10 HPFs of non-parenchymal areas were taken in order to compare with tumour parenchyma. Positive cells exhibited circumferential brown staining around a cell nucleus and cytoplasm (Figure 7-2). To ascertain levels of cellular infiltrate, HPFs were scored by the author counting all positive cells, and a subset were double scored independently by another individual (SLB, FF), to ensure inter-scorer reliability. The IHC, photography and scoring were performed blind to clinical parameters. The order of scoring was determined by randomly selecting photographs of HPFs.

7.2.4 Data analysis

Data was analysed within the R statistical environment (R Core Team, 2014). Matched primary and recurrent pairs and parenchymal versus non-parenchymal tumour areas were compared using the Wilcoxon signed-rank test. Associations between immune cell infiltration of the primary tumour and survival outcomes were analysed using the approaches to survival analysis described in Chapter 3.2.

7.3 Results

7.3.1 IHC cohort clinical features

The IHC cohort contained 55 patients with primary and matched recurrent specimens. An additional four relapsed patients had primary tissue available and were included in survival analyses. DNA methylation group was unavailable for 13 (22%) patients. 30 (51%) primary tumours were EPN_PFA, 11 (19%) EPN_RELA, 1 (2%) EPN_PFB and 2 (3%) EPN_YAP. Other diagnoses accounted for 2 (3%) of the primary tumours (1 EPN_MPE, 1 HGNET_MN1). Other than the IHC cohort being less likely to have been treated with radiotherapy (p=0.022), there were no differences to the clinical cohort described in Chapter 3 (Table 7-3).

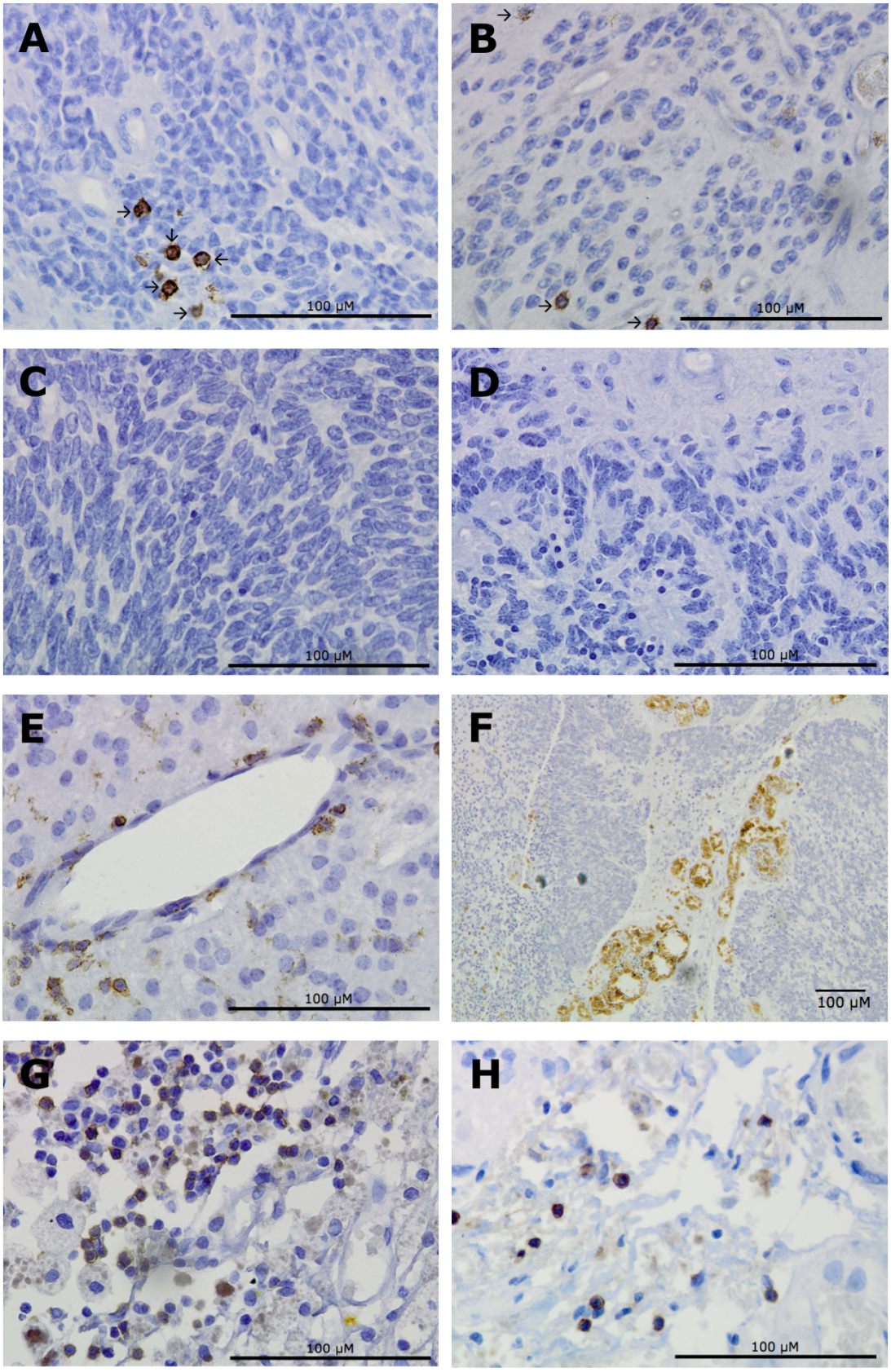
Parameter		**Clinical Cohort (n=188)**		**IHC Cohort (n=59)**		**P Value**
		Number	%	Number	%	
Age	<3 years	94	51	27	47	0.762
	3+ years	91	49	30	53	
	NK	3	-	2	-	
Gender	Male	105	58	30	56	0.876
	Female	77	42	24	44	
	NK	6	-	5	-	
Extent of Resection	GTR	76	45	25	46	0.877
	STR	93	55	29	54	
	NK	19	-	5	-	
Location	PF	136	73	39	66	0.233
	ST	44	23	19	32	
	SP	7	4	1	2	
	NK	1	-	-	-	
Grade	WHO II	85	52	27	47	0.541
	WHO III	78	48	31	53	
	NK	25	-	1	-	
Radiotherapy at diagnosis	Yes	104	59	23	41	0.022
	No	73	41	33	59	
	NK	11	-	3	-	
Median age		35 months		40 months		0.574
Median TTP		17 months		17 months		0.561
Median OS		61 months		80 months		0.267

Table 7-3: Clinical features of the IHC cohort compared to the overall clinical cohort described in Chapter 3. The only significant difference identified was an increased likelihood of treatment with radiotherapy in the IHC cohort.

7.3.2 Visual appearances of cellular markers and initial observations

For all markers, positive cells were indicated by a 'ring' of brown staining around an unstained, blue centre. For all markers, infiltration into the tumour parenchyma appeared subjectively less than infiltration into the non-parenchymal and stromal areas. In particular, many samples showed increased evidence of infiltration within and around tumour vasculature, but very little positivity elsewhere in the specimen. For CD4, CD8 and CD20 markers, there were large areas of absent staining within parenchymal areas; infiltrating immune cells were sparse. CD45 appeared to have greater levels of infiltration in parenchymal areas. This pattern appeared to be consistent in both primary and recurrent tumours. Examples of positive staining, negative staining and staining in different tumour areas are provided in Figure 7-2 (A)–(H).

Figure 7-2: (On next page) Representative images of IHC results. Positive cells were delineated by brown circumferential staining around a blue nucleus and cytoplasm. Examples within tumour parenchyma are given for CD20 (A) and CD8 (B) but CD4 and CD45 positive cells had identical appearances. Negative parenchymal areas demonstrated blue cell nuclei and cytoplasm with no brown markings (C) and (D). Assessment of spatial distribution of all markers indicated increased numbers of positive cells around blood vessels, here demonstrated for CD20 (E) and in stromal and other non-parenchymal areas compared to tumour parenchyma, here demonstrated for CD45 (F). Higher numbers of positive cells were also seen in areas of blood and necrosis. Examples provided for CD8 (G) and CD4 (H). (A-E), (G) and (H) at 400x magnification, (F) at 100x magnification.


A
100 µM
B
100 µM
C
100 µM
D
100 µM
E
100 µM
F
100 µM
G
100 µM
H
100 µM

7.3.3 Average levels of expression in tumour parenchyma

Numbers of CD20+, CD8+ and CD4+ cells were low in the primary tumour parenchyma with medians of zero (range 0-20), one (range 0-105) and zero (range 0-11) respectively. Numbers of CD45+ cells were higher, with a median of 19 (range 0-119) (Figure 7-3).

7.3.4 Spatial distribution of immune cells

There were significantly lower numbers of all cell markers in the parenchymal areas compared to the non-parenchymal areas across all primaries and first recurrences (Figure 7-3).

For the primary tumours, the median number of cells infiltrating parenchymal areas compared to non-parenchymal areas was 19 versus 70 for CD45+ ($p<0.001$), 0 versus 30 for CD20+ ($p<0.001$), 1 versus 16 for CD8+ ($p<0.001$) and 0 versus 13 for CD4+ ($p<0.001$). For the first recurrent tumours, the median number of cells infiltrating parenchymal areas compared to non-parenchymal areas was 10 versus 57 for CD45+ ($p<0.001$), 0 versus 21 for CD20+ ($p<0.001$), 1 versus 16 for CD8+ ($p<0.001$) and 0 versus 8 for CD4+ ($p<0.001$).

When subdivided based on location (PF and ST) and then DNA methylation subgroup (EPN_PFA and EPN_RELA), there remained statistically significantly more immune cells in the non-parenchymal areas than parenchymal areas for every cell marker in every location or DNA methylation defined subgroup, suggesting a generalisable effect across ependymoma subtypes.

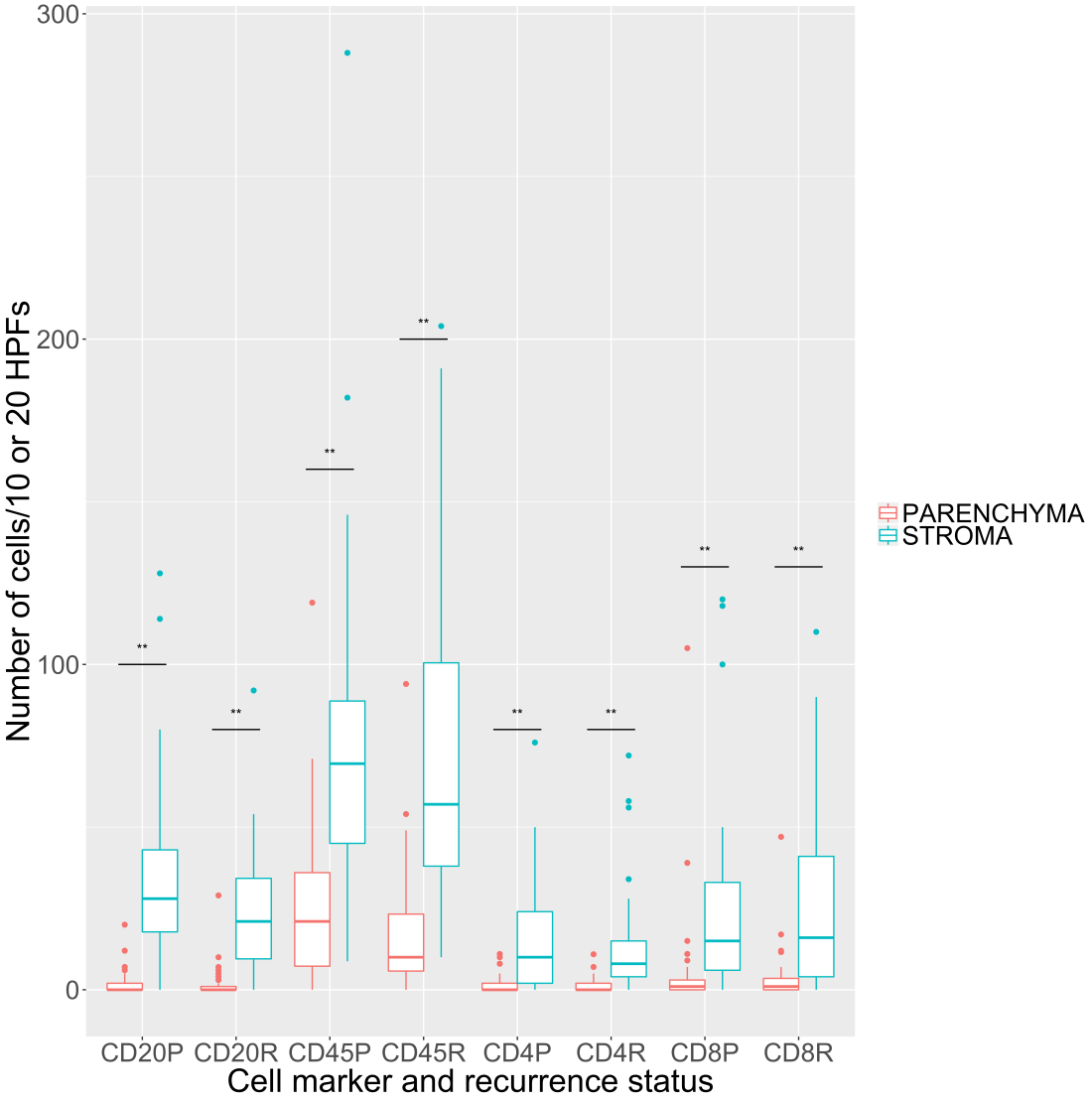


Figure 7-3: Graphical illustration of distribution of immune and inflammatory cells between parenchymal and non-parenchymal tumour areas. For all cellular markers, there were more cells in the tumour non-parenchymal regions than in the tumour parenchyma. Per 20 HPFs for CD4, CD8 and CD20 and 10 for CD45. P=Primary. R=Recurrence.

7.3.5 Clinical features and parenchymal infiltration

In view of the low numbers of CD20+, CD4+ and CD8+ cells in the parenchyma, OS and TTP were analysed based on comparing samples with no expression to samples with any expression. For CD45+ cells, values above and below the median were compared.

Intracranial tumour location, tumour grade, extent of resection and DNA methylation classification (EPN_PFA versus EPN_RELA tumours) were compared with expression of CD4, CD8 or CD20 and CD45 in the parenchyma (Table 7-4). The only statistically significant association found was between positive CD20 expression and higher tumour grade (p=0.041).

	Location			WHO Grade			Resection			DNA Methylation		
	PF	ST	p	II	III	p	GTR	STR	p	PFA	RELA	p
CD45: High	20	7	0.333	12	15	1.000	14	11	0.473	17	4	0.706
CD45: Low	15	11		11	14		10	14		13	5	
CD20: Pos	18	9	1.000	8	19	0.041	10	16	0.404	15	3	0.465
CD20: Neg	19	10		17	11		14	12		15	7	
CD8: Pos	22	12	1.000	17	17	0.560	15	17	1.000	18	7	0.715
CD8: Neg	15	7		8	13		9	11		12	3	
CD4: Pos	19	6	0.224	9	16	0.376	11	11	0.934	17	3	0.273
CD4: Neg	17	13		15	14		13	16		13	7	

Table 7-4: Associations between location, grade, resection status and DNA methylation group for positive and negative expression of CD20, CD4 and CD8 and high or low expression levels for CD45. Numbers in boxes indicate number of samples that were either positive or negative. There was a significant association between CD20 expression and grade III tumours. Primary tumours only. No other associations were identified. Chi-square test used for comparisons.

7.3.6 Cellular infiltration at recurrence

*All locations (*Figure 7-4*A and D)*

Paired analysis identified a significant decrease in the level of parenchymal infiltration with CD45+ cells at first recurrence. The median cell count at primary presentation was 19 cells per 10 HPFs compared to 10 at recurrence (1.9 fold decrease, p=0.002). No change was seen in the non-parenchymal areas (69.5 cells at primary and 57 at recurrence, p=0.703).

CD20+ infiltration in the parenchyma was not associated with a significant change at recurrence (median 0 at primary and recurrence, p=0.408). However, there was a significant fall in CD20+ cells in the non-parenchymal areas (median 28 cells at primary and 21 at recurrence, p=0.034).

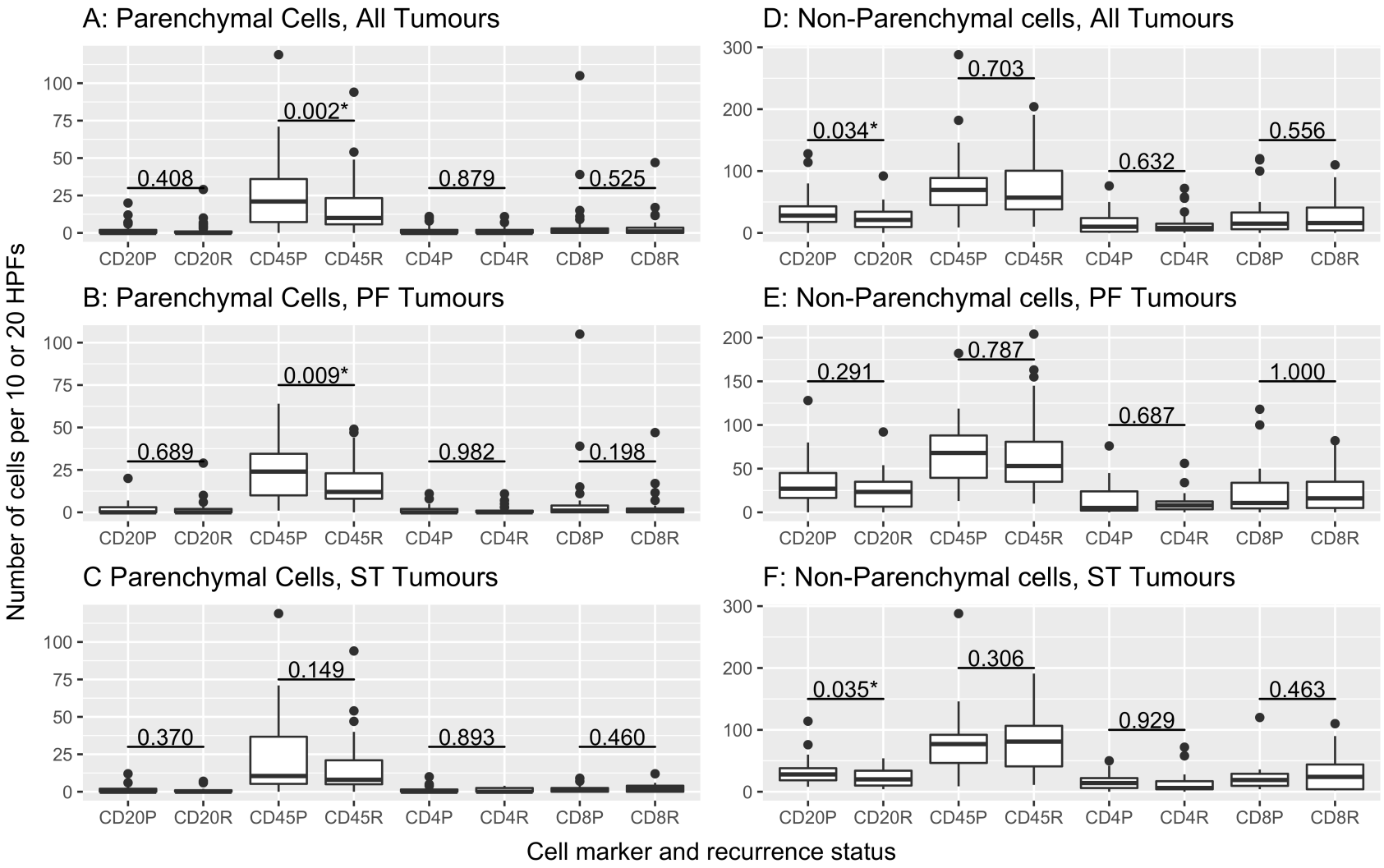
CD8+ and CD4+ cells showed no differences at primary versus first recurrence in both parenchyma and non-parenchyma (CD8+ parenchyma 1 at primary and recurrence p=0.525, non-parenchyma 15 at primary and 16 at recurrence p=0.556. CD4+ parenchyma 0 at primary and recurrence p=0.879, non-parenchyma 10 at primary and 8 at recurrence p=0.632).

*Posterior fossa (*Figure 7-4*B and E)*
When examining the PF tumours as a separate group, there remained a fall in parenchymal CD45+ cells from primary tumour to first recurrence (median 24 to 12 cells, p=0.009). There were no changes in any of the other cell markers in parenchymal or non-parenchymal regions. When examined by DNA methylation classification, there was a significant decrease in CD45+ cells from a median of 26 at primary to 10 at recurrence for the EPN_PFA tumours (p=0.0194).

*Supratentorial (*Figure 7-4*C and F)*
When examining the ST tumours as a separate group, there was a fall in non-parenchymal CD20+ cells from primary tumour to first recurrence (median 28 to 20 cells, p=0.035). There were no changes in any other cell markers in parenchymal or non-parenchymal regions. When examined by DNA methylation classification, no differences were identified for the EPN_RELA tumours.

As a potential association between radiotherapy and immune response was identified in Chapter 6, the immune markers were re-examined after stratification by treatment with radiotherapy. A fall in CD45 levels from primary to recurrence was identified in the parenchyma in the irradiated PF group (p=0.008), but no other significant differences were found between primary tumour and recurrence in irradiated and non-irradiated cohorts for CD4+, CD8+ or CD20+ cells.

Figure 7-4: (On next page). Boxplots illustrating changes in parenchymal immune cell infiltrate of CD20, CD45, CD4 and CD8 at primary (P) and first recurrence (R) across (A) all tumour locations, (B) PF samples and (c) ST samples. Alongside boxplots illustrating changes in non-parenchymal immune cell infiltrate of the same markers at primary and first recurrence across (D) all tumours locations, (E) PF samples, and (F) ST samples. Significant decreases in CD45 were identified at first recurrence in the parenchyma for all and PF tumours. Significant decreases in CD20 were seen at first recurrence in the non-parenchymal areas for all and the ST tumours. Numbers between indicate P-values.


A: Parenchymal Cells, All Tumours
B: Parenchymal Cells, PF Tumours
C Parenchymal Cells, ST Tumours
D: Non-Parenchymal cells, All Tumours
E: Non-Parenchymal cells, PF Tumours
F: Non-Parenchymal cells, ST Tumours
Number of cells per 10 or 20 HPFs
Cell marker and recurrence status
CD20P
CD20R
CD45P
CD45R
CD4P
CD4R
CD8P
CD8R
0.408
0.002*
0.879
0.525
0.689
0.009*
0.982
0.198
0.370
0.149
0.893
0.460
0.034*
0.703
0.632
0.556
0.291
0.787
0.687
1.000
0.035*
0.306
0.929
0.463

7.3.7 Parenchymal infiltration and clinical outcomes

In univariate analysis of all tumour types, a significant association was found between the presence of CD20+ or CD8+ cells in the parenchyma and poorer OS (p=0.007, p=0.009 respectively). No association was found between CD45+ or CD4+ cells and OS (p=0.939, p=0.788) (Figure 7-5). When examining time to first relapse, tumours with more CD20+ cells in the parenchyma recurred more rapidly (p=0.046) but no association was found with any other cellular infiltrate (CD45+ p= 0.947, CD8+ p= 0.797, CD4+=0.086) (Figure 7-6).

To assess whether CD20+ or CD8+ cells maintained their association with poorer OS in multivariate analysis, the clinical parameters tested in Chapter 3 were first tested to assess their significance in this reduced cohort. Out of extent of resection, tumour location, treatment with radiotherapy and tumour grade, only higher tumour grade was associated with significantly poorer OS (Grade II median 152 versus Grade III 47 months, p=0.044). Therefore, CD20+ and CD8+ survival analysis was combined with tumour grade data. The Schoenfeld test was non-significant (p=0.096), suggesting that the CPH test assumptions were met. The only factor which remained significantly associated with poorer OS after this multivariate analysis was infiltration with CD8+ cells. CD20 expression lost its significance, probably because it had a significant association with tumour grade (chi-square test, p=0.041) (Table 7-5).

Factor		Hazard Ratio	95% CI	P-value
Tumour grade	II	1.000	0.853-3.760	0.124
	III	1.790		
CD20 Expression	Yes	1.744	0.823-3.700	0.147
	No	1.000		
CD8 Expression	Yes	2.490	1.087-5.704	0.031
	No	1.000		

Table 7-5: Multivariate analysis of features reaching significance in univariate analysis combined with CD8 and CD20 expression. Only CD8 expression remained significantly associated with poorer OS in this model (p=0.031). 53 were patients included in this analysis.

When assessing time to progression, no clinical factors were significant in the univariate analysis for this cohort. Only CD20+ cells were associated with time in the univariate analysis of the IHC data, therefore no multivariate analysis was performed.

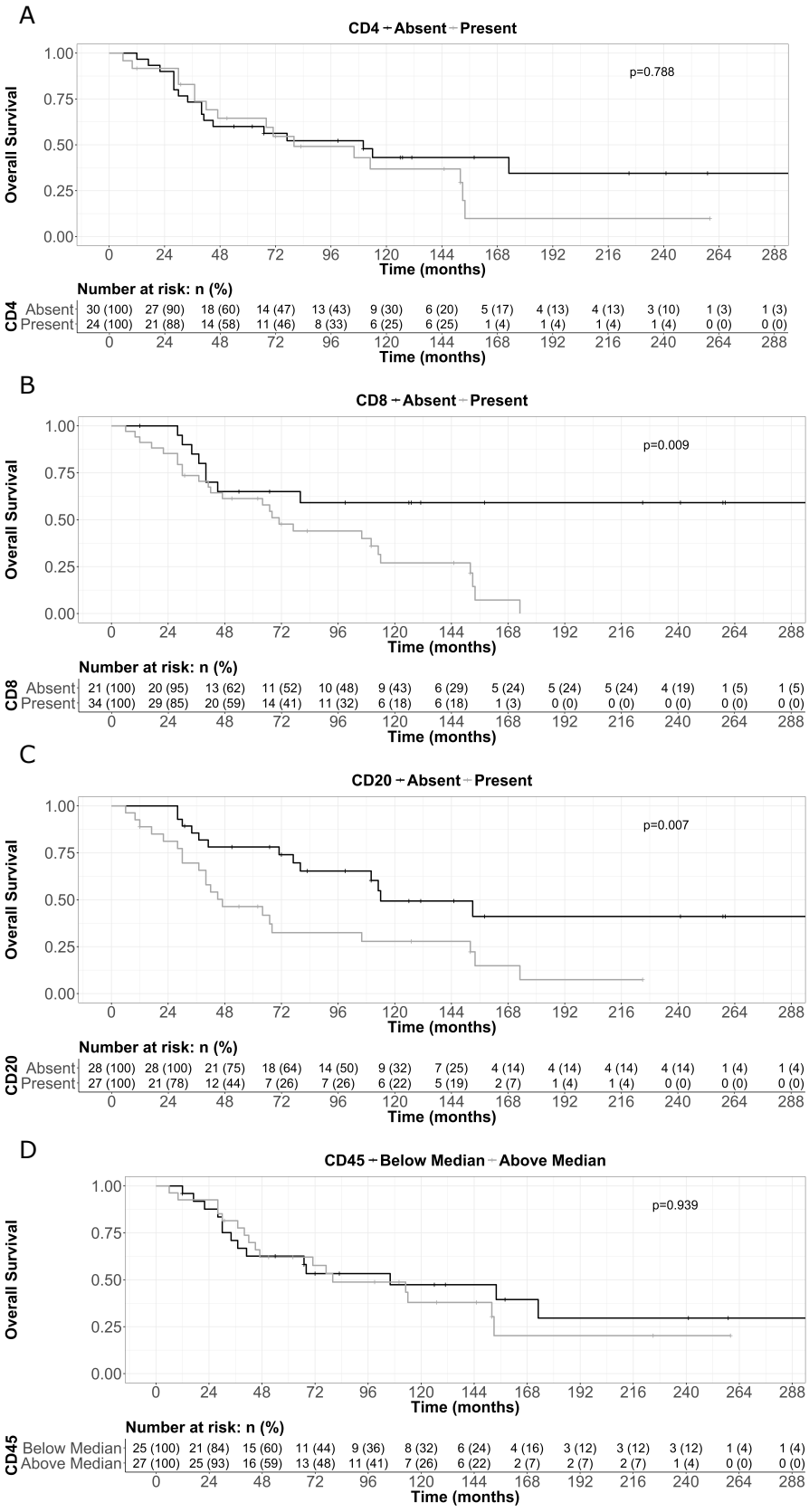


Figure 7-5: Univariate analyses of overall survival for patients with expression of (A) CD4, (B) CD8, (C) CD20 and (D) CD45. Significant associations were identified for CD20 and CD8, both of which were associated with poorer OS.

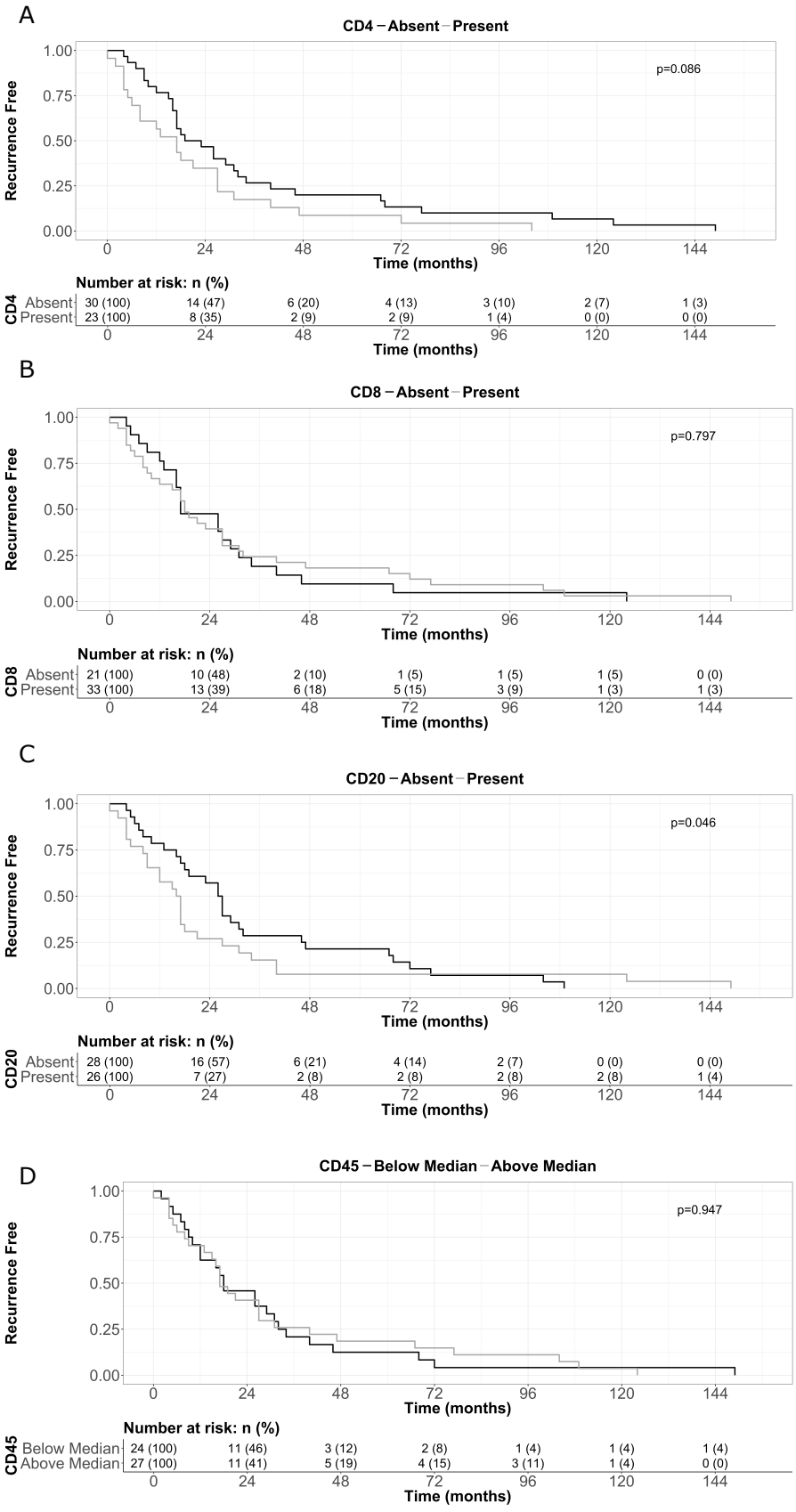


Figure 7-6: Univariate analyses of time to progression for patients with expression of (A) CD4, (B) CD8, (C) CD20 and (D) CD45. CD20 was associated with significantly more rapid progression.

7.4 Discussion

This chapter presented the results of an IHC and clinical analysis of immune cell infiltrate in a cohort of 55 paediatric ependymomas which recurred. The primary aims were to investigate:

(1) The spatial distribution of immune infiltrates;
(2) Whether there was an association between clinical factors and immune cell infiltration;
(3) Whether there was evidence of a change in the level of infiltration at recurrence compared to paired primary tumour;
(4) Whether there was an association between immune cell infiltration of the primary tumour and survival outcomes.

Spatial distribution of cellular infiltrates

Attention was paid to ensure that the counted cellular infiltrate represented regions of tumour parenchyma rather than surrounding stromal, haemorrhagic or necrotic areas. This contrasted with previous studies where this approach was not specified (Donson et al., 2009; Hoffman et al., 2014a). An understanding of how immune cells are distributed across tumour sections is important to be able to interpret how recurrent ependymoma correlates with previously described tumour immune phenotypes (Chen and Mellman, 2017).

For all immune cell markers, levels of tumour parenchymal infiltration were significantly lower than levels in the surrounding non-parenchymal areas. In fact, there were surprisingly few parenchymal CD20+, CD4+ or CD8+ cells, with medians of 0 or 1 cell per 20 HPFs. This suggested that levels of adaptive immune activity within these tumours may be low. This finding was consistent with the data presented regarding cytolytic activity and immunophenoscores in section 6.3. However, it was difficult to compare these results to previous studies in paediatric ependymoma because of the different approaches used in assessing expression levels (Donson et al., 2009; Hoffman et al., 2014a).

In line with a small study investigating PD-L1 expression in ependymoma (Dumont et al., 2017), observation of whole tumour specimens identified infiltrates around tumour vasculature, but minimal infiltration into the tumour itself. The presence of significantly more markers of the immune response outside of, compared to within, the tumour parenchyma, provided evidence to suggest that ependymoma is not a T-cell inflamed tumour, and is more likely to be immune excluded. Given that this pattern was still seen when tumours were

stratified by location and DNA methylation profile, this finding can be extended across the most common intracranial molecular subgroups.

Some may argue that the lack of immune cell markers within the ependymoma specimens is related to a lack of access for immune cells due to the immune privilege of the brain. This is contributed to by the impermeability of the blood brain barrier, lack of lymphatic supply to the parenchyma and lower levels of antigen presenting apparatus (Muldoon et al., 2013). However, this hypothesis is weakened by evidence that in gliomas this immune privilege is disrupted by remodelling of the ECM and vasculature, allowing immune cells to invade (Lee et al., 2009). Questions have also been raised about whether the brain microenvironment is really as immune privileged as previously thought (Carson et al., 2006). It would therefore be erroneous to assume that the lack of immune cell infiltration seen across this ependymoma cohort results from brain immune privilege. Future research questions must address other possible causes of immune cell exclusion from the tumour microenvironment, such as the composition of the ECM and chemokine profiles. There is also a possibility that this finding could represent immune suppression mediated tumour escape, rather than the effect of a competent blood brain barrier (Spranger, 2016); the change in cytokine profile at recurrence in the RNA-seq data may support this theory.

A search of www.clinicaltrials.gov for 'ependymoma' and 'immunotherapy' highlighted eight trials investigating the role of various immunotherapy techniques in brain tumours, including ependymoma. In particular, one study was investigating the role of checkpoint blockade. In an era where immunotherapeutic approaches to ependymoma are being investigated, it is critical to develop a good understanding of the underlying tumour biology. Given that the data from the present study suggests a low number of T-cells within the parenchyma, in addition to low expression of checkpoint molecules and low immunophenoscores (Chapter 6), immune checkpoint blockade may not be beneficial. In fact, it has previously been suggested that lack of T-cell infiltration will result in a lack of effect for checkpoint blockade (Spranger, 2016). The risk of proceeding with checkpoint inhibition trials in ependymoma, without first considering the level of T-cell infiltrate, is that it could be concluded that checkpoint blockade is ineffective, when actually it needs synergy with another treatment to obtain optimal efficacy. One approach to this problem is the use of immunogenic chemotherapy which is being investigated in an attempt to increase levels of T-cell infiltration into tumours (Pfirschke et al., 2016). Also, as this study has only

investigated tumours that go on to recur, it would be important to know more about the infiltration of tumours that do not recur. As many novel therapies are trialled at recurrence, if these non-recurrent tumours were more susceptible to checkpoint blockade, an important treatment opportunity would be missed.

Association between immune infiltrate and clinical factors

It was also important to consider whether baseline characteristics related to tumour grade, location and molecular subgroup were associated with differing levels of parenchymal immune cell infiltrate. The only significant association identified was between tumour grade and CD20+ cell infiltration ($p=0.041$). This must be interpreted with caution given the known difficulty in reaching consensus on designation of grade (Ellison et al., 2011). However, one strength of this study was that most samples underwent central pathology review in addition to local review at the originating centre. Assuming that all of the grade assignments were correct, a possible explanation for the finding is that B-cells may be stimulated by the increased levels of necrosis and inflammation in grade III tumours, resulting in cytokine release and chemotaxis to the affected area. However, if this hypothesis is true, one might expect to see an increase in other adaptive immune and inflammatory cells in grade III tumours, of which there was no evidence in this study.

Given that there were no associations between tumour location and immune cell infiltration, it was unsurprising that the two main, location based, DNA methylation subgroups, EPN_PFA and EPN_RELA, also demonstrated no significant differences in immune cell infiltration. This finding contrasted with a recent study into *PDL1* expression on T-cells in ependymoma, which suggested that EPN_RELA is enriched for this checkpoint marker and also associated with greater levels of CD4+ and CD8+ cells (Witt et al., 2018). However, the approach for determining which tumours had EPN_RELA versus other designations differed; the present study used DNA methylation profiling, whereas the study by Witt and colleagues used gene expression and other molecular profiles, so they are not directly comparable. Therefore, it remains difficult to conclude whether there are differences in the immune environment between intracranial locations and molecular subtypes in ependymoma. Further investigation is warranted to comprehensively delineate the immune environment in the different subgroups of the disease; some research groups have already begun to work towards this aim (Griesinger et al., 2015, 2017; Witt et al., 2018).

Paired analysis of immune infiltration at first recurrence

It was important to assess whether the gene expression changes at recurrence, identified in Chapter 6, were also evident at the protein level. In the PF1 group this may have been indicated by a fall in the level of immune and inflammatory cells. In the PF2, ST and radiotherapy treated groups this may have been indicated by an increase in adaptive immune cell recruitment at recurrence.

The majority of the PF location tumours that had RNA-seq data fell into the PF1 group. In this group, CD45+ cells were significantly reduced at recurrence. This provided initial validation of the changes seen in the RNA-seq data in PF1 tumours; however, it did not provide an explanation as to why this downregulation occurred.

One possibility for this downregulation is a process of tumour immune escape, resulting in the presence of less cellular infiltrate. A number of different immune escape mechanisms have been described in both T-cell inflamed and non T-cell inflamed tumours. Given that the results of this study suggest that ependymoma is not T-cell inflamed, this narrows down potential mechanisms to: a lack of innate immune sensing; failure to recruit effector T-cells; and alterations to the tumour microenvironment (Spranger, 2016). The change in inflammatory chemokine profiles at recurrence, in the PF1 subgroup in the RNA-seq data, provides supportive evidence that microenvironmental changes may be an important mechanism in this response. Additionally, the identification of ECM changes may result in both a lack of innate immune sensing and a failure to recruit effector T-cells. The knowledge about immune cell spatial distribution, combined with gene expression profiling, has therefore helped to narrow down possible avenues of further research into immune escape in ependymoma.

Unfortunately, there were too few PF2 tumours with RNA-seq classifications to perform a full subgroup analysis. This may explain why there were no significant increases at recurrence in the levels of CD20+, CD4+ or CD8+ cells in the PF tumours. However, there was also no evidence that ST tumours or tumours in patients who had received radiotherapy were associated with increased levels of immune infiltration at first recurrence.

It would have been advantageous to be able to include expanded cohorts of PF2 and ST tumours to validate the gene expression results with greater power. It is also possible that the statistical tests were not as reliable in these cohorts

because of the inherent variability in counting low numbers of cells. Although, if a genuine change was missed because of a small effect size between primary and recurrence, one would have to question whether it would be biologically plausible that this could impact overall patient outcomes. An alternative explanation for the lack of change seen is that, whilst the gene expression changes identified via the RNA sequencing were genuine, there was no resultant impact on an effective adaptive immune response. This theory would tend to be supported by the findings of the CYT data which indicated that immune effector activity remained low at recurrence in all tumour subtypes, despite an apparent upregulation of the adaptive immune response in some.

Association between immune infiltrate, survival and tumour progression
Univariate survival analyses associated the presence of CD20+ and CD8+ cells with worse OS in primary ependymomas which recurred. Survival analyses also identified CD20+ cells as being associated with a more rapid first progression. Multivariate analyses resulted in only the presence of CD8+ cells being associated with poorer OS. This finding is perhaps surprising given that CD8+ T-cell infiltration has previously been associated with improved outcomes in a number of meta-analyses in other malignancies (de Ruiter et al., 2017; Mao et al., 2016). However, these analyses did not all indicate whether they were based on multi- or univariate findings and were therefore at risk of confounding.

By excluding patients who did not recur, the present study selected for particularly poor outcomes. Therefore, it is possible that the CD8+ T-cell infiltration in this group of patients, represented a fundamentally different population of cells to those identified in the studies linking them with good prognosis. Immune cells, other than CD8+ cytotoxic T-cells, express CD8; in particular macrophages (Baba et al., 2006) and dendritic cells (Shortman, 2000). There are also a group of CD8+ suppressor cells, which have been linked with poor outcomes. They are associated with a CD8+CD27/CD28- phenotype and have been suggested to suppress anti-tumour responses (Filaci & Suciu-foca 2002, Maybruck et al. 2017). The association between CD8+ cells and poorer outcomes in recurrent ependymoma is a novel finding and as such requires confirmation in independent, prospective studies. Further assessment of molecular markers, including CD27 and CD28, present on the CD8+ cells is also required to establish any evidence of an immunosuppressive, and therefore pro-tumour, function.

Further work is also needed in order to correlate the outcomes of a non-recurrent dataset with the recurrent dataset to assess whether the clinical correlates remain the same.

Conclusions

This chapter has provided evidence that paediatric ependymomas that recur are immune excluded tumours with low levels of T- and B-cell infiltrates at both primary and recurrence. There were few associations with clinical factors other than a significant association between B-cells and tumour grade, which may potentially be explained by higher levels of necrosis. The downregulation of the inflammatory response identified in PF1 tumours in previous chapters was supported by the findings of the IHC analysis. There was no evidence of an upregulated immune response in recurrent ST tumours or those treated with radiotherapy which may relate to a small sample size. There was an apparent association between CD8+ cell infiltration and poorer OS; further investigation is required to assess for additional markers of CD8+ suppressor cells.

8 Final Discussion and Conclusions

This study aimed to undertake molecular profiling of recurrent ependymoma, combined with contemporary clinical data, to better understand recurrence biology and potential therapies. The specific aims were to:

(1) Collate and analyse a cohort of recurrent paediatric ependymoma cases to determine:
   a. Patterns of recurrence in the overall cohort and location determined subgroups;
   b. Factors impacting upon time to first recurrence and overall survival;
   c. Factors impacting upon progression and survival after first recurrence;
   d. Factors affecting risk of recurrence.

(2) Describe the clinical features of a cohort of DNA methylation defined cases which recurred;

(3) Support the results of the clinical analysis and RNA-seq analysis by generating DNA methylation profiles for samples with tissue availability;

(4) Undertake RNA sequencing of FFPE tumour specimens in order to expand the cohort for primary and recurrence analysis and validate the use of this technique against a cohort of FF specimens, in order to make recommendations for future research;

(5) Perform gene expression analysis of matched primary and recurrent pairs to determine changes in expression patterns at recurrence and correlate with molecular classifications where DNA methylation data available;

(6) Validate key expression changes using qPCR and IHC.

This study included a number of novel elements. The large clinical data analysis, supported by DNA methylation results, allowed for the stratification of cohorts of children with recurrence by clinical and molecular profiles. This is believed to be the largest series of children with exclusively recurrent disease presented to date. The gene expression study into matched primary and recurrent ependymoma built on previous findings showing a role for the immune system in this disease (Donson et al., 2009; Hoffman et al., 2014a). However, the use of RNA-seq allowed for a greater depth of analysis and potentially better understanding of how the underlying ontological terms related to tumour function. The use of independent FFPE and FF cohorts increased the statistical power and made the

results more generalisable across molecular cohorts. Combining the gene expression data with DNA methylation profiles enabled a better insight into the molecular subgroups and provided evidence that EPN_PFB expression profiles are not exclusive to EPN_PFB. Data from the DNA methylation profiling also contributed to a larger study delineating new subgroups of EPN_PFA ependymoma (Pajtler et al., 2017). Validation of RNA sequencing from FFPE material was successful; an approach not previously reported in ependymoma and only to a very limited extent in brain tissue (Esteve-Codina et al., 2017). For the first time, this study has indicated that recurrent ependymoma appears to have low immunogenic potential and the use of IHC analyses indicated that it is an immune-excluded tumour. The implications of this for studies into immunotherapy have been considered.

Clinical Features

The clinical analysis highlighted the abysmal medium to long term outcomes for children with recurrent ependymoma. Whilst there had been improvement over the decades, prognosis remained very poor. The research emphasised the clear and pressing need for the development of new and effective therapies for this illness.

The collation of a large, contemporaneous, cohort of clinical data from children with recurrent ependymoma permitted an up-to-date assessment of the clinical features associated with outcome in this devastating disease. The importance of preventing relapse was highlighted by a comparison with a cohort of data from children with isolated primary disease, which clearly demonstrated that recurrence was the most significant known indicator of poor outcomes. Patients with STR compared to GTR were at greater risk of recurrence but crucially, GTR did not prevent some children from recurring. It was evident that in patients who went on to relapse, previously identified prognostic factors at primary diagnosis, such as extent of resection and radiotherapy, did not impact upon OS. There was, however, a transient impact on progression if these interventions were delivered between first and second recurrence, but in the long-term this was lost. This study confirmed a previous suggestion that for some, ependymoma becomes a chronically relapsing disease (Zacharoulis et al., 2010).

The inclusion of 188 patients in the recurrent cohort increased the statistical power over other studies (Antony et al., 2014; Goldwein et al., 1990; Messahel et al., 2009; Zacharoulis et al., 2010). One of the advantages of these greater

numbers was the ability to subdivide the cohort based on intracranial tumour location. This was important as previous authors have indicated that ependymomas in different locations are likely to arise from different progenitor cells (Taylor et al., 2005) and exhibit different driver mutations (Johnson et al., 2010). Interestingly, it emerged that despite different origins, the clinical pattern of behaviour at recurrence was similar for PF and ST tumours, with the exception of statistically significant, but small, differences in time to progression. The exception to this was the spinal tumours which were associated with better outcomes, however the numbers in this analysis were small. The homogeneity of clinical outcomes in this study, despite location based stratification and the different clinical outcomes that have been previously attributed to different molecular subgroups (Pajtler et al., 2015), suggested that molecular classification might be more informative than location based classification.

Combining DNA methylation data with the clinical cohort

The dawning of the 'molecular era' of brain tumour research (Mack and Taylor, 2017) meant that an analysis of clinical data needed to be placed into context with a molecular definition. One of the strengths of this study was that it is one of the first to provide molecular annotations to a cohort specifically designed to analyse recurrent disease.

The findings seen in the PF and ST location clinical analyses were consistent with those for EPN_PFA and EPN_RELA subgroups. EPN_RELA tumours were associated with a more rapid time to first relapse than the EPN_PFA tumours. However, because EPN_PFA formed the majority of the PF cohort and EPN_RELA formed the majority of the ST cohort, it became difficult to discern the behaviour of the smaller molecular subgroups. Contrary to reports of EPN_PFB, EPN_MPE and EPN_YAP being less aggressive and therefore candidates for more conservative therapy (Pajtler et al., 2015; Ramaswamy et al., 2016), this study showed that they did relapse, but numbers were too low to draw robust conclusions regarding further implications for management. Larger collaborative efforts are needed to investigate the behaviour of EPN_PFB, EPN_YAP and EPN_MPE at both primary presentation and recurrence in childhood.

DNA methylation subgroup assignment almost never changed between primary and first recurrence. This confirmed the findings of the original study delineating nine subgroups in ependymoma (Pajtler et al., 2015). The two exceptions were: one tumour which switched from EPN_PFA to DNET; and one which switched from

DNET to EPN_PFA. Unfortunately, sufficient tissue was not available for re-review of these cases, but discussion with collaborators in Heidelberg indicated that this pattern could reflect contamination of the sample with normal brain tissue (Personal communication, Kristian Pajtler, 2017). It was therefore unclear whether the two cases that did switch represented a genuine change in subgroup. The fact that tumours did not change DNA methylation groups at recurrence shows that when patients developed a further tumour, it almost always appeared to be an ependymoma recurrence, rather than a treatment induced malignancy. This was evident even in those cases of very late relapse.

Expanding the role for nucleic acids extracted from FFPE

One of the challenges of this study was obtaining sufficient samples to include in the analysis of paired primary and recurrent disease. Collaboration with other investigators was an obvious way to improve this, but strategies were also employed to make use of an extensive archive of FFPE tissue. Whilst DNA studies, such as methylation arrays, have advanced the use of FFPE tissue (Pajtler et al., 2017, 2015; Ramaswamy et al., 2016), this has not been the same for RNA. By demonstrating that FFPE RNA sequencing is feasible in paediatric ependymoma, the potential for future gene expression work using this technique has increased.

Whilst it was clear that the FFPE cohort reflected the clustering and gene expression patterns expected from the dataset, there were still some limitations as demonstrated by increased variability between the FFPE and FF samples. The main areas for improvement were: to minimise the impact of bacterial contamination of blocks, stored for long periods in non-sterile environments; to optimise protocols to minimise the impact of over sequencing short RNA fragments; and to develop a better understanding of acceptable quality control cut-offs for FFPE RNA-seq data.

This study identified an attrition rate of approximately 20% of samples, largely due to over representation of bacterial reads. It could be argued that using RNA-seq on these blocks creates the potential to yield a great deal of new data, particularly for rare diseases where samples are hard to obtain. This must be balanced against the risk of sample loss which may be financially costly.

Compared to the FF tissue, the FFPE tissue was associated with shorter fragments of RNA and cDNA for sequencing. This resulted in some over-sequencing, with the presence of adapter content, and nucleotide bases that could not be called.

Future researchers must set their initial sequencing parameters with this in mind; reads of 100 bases are not needed for FFPE blocks.

A further difficulty with interpreting the FFPE RNA-seq data was a lack of information about what constitutes acceptable quality. Based on the results of the bioinformatic pipeline, many of the FFPE samples had quality control values below what was seen in the FF samples. However, as the final analysis still reflected FF and published data findings, a better understanding of acceptable quality control levels is needed. Until data quality metrics are standardised for FFPE tissue, confirmation of the findings of FFPE RNA-seq, with either a validation RNA-seq cohort or an orthogonal technique, is essential.

Ensuring the correct molecular classification of tumours

An important lesson learned from the combined analysis of DNA methylation and gene expression profiles, was that molecular subgroups of ependymoma should not be defined on the basis of gene expression profiling alone. Whilst previous authors have delineated EPN_PFA and EPN_PFB groups on the basis of gene expression (A. Griesinger et al., 2015; Hoffman et al., 2014a; Wani et al., 2012; Witt et al., 2011a), the profiles generated from this study have provided evidence that the EPN_PFB gene expression phenotype is not restricted to DNA methylation confirmed EPN_PFB. Whilst almost all of the tumours with EPN_PFA gene expression were EPN_PFA by DNA methylation profiling, the group consistent with EPN_PFB gene expression contained a mixture of molecular diagnoses, including EPN_PFA, EPN_PFB, EPN_MPE and EPN_YAP.

Hoffman and colleagues (2014) used gene expression profiling alone to delineate two distinct PF groups, and concluded that they were EPN_PFA and EPN_PFB ependymomas. As part of this study's collaboration with Hoffman and colleagues, additional DNA methylation classifications were obtained for a number of their PF tumours. All of the tumours in their EPN_PFB group, with available DNA methylation predictions, were in fact found to be EPN_PFA. This supported the findings from the FF and FFPE cohorts in the present study. This has shown that future research must, where possible, back up molecular subgroup information with the gold standard diagnostic approach of DNA methylation analysis. If this is not achieved, there is a danger that conclusions are reached based on incorrect subgrouping that may adversely influence clinical trial planning and future treatment decisions.

The evidence for multiple EPN_PFA gene expression groups in Chapter 6, alongside the clustering of the DNA methylation samples in Chapter 4, provided support for the existence of multiple EPN_PFA molecular subgroups. This data has contributed to further research in this area (Pajtler et al., 2017).

Gene expression changes at primary and recurrence

When stratified by molecular subgroup, the results of the matched primary and recurrent analyses showed strong overlaps between the FFPE and FF cohorts and provided further insight into the importance of molecular subgrouping. The differential expression results based on PF location in the two cohorts conflicted with one another. When analysed in more depth, it became apparent that the probable reason for this conflict was the different molecular composition of the posterior fossa location based subgroups. PF1 was the predominant cluster in the FFPE samples, whilst PF1 and PF2 were more balanced in the FF samples.

PF2 and ST samples demonstrated similar patterns of recurrence to one another. There was an upregulation of genes and ontology terms related to the adaptive immune response, and evidence of induction of the type I IFN pathway. The results from the PF1 group differed to this, demonstrating a downregulation of the innate immune and inflammatory responses characteristic of this subgroup at primary presentation. Importantly, in the PF1 subgroup, these downregulated changes appeared irrespective of treatment with radiotherapy.

The fact that the gene expression defined subgroups showed the most consistent changes at recurrence, when compared to location defined subgroups, suggested that recurrence pattern may actually be dependent on the original gene expression phenotype at primary diagnosis. Even though the PF2 subgroup contained several different DNA methylation predictions, the behaviour of this group was still the same across two independent cohorts.

RNA sequencing was performed on whole tumour samples rather than selected tumour cells. Therefore, the gene expression pattern would to some extent be influenced by the tumour microenvironment. The PF1 group in particular, was associated with terms related to the microenvironment, including the extracellular matrix, immune response, wound healing and inflammatory response. This, along with immune related changes at recurrence, raises the question as to whether there is a role for the tumour microenvironment in the mechanisms of recurrence.

Role of the immune system

In line with previous authors (Hoffman et al., 2014a; Donson et al., 2009) this study implicated the immune system in recurrent paediatric ependymoma. However, the use of: RNA sequencing; DNA methylation subgroup assignment; a wider sample set; and threshold free gene set enrichment, provided a greater resolution to the analysis.

PF1 tumours showed downregulation of innate immune response ontologies. This was supported by the fact that CYT activity and immunophenoscores did not change, as they represent markers of an adaptive response in an immunogenic environment. There was a significant downregulation of a number of ontologies related to inflammatory chemokines and immune cell taxis. This raised the possibility that the PF1 tumours exhibited some sort of tumour immune escape mechanism (Spranger, 2016). This provides a plausible explanation for how these PF1 tumours might recur. The finding of a fall in the inflammatory marker CD45 in the posterior fossa IHC validation cohort, also strongly supported the theory that the changes in gene expression in the PF1 subgroup were translated to a response at the protein level.

It was less clear how recurrence occurred in PF2 and ST tumours, which demonstrated an upregulation of the adaptive immune response. If this immune response was anti-tumour, one might expect to identify either an increase in tumour immunogenicity, stimulating an immune response; or an increase in cell killing, as the final step in an effective immune response. Additionally, via IHC, one might expect to see an increase in the number of adaptive immune cells in the tumour parenchyma at recurrence. However, these tumours showed no evidence of an associated increase in the immunophenoscore, cytolytic activity or immune cell infiltration at recurrence. This could be explained by at least three possible hypotheses. Firstly, the adaptive response seen was somehow pro-tumour in nature and stimulated further tumour growth. Secondly, the tumour overcame an anti-tumour response. Thirdly, there was an anti-tumour immune response, but the effector immune cells were unable to access tumour cells due to the nature of the ECM or BBB, thus confining the response to the surrounding stroma. This last hypothesis was supported by previous research on tumour immunophenotypes including immune excluded and immune desert tumours (Chen and Mellman, 2017). The IHC data was highly suggestive of adaptive immune cell exclusion, with significantly more cells seen in non-parenchymal

areas, such as resection cavities and necrosis, than parenchymal areas. Further work to investigate these hypotheses is warranted.

Changes at recurrence largely occurred irrespective of whether radiotherapy was delivered after primary treatment. However, in the PF1 group there was evidence of an association between upregulation of the type I IFN pathway and radiotherapy. A technical validation by qPCR demonstrated that this change was not artefactual. Radiotherapy is known to invoke a type I IFN response in other cancers (Lim et al., 2014) and the IFN pathway has been suggested to have both positive and negative impacts upon cancer survival (Bracci et al., 2017; Minn, 2015; Parker et al., 2016; Weichselbaum et al., 2008; Zitvogel et al., 2015). Consequently, this needs more extensive investigation to confirm whether radiotherapy was responsible for this change. If this association is confirmed then consideration of how this may impact patient management is required.

A further important observation was the low level of expression of immune checkpoint molecules across FF and FFPE datasets. Given that there is increasing interest in immune checkpoint blockade and other immune modulating treatments as therapies for cancer (Anderson, 2014; Charoentong et al., 2017; Connolly et al., 2016; Liu and Zeng, 2012), a better understanding of their presence in ependymoma was warranted. Despite ongoing clinical trials into checkpoint blockade in ependymoma, remarkably, evidence was only found of two conflicting studies investigating whether these molecules are expressed at the protein level (Dumont et al. 2017, Witt et al. 2018). The current study supported the findings of Dumont, by demonstrating that at the gene expression level, checkpoint markers were lowly expressed. This raises questions over the potential efficacy of checkpoint blockade therapy in ependymoma. However, there may still be a role for other immunotherapies in this disease. Modalities such as Chimeric Antigen Receptor (CAR) T-cell therapy might be considered, particularly in light of some emerging evidence about its potential effectiveness in GBM (O'Rourke et al., 2017). A better understanding of the underlying ependymoma immune environment is therefore critical in considering different treatment modalities, and preventing both unanticipated negative consequences for patients, and the unnecessary use of finite resources.

Immunohistochemistry

The IHC cohort provided an initial insight into the relationship between immune cell infiltrate and survival in the recurrent cohort. It was surprising that CD8+ cell

infiltration appeared to be associated with worse outcomes in multivariate analysis and contradicted findings of studies of CD8+ cell infiltration in other tumours (Mao et al., 2016). This finding needs confirmation in an independent cohort. Further IHC is also needed to identify whether these CD8+ cells are indeed CD8+ effector cells capable of cell killing, or whether they represent another CD8+ subset which could be associated with suppressor activity. CD20+ cells were found to be associated with worse outcomes in a univariate analysis but this was lost in a multivariate approach.

The spatial distribution of the immune cells was highly suggestive that all subtypes of ependymoma are T-cell excluded tumours. Possible mechanisms for this include: the chemokine profile of the tumour; properties of the tumour's ECM; or functioning of the BBB (Muldoon et al., 2013). The changes in chemokines between primary and recurrence, and between molecular subgroups in the RNA sequencing dataset, may begin to provide some explanation for the mechanisms of immune exclusion in these tumours. The findings of the immune cell spatial distribution needs validation in another dataset, as this appears to be the first time this finding has been described in the literature in paediatric ependymoma. However, if the spatial distribution and behaviour as an immune excluded tumour are confirmed, this would provide supporting evidence as to how immune escape mechanisms indicated by the gene expression analysis may operate.

Recommendations for future work

Clinical Data

Whilst the behaviour of the more common DNA methylation subgroups is becoming better understood at primary diagnosis (Pajtler et al., 2015; Ramaswamy et al., 2016), and now at recurrence; further understanding is needed of the behaviour of the less frequently occurring subgroups such as EPN_PFB and EPN_YAP. The fact that both of these groups appeared in the recurrent dataset suggested that outcomes may not be as good as previously suggested. Large, international, collaborative efforts are required to generate sufficiently sized cohorts with these molecular subgroups, in order to evaluate the true natural history of these diseases, before any decisions are made on altering their therapeutic burden.

DNA Methylation Profiling

Whilst molecular subgroups remained largely unchanged at recurrence, there was evidence that correlation scores were lower in the relapsed tumours. Further investigation of the DNA methylation cohort is therefore required to analyse whether there are significantly differentially methylated probes at recurrence and whether they are related to the immune response.

Immune studies

Whilst associations between recurrent disease and changes in the immune response have been identified in this retrospective study, confirmation of these findings are needed in a prospective trial. The BIOMarkers of Ependymoma in Children and Adolescents (BIOMECA) arm of the current European ependymoma clinical trial (SIOP Ependymoma II) may provide an opportunity to investigate this further. A prospective design would also be able to specifically ask questions about the association between therapeutic interventions, such as radiotherapy, and underlying biological changes.

Patient based studies may be limited to observational approaches. Therefore, it would be beneficial to demonstrate any therapeutic associations in a suitable animal model, for example a mouse model of ependymoma. Cell culture work is often used for follow up of biological findings. However, this is limited by the lack of tumour microenvironment and immune infiltrate. In spite of this, work has begun to identify whether secretion of the chemokines identified in the RNA-seq analysis is derived from tumour or immune cells. Additionally, it may be interesting to investigate the response of ependymoma cell lines to inflammatory chemokines.

A further question relates to whether there is a change in neoantigen load in recurrent ependymoma compared to matched primary disease. Increased neoantigen expression has been associated with an increased host immune response to tumour (Charoentong et al., 2017; Rooney et al., 2015). RNA sequencing data cannot provide detail about the overall level of neoantigen load because it does not sequence across the entire genome. Therefore, whole genome or exome sequencing of matched primary and recurrent tumours are needed to estimate changes in the neoantigen burden, and to infer whether it may be responsible for driving an immune response in the PF2 and ST samples.

Non-recurrent tumours

It is unclear whether non-recurrent ependymomas represent different entities to recurrent ependymomas. This study has focussed on the biology of a cohort of recurrent tumours; a biological comparison with a cohort of non-recurrent tumours is required. Work is underway to identify this cohort in order to undertake RNA sequencing to compare gene expression patterns. IHC is also being undertaken to identify whether the immune infiltrate of the non-recurrent tumours differs to that of the recurrent tumours. Work by previous authors (Donson et al., 2009) has found some weak associations between immune infiltrate and recurrence status, but was probably underpowered to detect significant differences.

Final Conclusions

The key strengths of this study were the size and follow up duration of the clinical data, coupled with molecular analysis using both gene expression and DNA methylation profiles. The use of RNA-Seq from FFPE in a paediatric brain tumour cohort was novel. This allowed the inclusion of more samples and validated the approach for future work. The combination of these entities meant that the study was able to robustly reflect the recent molecular developments in ependymoma research. No other studies were found during the literature search which linked these elements in an analysis of paired primary and recurrent disease.

Clinical behaviour of recurrent disease was similar across intracranial molecular subtypes. Standard therapy did not reliably prevent recurrence and only provided transient benefits after first relapse. The main factor determining overall survival in this disease was recurrence status.

Molecular analyses confirmed a strong association between recurrence and the immune system, but raised new questions about the ability of immune cells to interact directly with ependymoma cells in view of their spatial distribution. Furthermore, protein level analyses indicated that ependymoma is an immune excluded tumour.

Therapy appeared to have minimal impact on tumours with an inflammatory phenotype at primary presentation (PF1) and there was some evidence that these tumours may undergo immune escape. It was possible that radiotherapy had some impact on the type I IFN pathway in this subgroup, but this needs further confirmation.

Development of an adaptive immune response in PF2 and ST molecular subgroups did not prevent recurrence and the origin and implications of this response require further investigation.

Further research is warranted to fully understand the underlying tumour biology, and to investigate the role of the immune system in ependymoma. This is vital to guide further treatment interventions, including whether tumour modulating immunotherapy may be a viable option in this devastating disease.

References

Addeo, A., Rinaldi, C.R., 2013. "Treatment with ipilimumab: A case report of complete response in a metastatic malignant melanoma patient". *Case Rep. Oncol.* **6**(2): 285–288.

Agaoglu, F.Y., Ayan, I., et al., 2005. "Ependymal tumors in childhood". *Pediatr. Blood Cancer* **45**(3): 298–303.

Aguirre-Ghiso, J. a, 2007. "Models, mechanisms and clinical evidence for cancer dormancy.". *Nat. Rev. Cancer* **7**(11): 834–46.

Ailon, T., Dunham, C., et al., 2014. "The role of resection alone in select children with intracranial ependymoma: the Canadian Pediatric Brain Tumour Consortium experience". *Child's Nerv. Syst.* **31**(1): 57–65.

Aissani, B., Bernardi, G., 1991. "CpG islands, genes and isochores in the genomes of vertebrates". *Gene* **106**(2): 185–195.

Aizer, A. a., Ancukiewicz, M., et al., 2013. "Natural history and role of radiation in patients with supratentorial and infratentorial WHO grade II ependymomas: Results from a population-based study". *J. Neurooncol.* **115**(3): 411–419.

Allen, J., Packer, R., et al., 1991. "Recombinant interferon beta: a phase I-II trial in children with recurrent brain tumors.". *J. Clin. Oncol.* **9**(5): 783–788.

Almeida, L.G., Sakabe, N.J., et al., 2009. "CTdatabase: A knowledge-base of high-throughput and curated data on cancer-testis antigens". *Nucleic Acids Res.* **37**(SUPPL. 1): 2007–2010.

Ameur, A., Zaghlool, A., et al., 2011. "Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain". *Nat. Struct. Mol. Biol.* **18**(12): 1435–1440.

Amini, P., Ettlin, J., et al., 2017. "An optimised protocol for isolation of RNA from small sections of laser-capture microdissected FFPE tissue amenable for next-generation sequencing". *BMC Mol. Biol.* **18**(1): 22.

Amirian, E.S., Armstrong, T.S., et al., 2012. "Predictors of Survival among Pediatric and Adult Ependymoma Cases: A Study Using Surveillance, Epidemiology, and End Results Data from 1973 to 2007". *Neuroepidemiology* **39**(2): 116–124.

Anderson, A.C., 2014. "Tim-3: An Emerging Target in the Cancer Immunotherapy Landscape". *Cancer Immunol. Res.* **2**(5): 393–398.

Andrews, S., 2010. "FastQC: a quality control tool for high throughput sequence data.".

Ansorge, W.J., 1991. "New nucleic acid sequencing method - using one labelled nucleotide at one time in cycles comprising elongation, wash, label detection and removal of the label, then repeating". 4141178.

Ansorge, W.J., 2009. "Next-generation DNA sequencing techniques". *N. Biotechnol.* **25**(4): 195–203.

Antony, R., Wong, K.E., et al., 2014. "A retrospective analysis of recurrent intracranial ependymoma.". *Pediatr. Blood Cancer* **61**(7): 1195–201.

Aryee, M.J., Jaffe, A.E., et al., 2014. "Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays". *Bioinformatics* **30**(10): 1363–1369.

Auerbach, S.S., Phadke, D.P., et al., 2015. "RNA-Seq-based toxicogenomic assessment of fresh frozen and formalin-fixed tissues yields similar mechanistic insights". *J. Appl. Toxicol.* **35**(7): 766–780.

Baba, T., Ishizu, A., et al., 2006. "CD4+/CD8+ macrophages infiltrating at inflammatory sites: a population of monocytes/macrophages with a cytotoxic phenotype". *Blood* **107**(5): 2004–2012.

Bayliss, J., Mukherjee, P., et al., 2016. "Lowered H3K27me3 and DNA hypomethylation define poorly prognostic pediatric posterior fossa ependymomas". *Sci. Transl. Med.* **8**(366): 366ra161.

BDBiosciences, 2016. Human and Mouse CD Marker Handbook. BD Biosciences.

Becher, O.J., Millard, N.E., et al., 2017. "A phase i study of single-agent perifosine for recurrent or refractory pediatric CNS and solid tumors". *PLoS One* **12**(6): 1–15.

Birch, J.M., Marsden, H.B., 1987. "A classification scheme for childhood cancer". *Int. J. Cancer* **40**(5): 620–624.

Blair, G.E., Cook, G.P., 2008. "Cancer and the immune system: an overveiw". *Oncogene* **27**(45): 5868.

Blow, N., 2007. "Tissue preparation: Tissue issues.". *Nature* **448**(7156): 959–963.

Bolger, A.M., Lohse, M., et al., 2014. "Genome analysis Trimmomatic : a flexible trimmer for Illumina sequence data". *Bioinformatics* **30**(15): 2114–2120.

Bouffet, E., Hawkins, C.E., et al., 2012. "Survival benefit for pediatric patients with recurrent ependymoma treated with reirradiation.". *Int. J. Radiat. Oncol. Biol. Phys.* **83**(5): 1541–8.

Bouffet, E., Perilongo, G., et al., 1998. "Intracranial ependymomas in children: A critical review of prognostic factors and a plea for cooperation". *Med. Pediatr. Oncol.* **30**(6): 319–331.

Bracci, L., Sistigu, A., et al., 2017. "The added value of type I interferons to cytotoxic treatments of cancer". *Cytokine Growth Factor Rev.* **36**: 89–97.

Brocker, C., Thompson, D., et al., 2010. "Evolutionary divergence and functions of the human interleukin (IL) gene family". *Hum. Genomics* **5**(1): 30–55.

Brown, G.R., Hem, V., et al., 2015. "Gene: A gene-centered information resource at NCBI". *Nucleic Acids Res.* **43**(D1): D36–D42.

Cage, T. a, Clark, A.J., et al., 2013. "A systematic review of treatment outcomes in pediatric patients with intracranial ependymomas.". *J. Neurosurg. Pediatr.* **11**(6): 673–81.

Cancer Research UK, 2015a. UK Childhood Cancer Statistics [WWW Document]. URL http://publications.cancerresearchuk.org/downloads/Product/CS_KF_CHILDHOOD.pdf (accessed 2.9.15).

Cancer Research UK, 2015b. Children's brain, other CNS and intracranial tumour incidence [WWW Document]. *UK Cancer Stat.* URL http://www.cancerresearchuk.org/health-professional/cancer-statistics/childrens-cancers/incidence#heading-Five (accessed 11.1.17).

Carlson, M., 2017. "org.Hs.eg.db: Genome wide annotation for Human".

Carson, M.J., Doose, J.M., et al., 2006. "CNS immune privilege: Hiding in plain sight". *Immunol. Rev.* **213**(1): 48–65.

Carter, M., Nicholson, J., et al., 2002. "Genetic abnormalities detected in ependymomas by comparative genomic hybridisation.". *Br. J. Cancer* **86**(6): 929–939.

Catakovic, K., Klieser, E., et al., 2017. "T cell exhaustion: from pathophysiological basics to tumor immunotherapy". *Cell Commun. Signal.* **15**(1): 1.

CCLG, 2017. The Children's Cancer and Leukaemia Group (CCLG) Biobank [WWW Document]. *CCLG Website*. URL http://www.cclg.org.uk/tissue-bank/ (accessed 11.1.17).

Chalkley, R., Hunter, C., 1975. "Histone-histone propinquity by aldehyde fixation of chromatin.". *Proc. Natl. Acad. Sci. U. S. A.* **72**(4): 1304–1308.

Charoentong, P., Angelova, M., et al., 2017. "Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade Resource Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkp". *Cell Rep.* **18**(1): 248–262.

Chen, D.S., Mellman, I., 2017. "Elements of cancer immunity and the cancer–immune set point". *Nature* **541**(7637): 321–330.

Cheng, L., Ruan, Z., 2015. "Tim-3 and Tim-4 as the potential targets for antitumor therapy". *Hum. Vaccines Immunother.* **11**(10): 2458–2462.

Ching, T., Huang, S., et al., 2014. "Power analysis and sample size estimation for RNA-Seq differential expression". *RNA* **20**(11): 1–13.

Clarke, M.F., Dick, J.E., et al., 2006. "Cancer stem cells--perspectives on current

status and future directions: AACR Workshop on cancer stem cells.". *Cancer Res.* **66**(19): 9339–44.

Cock, P.J.A., Fields, C.J., et al., 2010. "The Sanger FASTQ file format for sequences with quality scores , and the Solexa / Illumina FASTQ variants". *Nucleic Acids Res.* **38**(6): 1767–1771.

Connolly, K.A., Belt, B.A., et al., 2016. "Increasing the efficacy of radiotherapy by modulating the CCR2/CCR5 chemokine axes". *Oncotarget* **7**(52): 86522–86535.

Corallo, D., Schiavinato, A., et al., 2013. "Emilin3 is required for notochord sheath integrity and interacts with Scube2 to regulate notochord-derived Hedgehog signals". *Development* **140**(22): 4594–4601.

Corallo, D., Schiavinato, A., et al., 2017. "EMILIN3, an extracellular matrix molecule with restricted distribution in skin". *Exp. Dermatol.* **26**(5): 435–438.

Dardis, C., 2016. "survMisc: Miscellaneous Functions for Survival Data".

de Ruiter, E.J., Ooft, M.L., et al., 2017. "The prognostic role of tumor infiltrating T-lymphocytes in squamous cell carcinoma of the head and neck: A systematic review and meta-analysis". *Oncoimmunology* **6**(11): e1356148.

Del Fabbro, C., Scalabrin, S., et al., 2013. "An extensive evaluation of read trimming effects on illumina NGS data analysis". *PLoS One* **8**(12): 1–13.

DeWire, M., Fouladi, M., et al., 2015. "An open-label, two-stage, phase II study of bevacizumab and lapatinib in children with recurrent or refractory ependymoma: a collaborative ependymoma research network study (CERN)". *J. Neurooncol.* **123**(1): 85–91.

Donson, A.M., Birks, D.K., et al., 2009. "Immune gene and cell enrichment is associated with a good prognosis in ependymoma.". *J. Immunol.* **183**(11): 7428–7440.

Duffner, P., Horowitz, M., et al., 1993. "Postoperative chemotherapy and delayed radiation in children less than three years of age with malignant brain tumours". *N. Engl. J. Med.* **328**(24): 1725–31.

Duffner, P.K., Krischer, J.P., et al., 1998. "Prognostic Factors in Infants and Very Young Children with Intracranial Ependymomas". *Pediatr. Neurosurg.* **28**(4): 215–222.

Dumont, B., Forest, F., et al., 2017. "PD1 and PD-L1 in ependymoma might not be therapeutic targets". *Clin. Neuropathol.* **36**: 90–92.

Dunn, G.P., Bruce, A.T., et al., 2005. "A critical function for type I interferons in cancer immunoediting". *Nat. Immunol.* **6**(7): 722–729.

Dvorak, H.F., 1986. "Tumors: Wounds That Do Not Heal". *N. Engl. J. Med.*

315(26): 1650–1659.

Dyer, S., Prebble, E., et al., 2002. "Genomic imbalances in pediatric intracranial ependymomas define clinically relevant groups.". *Am. J. Pathol.* **161**(6): 2133–2141.

Eden, E., Navon, R., et al., 2009. "GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists". *BMC Bioinformatics* **10**(1): 48.

Ellison, D.W., Kocak, M., et al., 2011. "Histopathological grading of pediatric ependymoma: reproducibility and clinical relevance in European trial cohorts.". *J. Negat. Results Biomed.* **10**(1): 7.

ENCODE, 2016. ENCODE Guidelines and Best Practices for RNA-seq [WWW Document]. *ENCODE RNA-seq Guidel.* URL https://www.encodeproject.org/documents/cede0cbe-d324-4ce7-ace4-f0c3eddf5972/@@download/attachment/ENCODE Best Practices for RNA_v2.pdf (accessed 5.22.17).

Esche, C., Stellato, C., et al., 2005. "Chemokines: key players in innate and adaptive immunity". *J. Invest. Dermatol.* **125**(4): 615–628.

Esmatabadi, M.J.D., Bakhshinejad, B., et al., 2016. "Therapeutic resistance and cancer recurrence mechanisms: Unfolding the story of tumour coming back". *J. Biosci.* **41**(3): 497–506.

Esteve-Codina, A., Arpi, O., et al., 2017. "A comparison of RNA-Seq results from paired formalin-fixed paraffin-embedded and fresh-frozen glioblastoma tissue samples". *PLoS One* **12**(1): 1–18.

Evans, A.E., Anderson, J.R., et al., 1996. "Adjuvant Chemotherapy of Childhood Posterior Fossa Ependymoma: Cranio-Spinal Irradiation With or Without Adjuvant CCN U, Vincristine, and Prednisone: A Childrens Cancer Group Study". *Med. Pediatr. Oncol.* **27**(1): 8–14.

Ewing, B., Green, P., 1998. "Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities". *Genome Res.* **8**(3): 186–94.

Ewing, B., Hillier, L., et al., 1998. "Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment". *Genome Res.* **8**(3): 251–9.

Fang, Z., Cui, X., 2011. "Design and validation issues in RNA-seq experiments". *Brief. Bioinform.* **12**(3): 280–287.

Farkona, S., Diamandis, E.P., et al., 2016. "Cancer immunotherapy: the beginning of the end of cancer?". *BMC Med.* **14**(1): 73.

Figarella-Branger, D., Civatte, M., et al., 2000. "Prognostic factors in intracranial ependymomas in children.". *J. Neurosurg.* **93**(4): 605–613.

Filaci, G., Suciu-foca, N., 2002. "CD8+ T suppressor cells are back to the game : are they players in autoimmunity ?". *Autoimmun. Rev.* **1**(5): 279–283.

Fine, J.P., Gray, R.J., 1999. "A Proportional Hazards Model for the Subdistribution of a Competing Risk". *J. Am. Stat. Assoc.* **94**(446): 496–509.

Fisher, R.A., 1925. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh.

Fisher, R.A., Mosteller, F., 1948. "Questions and answers #14". *Am. Stat.* **2**(5): 30–31.

Fleming Thomas R. O'Sullivan Margaret, H.D.P., 1987. "Supremum Versions of the Log Rank and Generalized Wilcoxon Statistics". *J. Am. Stat. Assoc.* **82**(397): 312–320.

Foreman, N.K., Love, S., et al., 1997. "Second-look Surgery for Incompletely Resected Fourth Ventricle Ependymomas: Technical Case Report". *Neurosurgery* **40**(4): 856–860.

Forment, J. V, Kaidi, A., et al., 2012. "Chromothripsis and cancer: causes and consequences of chromosome shattering.". *Nat. Rev. Cancer* **12**(10): 663–70.

Fortin, J.P., Triche, T.J., et al., 2017. "Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi". *Bioinformatics* **33**(4): 558–560.

Fox, C.H., Johnson, F.B., et al., 1985. "Formaldehyde fixation.". *J. Histochem. Cytochem.* **33**(8): 845–853.

Franca, L.T.C., Carrilho, E., et al., 2002. "A review of DNA sequencing techniques.". *Q. Rev. Biophys.* **35**(2): 169–200.

French, S.W., Mendoza, A.S., et al., 2017. "The role of the IL-8 signaling pathway in the in fi ltration of granulocytes into the livers of patients with alcoholic hepatitis". *Exp. Mol. Pathol.* **103**(2): 137–140.

Furth, J., Kahn, M.C., 1937. "The transmission of leukaemia of mice with a single cell". *Cancer Res.* **31**(2): 276–282.

Garvin, J., Selch, M.T., et al., 2012. "Phase II Study of Pre-Irradiation Chemotherapy for Childhood Intracranial Ependymoma. Children's Cancer Group Protocol 9942: A Report From the Children's Oncology Group". *Pediatr. Blood Cancer* **59**(7): 1183–1189.

Gatta, G., Botta, L., et al., 2014. "Childhood cancer survival in Europe 1999-2007: Results of EUROCARE-5-a population-based study". *Lancet Oncol.* **15**(1): 35–47.

Geyer, J.R., Sposto, R., et al., 2005. "Multiagent Chemotherapy and Deferred Radiotherapy in Infants With Malignant Brain Tumors : A Report From the Children ' s Cancer Group". *J. Clin. Oncol.* **23**(30): 9–11.

Gilbertson, R.J., Bentley, L., et al., 2002. "Advances in Brief ERBB Receptor

Signaling Promotes Ependymoma Cell Proliferation and Represents a Potential Novel Therapeutic Target for This Disease". *Clin. Cancer Res.* **8**(10): 3054–3064.

Gimsa, U., Mitchison, N.A., et al., 2013. "Immune privilege as an intrinsic CNS property: Astrocytes protect the CNS against T-cell-mediated neuroinflammation". *Mediators Inflamm.* **2013**: 320519.

Godfraind, C., Kaczmarska, J.M., et al., 2012. "Distinct disease-risk groups in pediatric supratentorial and posterior fossa ependymomas". *Acta Neuropathol.* **124**(2): 247–257.

Goldwein, J.W., Glauser, T. a., et al., 1990. "Recurrent intracranial ependymomas in children. Survival, patterns of failure, and prognostic factors". *Cancer* **66**(3): 557–563.

Graw, S., Meier, R., et al., 2015. "Robust gene expression and mutation analyses of RNA-sequencing of formalin-fixed diagnostic tumor samples". *Sci. Rep.* **5**: 12335.

Griesinger, A.A.M., Josephson, R.R.J., et al., 2015. "Interleukin-6/STAT3 pathway signaling drives an inflammatory phenotype in Group A ependymoma". *Cancer Immunol. Res.* **3**(10): 1165–1174.

Griesinger, A.M., Witt, D.A., et al., 2017. "NF-κB upregulation through epigenetic silencing of LDOC1 drives tumor biology and specific immunophenotype in Group A ependymoma". *Neuro. Oncol.* **19**(10): 1350–1360.

Grill, J., Le Deley, M.C., et al., 2001. "Postoperative chemotherapy without irradiation for ependymoma in children under 5 years of age: a multicenter trial of the French Society of Pediatric Oncology.". *J. Clin. Oncol.* **19**(5): 1288–96.

Grizzi, F., Gaetani, P., et al., 2006. "Sperm protein 17 is expressed in human nervous system tumours". *BMC Cancer* **6**: 23.

Groeneveld, T.W.L., Oroszlan, M., et al., 2005. "Interactions of the Extracellular Matrix Proteoglycans Decorin and Biglycan with C1q and Collectins". *J. Immunol.* **175**(7): 4715–4723.

Group, T.S.F.S.W., 2016. Sequence Alignment / Map Format Specification [WWW Document]. URL https://samtools.github.io/hts-specs/SAMv1.pdf (accessed 11.22.16).

Grundy, R.G., Wilne, S. a, et al., 2007. "Primary postoperative chemotherapy without radiotherapy for intracranial ependymoma in children: the UKCCSG/SIOP prospective study.". *Lancet. Oncol.* **8**(8): 696–705.

Guo, Y., Wu, J., et al., 2016. "RNA Sequencing of Formalin-Fixed, Paraffin-Embedded Specimens for Gene Expression Quantification and Data Mining".

Int. J. Genomics **2016**: 9837310.

Gururangan, S., Fangusaro, J., et al., 2012. "Lack of efficacy of bevacizumab + irinotecan in cases of pediatric recurrent ependymoma — a Pediatric Brain Tumor Consortium study" **14**(11): 1404–1412.

Haile, S., Pandoh, P., et al., 2017. "Automated high throughput nucleic acid purification from formalin-fixed paraffinembedded tissue samples for next generation sequence analysis". *PLoS One* **12**(6): 1–15.

Hanahan, D., Weinberg, R. a., 2000. "The hallmarks of cancer". *Cell* **100**(1): 57–70.

Hanahan, D., Weinberg, R. a., 2011. "Hallmarks of cancer: The next generation". *Cell* **144**(5): 646–674.

He, S.-J., Gu, Y.-Y., et al., 2014. "High expression and frequently humoral immune response of melanoma-associated antigen D4 in glioma". *Int. J. Clin. Exp. Pathol.* **7**(5): 2350–2360.

Hedegaard, J., Thorsen, K., et al., 2014. "Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue". *PLoS One* **9**(5).

Herbst, R.S., Soria, J.C., et al., 2014. "Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients". *Nature* **515**(7528): 563–567.

Hester, S.D., Bhat, V., et al., 2016. "Editor's Highlight: Dose–Response Analysis of RNA-Seq Profiles in Archival Formalin-Fixed Paraffin-Embedded Samples". *Toxicol. Sci.* **154**(2): 202–213.

Hirose, Y., Aldape, K., et al., 2001. "Chromosomal abnormalities subdivide ependymal tumors into clinically relevant groups.". *Am. J. Pathol.* **158**(3): 1137–1143.

Hodi, F.S., O'Day, S.J., et al., 2010. "Improved Survival with Ipilimumab in Patients with Metastatic Melanoma". *N. Engl. J. Med.* **363**(8): 711–723.

Hoffman, L.M., Donson, A.M., et al., 2014a. "Molecular sub-group-specific immunophenotypic changes are associated with outcome in recurrent posterior fossa ependymoma.". *Acta Neuropathol.* **127**(5): 731–45.

Hoffman, L.M., Plimpton, S.R., et al., 2014b. "Fractionated stereotactic radiosurgery for recurrent ependymoma in children". *J Neurooncol* **116**(1): 107–111.

Hong, G., Zhang, W., et al., 2013. "Separate enrichment analysis of pathways for up- and downregulated genes". *J. R. Soc. Interface* **11**(92): 20130950–20130950.

Howlader N, Noone AM, Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL,

Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, C.K. (eds)., 2013. "SEER Cancer Statistics Review, 1975-2011". *Natl. Cancer Institute. Bethesda, MD*.

Huang, D.W., Sherman, B.T., et al., 2009a. "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.". *Nat. Protoc.* **4**(1): 44–57.

Huang, D.W., Sherman, B.T., et al., 2009b. "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.". *Nucleic Acids Res.* **37**(1): 1–13.

Hurwitz, C.A., Strauss, L.C., et al., 2001. "Paclitaxel for the treatment of progressive or recurrent childhood brain tumors: a pediatric oncology phase II study.". *J. Pediatr. Hematol. Oncol.* **23**(5): 277–281.

Ichim, C. V, Coley, W., et al., 2005. "Revisiting immunosurveillance and immunostimulation: Implications for cancer immunotherapy". *J. Transl. Med.* **3**(1): 8.

Illumina, 2014. FastQ Files [WWW Document]. URL http://support.illumina.com/content/dam/illumina-support/help/BaseSpaceHelp_v2/Content/Vault/Informatics/Sequencing_Analysis/BS/swSEQ_mBS_FASTQFiles.htm (accessed 11.21.16).

Illumina, 2015. "An Introduction to Next-Generation Sequencing Technology". *Illumina.com* (illumina): 1–16.

Indelicato, D.J., Bradley, J.A., et al., 2017. "Outcomes following proton therapy for pediatric ependymoma". *Acta Oncol. (Madr).* 1–5.

Jaing, T.-H., Wang, H.-S., et al., 2004. "Multivariate Analysis of Clinical Prognostic Factors in Children with Intracranial Ependymomas". *J. Neurooncol.* **68**(3): 255–261.

Jakacki, R.I., Foley, M.A., et al., 2016. "Single-agent erlotinib versus oral etoposide in patients with recurrent or refractory pediatric ependymoma: a randomized open-label study". *J. Neurooncol.* **129**(1): 131–138.

Jandial, R., Hoisang, U., et al., 2008. "Brain tumor stem cells and the tumor microenvironment". *Neurosurg Focus* **24**(3): 1–6.

Jeuken, J.W.M., Sprenger, S.H.E., et al., 2002. "Correlation between localization, age, and chromosomal imbalances in ependymal tumours as detected by CGH". *J. Pathol.* **197**(2): 238–244.

Jiang, D., Liang, J., et al., 2005. "Regulation of lung injury and repair by Toll-like receptors and hyaluronan". *Nat. Med.* **11**(11): 1173–1179.

Johnson, R. a, Wright, K.D., et al., 2010. "Cross-species genomics matches driver mutations and cell compartments to model ependymoma.". *Nature*

466(7306): 632–6.

Jones, D.T.W., Kocialkowski, S., et al., 2008. "Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas". *Cancer Res.* **68**(21): 8673–8677.

Jovanović, B., Sheng, Q., et al., 2017. "Comparison of triple-negative breast cancer molecular subtyping using RNA from matched fresh-frozen versus formalin-fixed paraffin-embedded tissue". *BMC Cancer* **17**(1): 241.

Just, P., Letourneur, F., et al., 2016. "Identification by FFPE RNA-Seq of a New Recurrent Inversion Leading to RBM10-TFE3 Fusion in Renal Cell Carcinoma with Subtle TFE3 Break-Apart FISH Pattern". *Genes. Chromosomes Cancer* **55**(6): 541–548.

Kaatsch, P., 2010. "Epidemiology of childhood cancer". *Cancer Treat. Rev.* **36**(4): 277–285.

Kaatsch, P., Steliarova-Foucher, E., et al., 2006. "Time trends of cancer incidence in European children (1978-1997): Report from the Automated Childhood Cancer Information System project". *Eur. J. Cancer* **42**(13): 1961–1971.

Kassambara, A., Kosinski, M., 2017. "survminer: Drawing Survival Curves using "ggplot2"".

Kieran, M.W., Chi, S.N., et al., 2015. "Tumors of the Brain and Spinal Cord". In: Orkin, S.H., Fisher, D.E., Ginsburg, D., Look, T.A., Lux, S.E., Nathan, D.G. (Eds.), Nathan and Oski's Hematology and Oncology of Infancy and Childhood. Elsevier Inc., Philadelphia, pp. 1779–1885.

Kilday, J.P., Rahman, R., et al., 2009. "Pediatric ependymoma: biological perspectives.". *Mol. Cancer Res.* **7**(6): 765–86.

Kilday, J.P., Mitra, B., et al., 2012. "Copy number gain of 1q25 predicts poor progression-free survival for pediatric intracranial ependymomas and enables patient risk stratification: A prospective european clinical trial cohort analysis on behalf of the Children's Cancer Leukaemia Group (CCLG". *Clin. Cancer Res.* **18**(3): 2001–2011.

Kim, D., Pertea, G., et al., 2013. "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.". *Genome Biol.* **14**(4): R36.

Kim, J.M., Chen, D.S., 2016. "Immune escape to PD-L1/PD-1 blockade: seven steps to success (or failure)". *Ann. Oncol.* **27**(8): 1492–1504.

Korshunov, A., Neben, K., et al., 2003. "Gene expression patterns in ependymomas correlate with tumor location, grade, and patient age.". *Am. J. Pathol.* **163**(5): 1721–1727.

Korshunov, A., Witt, H., et al., 2010. "Molecular staging of intracranial

ependymoma in children and adults". *J. Clin. Oncol.* **28**(19): 3182–3190.

Koshy, M., Rich, S., et al., 2011. "Post-operative radiation improves survival in children younger than 3 years with intracranial ependymoma". *J. Neurooncol.* **105**(3): 583–590.

Kratz, A., Carninci, P., 2014. "The devil in the details of RNA-seq". *Nat. Biotechnol.* **25**(9): 882–884.

Krishnadas, D.K., Shusterman, S., et al., 2015. "A phase I trial combining decitabine/dendritic cell vaccine targeting MAGE-A1, MAGE-A3 and NY-ESO-1 for children with relapsed or therapy-refractory neuroblastoma and sarcoma". *Cancer Immunol. Immunother.* **64**(10): 1251–1260.

Lawrence, M., Huber, W., et al., 2013. "Software for Computing and Annotating Genomic Ranges". *PLoS Comput. Biol.* **9**(8): 1–10.

Lee, J., Lund-smith, C., et al., 2009. "Glioma-induce remodeling of the neurovascular unit". *Brain Res.* **1288**: 125–134.

Lee, S., Margolin, K., 2011. "Cytokines in cancer immunotherapy". *Cancers (Basel).* **3**(4): 3856–3893.

Lehman, N.L., 2008. "Central nervous system tumors with ependymal features: A broadened spectrum of primarily ependymal differentiation?". *J. Neuropathol. Exp. Neurol.* **67**(3): 177–188.

Li, A.M., Dunham, C., et al., 2015. "EZH2 expression is a prognostic factor in childhood intracranial ependymoma: A Canadian Pediatric Brain Tumor Consortium study". *Cancer* **121**(9): 1499–1507.

Li, F., Huang, Q., et al., 2010. "Apoptotic cells activate the "phoenix rising" pathway to promote wound healing and tissue regeneration". *Sci. Signal.* **3**(110): ra13.

Li, P., Conley, A., et al., 2014. "Whole-Transcriptome profiling of formalin-fixed, paraffin-embedded renal cell carcinoma by RNA-seq". *BMC Genomics* **15**(1): 1087.

Li, X., Yan, J., et al., 2017. "Serum immunoreactivity of cancer/testis antigen OY-TES-1 and its tissues expression in glioma". *Oncol. Lett.* **13**(5): 3080–3086.

Li, Z., 2013. "CD133: A stem cell biomarker and beyond". *Exp. Hematol. Oncol.* **2**(1): 1–8.

Liao, Y., Smyth, G.K., et al., 2014. "FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features". *Bioinformatics* **30**(7): 923–930.

Lim, J.Y.H., Gerber, S.A., et al., 2014. "Type i interferons induced by radiation therapy mediate recruitment and effector function of CD8+ T cells". *Cancer Immunol. Immunother.* **63**(3): 259–271.

Liu, Y., Noon, A.P., et al., 2014. "Next-generation RNA Sequencing of Archival Formalin-fixed Paraffin-embedded Urothelial Bladder Cancer". *Eur. Urol.* **66**(6): 982–986.

Liu, Y., Zeng, G., 2012. "Cancer and Innate Immune System interactions: Translational Potentials for Cancer Immunotherapy". *J. Immunother.* **35**(4): 299–308.

Liu, Y., Zhou, J., et al., 2014. "RNA-seq differential expression studies: More sequence or more replication?". *Bioinformatics* **30**(3): 301–304.

Lobón, M.J., Bautista, F., et al., 2016. "Re-irradiation of recurrent pediatric ependymoma: modalities and outcomes: a twenty-year survey.". *Springerplus* **5**(1): 879.

Louis, D.N., Ohgaki, H., et al., 2007. "The 2007 WHO classification of tumours of the central nervous system.". *Acta Neuropathol.* **114**(2): 97–109.

Louis, D.N., Perry, A., et al., 2016. "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary". *Acta Neuropathol.* **131**(6): 1–18.

Love, M.I., Huber, W., et al., 2014. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". *Genome Biol.* **15**(12): 550.

Lukashova-v.Zangen I., I., Kneitz, S., et al., 2007. "Ependymoma gene expression profiles associated with histological subtype, proliferation, and patient survival". *Acta Neuropathol.* **113**: 325–337.

Mack, S.C., Taylor, M.D., 2017. "Put away your miroscopes: the ependymoma molecular era has begun". *Curr. Opin. Oncol.* **29**(6): 443–447.

Mack, S.C., Taylor, M.D., 2009. "The genetic and epigenetic basis of ependymoma.". *Child's Nerv. Syst.* **25**(10): 1195–201.

Mack, S.C., Witt, H., et al., 2013. "Emerging insights into the ependymoma epigenome.". *Brain Pathol.* **23**(2): 206–9.

Mack, S.C., Witt, H., et al., 2014. "Epigenomic alterations define lethal CIMP-positive ependymomas of infancy.". *Nature* **506**(7489): 445–50.

Mansur, D.B., Perry, A., et al., 2005. "Postoperative radiation therapy for grade II and III intracranial ependymoma.". *Int. J. Radiat. Oncol. Biol. Phys.* **61**(2): 387–91.

Mao, Y., Qu, Q., et al., 2016. "The prognostic value of tumor-infiltrating lymphocytes in breast cancer: A systematic review and meta-analysis". *PLoS One* **11**(4): 1–13.

Margulies, M., Egholm, M., et al., 2005. "Genome sequencing in microfabricated high-density picolitre reactors.". *Nature* **437**(7057): 376–380.

Marinoff, A.E., Ma, C., et al., 2017. "Rethinking childhood ependymoma: a

retrospective, multi-center analysis reveals poor long-term overall survival". *J. Neurooncol.* **135**(1): 201–211.

Marioni, J.C., Mason, C.E., et al., 2008. "RNA-seq : An assessment of technical reproducibility and comparison with gene expression arrays". *Genome Res.* **18**(9): 1509–1517.

Massimino, M., Gandola, L., et al., 2004. "Hyperfractionated radiotherapy and chemotherapy for childhood ependymoma: Final results of the first prospective AIEOP (Associazione Italiana di Ematologia-Oncologia Pediatrica) study". *Int. J. Radiat. Oncol. Biol. Phys.* **58**(5): 1336–1345.

Massimino, M., Gandola, L., et al., 2011. "Infant ependymoma in a 10-year AIEOP (Associazione Italiana Ematologia Oncologia Pediatrica) experience with omitted or deferred radiotherapy". *Int. J. Radiat. Oncol. Biol. Phys.* **80**(3): 807–814.

Massimino, M., Miceli, R., et al., 2016. "Final results of the second prospective AIEOP protocol for pediatric intracranial ependymoma". *Neuro. Oncol.* **18**(10): 1451–1460.

Masuda, N., Ohnishi, T., et al., 1999. "Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples". *Nucleic Acids Res.* **27**(22): 4436–4443.

Maybruck, B.T., Pfannenstiel, L.W., et al., 2017. "Tumor-derived exosomes induce CD8+ T cell suppressors". *J. Immunother. Cancer* **5**(1): 65.

McGuire, C.S., Sainani, K.L., et al., 2009. "Both Location and Age Predict Survival in Ependymoma: A SEER Study". *Pediatr. Blood Cancer* **52**(1): 65–69.

Medrano, R.F. V, Hunger, A., et al., 2017. "Immunomodulatory and antitumor effects of type I interferons and their application in cancer therapy". *Oncotarget* **8**(41): 71249–71284.

Mendrzyk, F., Korshunov, A., et al., 2006. "Identification of gains on 1q and epidermal growth factor receptor overexpression as independent prognostic markers in intracranial ependymoma". *Clin. Cancer Res.* **12**(16): 2070–2079.

Merchant, T.E., Boop, F.A., et al., 2008. "A Retrospective Study of Surgery and Reirradiation for Recurrent Ependymoma". *Int. J. Radiat. Oncol.* **71**(1): 87–97.

Merchant, T.E., Li, C., et al., 2009. "Conformal radiotherapy after surgery for paediatric ependymoma: a prospective study.". *Lancet Oncol.* **10**(3): 258–66.

Messahel, B., Ashley, S., et al., 2009. "Relapsed intracranial ependymoma in children in the UK: patterns of relapse, survival and therapeutic outcome.".

Eur. J. Cancer **45**(10): 1815–23.

Minn, A.J., 2015. "Interferons and the Immunogenic Effects of Cancer Therapy". *Trends Immunol.* **36**(11): 725–737.

Mittempergher, L., de Ronde, J.J., et al., 2011. "Gene expression profiles from formalin fixed paraffin embedded breast cancer tissue are largely comparable to fresh frozen matched tissue". *PLoS One* **6**(2): e17163.

Modena, P., Lualdi, E., et al., 2006. "Identification of tumor-specific molecular signatures in intracranial ependymoma and association with clinical characteristics.". *J. Clin. Oncol.* **24**(33): 5223–33.

Mohankumar, K.M., Currle, D.S., et al., 2015. "An in vivo screen identifies ependymoma oncogenes and tumor-suppressor genes". *Nat. Genet.* **47**(8): 878–887.

Morlan, J.D., Qu, K., et al., 2012. "Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue". *PLoS One* **7**(8): 1–8.

Morris, T., Teschendorff, A., et al., 2014. "ChAMP: 450k Chip Analysis Methylation Pipeline". *Bioinformatics* **30**(3): 428–430.

Mortazavi, A., Williams, B. a, et al., 2008. "Mapping and quantifying mammalian transcriptomes by RNA-Seq.". *Nat. Methods* **5**(7): 621–628.

Morton, M., Bai, X., et al., 2014. "Identification of mRNAs and lincRNAs associated with lung cancer progression using next-generation RNA sequencing from laser micro-dissected archival FFPE tissue specimens" **85**(1): 31–39.

Morwood, S.R., Nicholson, L.B., 2006. "Modulation of the immune response by extracellular matrix proteins". *Arch. Immunol. Ther. Exp. (Warsz).* **54**(6): 367–374.

Muldoon, L.L., Alvarez, J.I., et al., 2013. "Immunologic privilege in the central nervous system and the blood-brain barrier". *J. Cereb. Blood Flow Metab.* **33**(1): 13–21.

Nahta, R., Esteva, F.J., 2003. "HER-2-Targeted Therapy : Lessons Learned and Future Directions". *Clin. Cancer Res.* **9**(14): 5078–5084.

Nambirajan, A., Sharma, M.C., et al., 2014. "Study of Stem cell marker Nestin and its correlation with Vascular endothelial growth factor and microvascular density in Ependymomas.". *Neuropathol. Appl. Neurobiol.* **40**(6): 714–725.

Northcott, P. a, Pfister, S.M., et al., 2015. "Next-generation (epi)genetic drivers of childhood brain tumours and the outlook for targeted therapies". *Lancet Oncol.* **16**(6): e293–e302.

Norton, N., Sun, Z., et al., 2013. "Gene expression, single nucleotide variant and fusion transcript discovery in archival material from breast tumors". *PLoS*

One **8**(11): 1–11.

O'Rourke, D.M., Nasrallah, M.P., et al., 2017. "A single dose of peripherally infused EGFRvIII-directed CAR T cells mediates antigen loss and induces adaptive resistance in patients with recurrent glioblastoma". *Sci. Transl. Med.* **9**(399): eaaa0984.

Pages, H., Carlson, M., et al., 2017. "AnnotationDbi: Annotation Database Interface".

Pajtler, K.W., Lin, T., et al., 2017. "EPN31: Molecular refinement of pediatric posterior fossa ependymoma". *Neuro. Oncol.* **18**: iii30-iii39.

Pajtler, K.W., Witt, H., et al., 2015. "Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups". *Cancer Cell* **27**(5): 728–743.

Palm, T., Figarella-Branger, D., et al., 2009. "Expression profiling of ependymomas unravels localization and tumor grade-specific tumorigenesis". *Cancer* **115**(17): 3955–3968.

Park, B., Yee, C., et al., 2014. "The effect of radiation on the immune response to cancers". *Int. J. Mol. Sci.* **15**(1): 927–943.

Parker, B.S., Rautela, J., et al., 2016. "Antitumour actions of interferons: implications for cancer therapy.". *Nat. Rev. Cancer* **16**(3): 131–44.

Parker, M., Mohankumar, K.M., et al., 2014. "C11orf95-RELA fusions drive oncogenic NF-κB signalling in ependymoma.". *Nature* **506**(7489): 451–5.

Paulino, A.C., Wen, B.-C., et al., 2002. "Intracranial ependymomas: an analysis of prognostic factors and patterns of failure.". *Am. J. Clin. Oncol.* **25**(2): 117–122.

Perica, K., Varela, J.C., et al., 2015. "Adoptive T Cell Immunotherapy For Cancer". *Rambam Maimonides Med. J.* **6**(1): e0004.

Perilongo, G., Massimino, M., et al., 1997. "Analyses of prognostic factors in a restrospective review of 92 children with ependymoma: Italian Pediatric Neuro-oncology Group". *Med. Pediatr. Oncol.* **29**(2): 79–85.

Peris-Bonet, R., Martínez-García, C., et al., 2006. "Childhood central nervous system tumours--incidence and survival in Europe (1978-1997): report from Automated Childhood Cancer Information System project.". *Eur. J. Cancer* **42**(13): 2064–80.

Perry, V.H., Teeling, J., 2013. "Microglia and macrophages of the central nervous system: The contribution of microglia priming and systemic inflammation to chronic neurodegeneration". *Semin. Immunopathol.* **35**(5): 601–612.

Peyre, M., Commo, F., et al., 2010. "Portrait of ependymoma recurrence in children: biomarkers of tumor progression identified by dual-color

microarray-based gene expression analysis.". *PLoS One* **5**(9): e12932.

Pfaffl, M.W., 2001. "A new mathematical model for relative quantification in". *Nucleic Acids Res.* **29**(9): 2003–2007.

Pfirschke, C., Engblom, C., et al., 2016. "Immunogenic Chemotherapy Sensitizes Tumors to Checkpoint Blockade Therapy". *Immunity* **44**(2): 343–354.

Pietsch, T., Wohlers, I., et al., 2014. "Supratentorial ependymomas of childhood carry C11orf95-RELA fusions leading to pathological activation of the NF-κB signaling pathway.". *Acta Neuropathol.* **127**(4): 609–11.

Pollack, I., Gerszten, P.C., et al., 1995. "Intracranial Ependymomas of Childhood: Long Term outcome and Prognostic Factors". *Neurosurgery* **37**(4): 655–667.

Puget, S., Grill, J., et al., 2009. "Candidate genes on chromosome 9q33-34 involved in the progression of childhood ependymomas". *J. Clin. Oncol.* **27**(11): 1884–1892.

R Core Team, 2014. "R: A language and environment for statistical computing.". *R Found. Stat. Comput. Vienna, Austria*.

Raman, D., Baugher, P.J., et al., 2007. "Role of chemokines in tumor growth". *Cancer Lett.* **256**(2): 137–165.

Ramaswamy, V., Hielscher, T., et al., 2016. "Therapeutic Impact of Cytoreductive Surgery and Irradiation of Posterior Fossa Ependymoma in the Molecular Era : A Retrospective Multicohort Analysis" **34**(21): 2468–2477.

Ransohoff, R.M., Kivisäkk, P., et al., 2003. "Three or more routes for leukocyte migration into the central nervous system". *Nat. Rev. Immunol.* **3**(7): 569–581.

Rau, A., Marot, G., et al., 2014. "Differential meta-analysis of RNA-seq data from multiple studies". *BMC Bioinformatics* **15**(1): 91.

Reardon, D. a., Entrekin, R.E., et al., 1999. "Chromosome arm 6q loss is the most common recurrent autosomal alteration detected in primary pediatric ependymoma". *Genes Chromosom. Cancer* **24**(3): 230–237.

Reeves, G., Todd, I., 2004. Lecture Notes on Immunology. Blackwell Publishing, Oxford.

Rickert, C.H., Paulus, W., 2001. "Epidemiology of central nervous system tumors in childhood and adolescence based on the new WHO classification.". *Childs. Nerv. Syst.* **17**(9): 503–11.

Ridley, L., Rahman, R., et al., 2008. "Multifactorial analysis of predictors of outcome in pediatric intracranial ependymoma.". *Neuro. Oncol.* **10**(5): 675–89.

Robasky, K., Lewis, N.E., et al., 2014. "The role of replicates for error mitigation in next-generation sequencing.". *Nat. Rev. Genet.* **15**(1): 56–62.

Robertson, P.L., Zeltzer, P.M., et al., 1998. "Survival and prognostic factors following radiation therapy and chemotherapy for ependymomas in children: a report of the Children's Cancer Group.". *J. Neurosurg.* **88**(4): 695–703.

Robinson, M.D., McCarthy, D.J., et al., 2010. "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data". *Bioinformatics* **26**(1): 139–140.

Rooney, M.S., Shukla, S.A., et al., 2015. "Molecular and genetic properties of tumors associated with local immune cytolytic activity". *Cell* **160**(1–2): 48–61.

Rouschop, K.M.A., 2005. "Protection against Renal Ischemia Reperfusion Injury by CD44 Disruption". *J. Am. Soc. Nephrol.* **16**(7): 2034–2043.

Rousseau, A., Idbaih, A., et al., 2010. "Specific chromosomal imbalances as detected by array CGH in ependymomas in association with tumor location, histological subtype and grade". *J. Neurooncol.* **97**(3): 353–364.

Royal College of Paediatrics and Child Health, National Children's Bureau, 2014. Why children die: Death in infants, children and young people in the UK Part B.

Rudà, R., Reifenberger, G., et al., 2017. "EANO guidelines for the diagnosis and treatment of ependymal tumors". *Neuro. Oncol.* 1–12.

Sanger, F., Coulson, A.R., 1975. "A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.". *J. Mol. Biol.* **94**(3): 441–448.

Sanger, F., Nicklen, S., et al., 1977. "DNA sequencing with chain-terminating inhibitors.". *Proc. Natl. Acad. Sci. U. S. A.* **74**(12): 5463–5467.

Sayour, E.J., Mitchell, D.A., 2017. "Immunotherapy for pediatric brain tumors". *Brain Sci.* **7**(10).

Schaefer, L., 2010. "Extracellular matrix molecules: Endogenous danger signals as new drug targets in kidney diseases". *Curr. Opin. Pharmacol.* **10**(2): 185–190.

Schroeder, A., Mueller, O., et al., 2006. "The RIN: an RNA integrity number for assigning integrity values to RNA measurements.". *BMC Mol. Biol.* **7**: 3.

Schuster, S.C., 2008. "Next-generation sequencing transforms today's biology.". *Nat. Methods* **5**(1): 16–18.

Schwalbe, E.C., Lindsey, J.C., et al., 2017. "Novel molecular subgroups for clinical classification and outcome prediction in childhood medulloblastoma: a cohort study". *Lancet Oncol.* **18**(7): 958–971.

Scotting, P.J., Walker, D. a, et al., 2005. "Childhood solid tumours: a developmental disorder". *Nat. Rev. Cancer* **5**(6): 481–488.

Serrano, A., Castro-Vega, I., et al., 2011. "Role of Gene Methylation in Antitumor Immune Response: Implication for Tumor Progression". *Cancers (Basel).* **3**(2): 1672–1690.

Shanmugavadivel, D., Walker, D., et al., 2015. "HeadSmart: are you brain tumour aware?". *Paediatr. Child Health (Oxford).* **26**(2): 81–86.

Sharma, P., Hu-Lieskovan, S., et al., 2017. "Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy". *Cell* **168**(4): 707–723.

Sharma, P., Wagner, K., et al., 2011. "Novel cancer immunotherapy agents with survival benefits: recent success and next steps". *Nat. Rev. Cancer* **11**(11): 805–812.

Shi, W., 2014. "Rsubread package : high-performance read alignment , quantification and mutation discovery Read alignment" 1–10.

Shortman, K., 2000. "Dendritic cells: Multiple subtypes, multiple origins, multiple functions". *Immunol Cell Biol* **78**(2): 161–165.

Shortman, R.I., Lowis, S.P., et al., 2014. "Cognitive Function in Children With Brain Tumors in the First Year After Diagnosis Compared to Healthy Matched Controls". *Pediatr. Blood Cancer* **61**(3): 464–472.

Silginer, M., Nagy, S., et al., 2017. "Autocrine activation of the IFN signaling pathway may promote immune escape in glioblastoma". *Neuro. Oncol.* **19**(10): 1338–1349.

Simpson, A.J.G., Caballero, O.L., et al., 2005. "Cancer/testis antigens, gametogenesis and cancer". *Nat. Rev. Cancer* **5**(8): 615–625.

Sinicropi, D., Qu, K., et al., 2012. "Whole transcriptome RNA-seq analysis of breast cancer recurrence risk using formalin-fixed paraffin-embedded tumor tissue". *PLoS One* **7**(7): 1–10.

Snider, C.A., Yang, K., et al., 2017. "Impact of radiation therapy and extent of resection for ependymoma in young children: A population-based study.". *Pediatr. Blood Cancer* **65**(3): e26880.

Sowar, K., Straessle, J., et al., 2006. "Predicting which children are at risk for ependymoma relapse.". *J. Neurooncol.* **78**(1): 41–6.

Spiegler, B.J., Bouffet, E., et al., 2004. "Change in neurocognitive functioning after treatment with cranial radiation in childhood.". *J. Clin. Oncol.* **22**(4): 706–13.

Spranger, S., 2016. "Mechanisms of tumor escape in the context of the T-cell-inflamed and the non-T-cell-inflamed tumor microenvironment". *Int. Immunol.* **28**(8): 383–391.

Steliarova-Foucher, E., Stiller, C., et al., 2004. "Geographical patterns and time trends of cancer incidence and survival among children and adolescents in

Europe since the 1970s (the ACCISproject): an epidemiological study.". *Lancet* **364**(9451): 2097–2105.

Steliarova-Foucher, E., Stiller, C., et al., 2005. "International classification of childhood cancer, third edition". *Cancer* **103**(7): 1457–1467.

Strother, D.R., Lafay-Cousin, L., et al., 2014. "Benefit from prolonged dose-intensive chemotherapy for infants with malignant brain tumors is restricted to patients with ependymoma: a report of the Pediatric Oncology Group randomized controlled trial 9233/34.". *Neuro. Oncol.* **16**(3): 457–65.

Sturm, D., Orr, B.A., et al., 2016. "New Brain Tumor Entities Emerge from Molecular Classification of CNS-PNETs Resource New Brain Tumor Entities Emerge from Molecular Classification of CNS-PNETs". *Cell* **164**(5): 1060–1072.

Subramanian, A., Tamayo, P., et al., 2005. "GSEA : Gene set enrichment analysis Gene set enrichment analysis : A knowledge-based approach for interpreting genome-wide expression profiles". *PNAS* **102**(43): 15545-1555-.

Supek, F., Bošnjak, M., et al., 2011. "Revigo summarizes and visualizes long lists of gene ontology terms". *PLoS One* **6**(7).

Tabori, U., Wong, V., et al., 2008. "Telomere maintenance and dysfunction predict recurrence in paediatric ependymoma.". *Br. J. Cancer* **99**(7): 1129–1135.

Tarazona, S., García-Alcalde, F., et al., 2011. "Differential expression in RNA-seq: A matter of depth". *Genome Res.* **21**(12): 2213–2223.

Taylor, M.D., Northcott, P. a, et al., 2012. "Molecular subgroups of medulloblastoma: the current consensus.". *Acta Neuropathol.* **123**(4): 465–72.

Taylor, M.D., Poppleton, H., et al., 2005. "Radial glia cells are candidate stem cells of ependymoma.". *Cancer Cell* **8**(4): 323–35.

Taylor, T., Jaspan, T., et al., 2008. "Radiological classification of optic pathway gliomas: experience of a modified functional classification system.". *Br. J. Radiol.* **81**(970): 761–6.

Teschendorff, A.E., Marabita, F., et al., 2013. "A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data". *Bioinformatics* **29**(2): 189–196.

Therneau, T., 2015. "Survival: A Package for Survival Analysis in S_".

Tihan, T., Zhou, T., et al., 2008. "The prognostic value of histological grading of posterior fossa ependymomas in children: a Children's Oncology Group study and a review of prognostic factors.". *Mod. Pathol.* **21**: 165–177.

Tsang, D.S., Burghen, E., et al., 2018. "Outcomes After Reirradiation for

Recurrent Pediatric Intracranial Ependymoma". *Int. J. Radiat. Oncol. Biol. Phys.* **100**(2): 507–515.

Venkatramani, R., Ji, L., et al., 2013. "Outcome of infants and young children with newly diagnosed ependymoma treated on the "head Start" III prospective clinical trial". *J. Neurooncol.* **113**(2): 285–291.

Venter, J., Smith, H., et al., 1996. "A new strategy for genome sequencing.". *Nature* **381**(6581): 364–366.

Venter, J.C., Adams, M.D., et al., 2001. "The sequence of the human genome.". *Science* **291**(5507): 1304–1351.

Vinchon, M., Leblond, P., et al., 2005. "Intracranial ependymomas in childhood: Recurrence, reoperation, and outcome". *Child's Nerv. Syst.* **21**(3): 221–226.

von Ahlfen, S., Missel, A., et al., 2007. "Determinants of RNA quality from FFPE samples". *PLoS One* **2**(12): 1–7.

Vora, A., Goulden, N., et al., 2013. "Treatment reduction for children and young adults with low-risk acute lymphoblastic leukaemia defined by minimal residual disease (UKALL 2003): a randomised controlled trial.". *Lancet. Oncol.* **14**(3): 199–209.

Vukmirovic, M., Herazo-Maya, J.D., et al., 2017. "Identification and validation of differentially expressed transcripts by RNA-sequencing of formalin-fixed, paraffin-embedded (FFPE) lung tissue from patients with Idiopathic Pulmonary Fibrosis". *BMC Pulm. Med.* **17**(1): 15.

Wang, C., Gong, B., et al., 2014. "The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance". *Nat. Biotechnol.* **32**(9): 926–935.

Wang, Z., Gerstein, M., et al., 2009. "RNA-Seq: a revolutionary tool for transcriptomics.". *Nat. Rev. Genet.* **10**(1): 57–63.

Wani, K., Armstrong, T.S., et al., 2012. "A prognostic gene expression signature in infratentorial ependymoma.". *Acta Neuropathol.* **123**(5): 727–38.

Warburg, O., 1956. "On the Origin of Cancer Cells". *Science (80-.).* **123**(3191): 309–314.

Weichselbaum, R.R., Ishwaran, H., et al., 2008. "An interferon-related gene signature for DNA damage resistance is a predictive marker for chemotherapy and radiation for breast cancer.". *Proc. Natl. Acad. Sci. U. S. A.* **105**(47): 18490–5.

Weinberg, R.A., Hahn, W.C., et al., 1999. "Creation of human tumour cells with defined genetic elements.". *Nature* **400**(6743): 464–468.

Wetmore, C., Daryani, V.M., et al., 2016. "Phase II evaluation of sunitinib in the treatment of recurrent or refractory high-grade glioma or ependymoma in

children: a children's Oncology Group Study ACNS1021.". *Cancer Med.* **5**(7): 1416–1424.

Wetterstrand, K., n.d. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) [WWW Document]. URL www.genome.gov/sequencingcosts (accessed 4.9.15).

Whiteside, T.L., 2008. "The tumor microenvironment and its role in promoting tumor growth". *Oncogene* **27**(45): 5904–5912.

Williams, C., Pontén, F., et al., 1999. "A high frequency of sequence alterations is due to formalin fixation of archival specimens.". *Am. J. Pathol.* **155**(5): 1467–1471.

Williams, C.R., Baccarella, A., et al., 2016. "Trimming of sequence reads alters RNA-Seq gene expression estimates". *BMC Bioinformatics* **17**: 103.

Wilne, S., Koller, K., et al., 2010. "The diagnosis of brain tumours in children: a guideline to assist healthcare professionals in the assessment of children who may have a brain tumour.". *Arch. Dis. Child.* **95**(7): 534–539.

Witt, D.A., Donson, A.M., et al., 2018. "Specific expression of PD-L1 in RELA-fusion supratentorial ependymoma: Implications for PD-1-targeted therapy". *Pediatr. Blood Cancer* e26960.

Witt, H., Mack, S.C., et al., 2011. "Delineation of two clinically and molecularly distinct subgroups of posterior fossa ependymoma.". *Cancer Cell* **20**(2): 143–57.

Woo, S.-R., Corrales, L., et al., 2015. "Innate Immune Recognition of Cancer". *Annu. Rev. Immunol.* **33**(1): 445–474.

Wright, K.D., Daryani, V.M., et al., 2015. "Phase I study of 5-fluorouracil in children and young adults with recurrent ependymoma". *Neuro. Oncol.* **17**(12): 1620–1627.

Xiao, Y.L., Kash, J.C., et al., 2013. "High-throughput RNA sequencing of a formalin-fixed, paraffin-embedded autopsy lung tissue sample from the 1918 influenza pandemic". *J. Pathol.* **229**(4): 535–545.

Yaddanapudi, K., Mitchell, R.A., et al., 2013. "Cancer vaccines". *Oncoimmunology* **2**(3): e23403.

Yakovleva, A., Plieskatt, J.L., et al., 2017. "Fit for genomic and proteomic purposes: Sampling the fitness of nucleic acid and protein derivatives from formalin fixed paraffin embedded tissue.". *PLoS One* **12**(7): e0181756.

Yeung, Y.T., McDonald, K.L., et al., 2013. "Interleukins in glioblastoma pathophysiology: Implications for therapy". *Br. J. Pharmacol.* **168**(3): 591–606.

Yock, T.I., Bhat, S., et al., 2014. "Quality of life outcomes in proton and photon

treated pediatric brain tumor survivors.". *Radiother. Oncol.* **113**(1): 89–94.

Yu, S., Li, A., et al., 2017. "Chimeric antigen receptor T cells: a novel therapy for solid tumors". *J. Hematol. Oncol.* **10**(1): 78.

Zacharoulis, S., Ashley, S., et al., 2010. "Treatment and outcome of children with relapsed ependymoma: A multi-institutional retrospective analysis". *Child's Nerv. Syst.* **26**(7): 905–911.

Zhao, W., He, X., et al., 2014. "Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling.". *BMC Genomics* **15**(1): 419.

Zheng, P.P., Pang, J.C.S., et al., 2000. "Comparative genomic hybridization detects losses of chromosomes 22 and 16 as the most common recurrent genetic alterations in primary ependymomas". *Cancer Genet. Cytogenet.* **122**: 18–25.

Zheng, Y., Zha, Y., et al., 2013. "Egr2-dependent gene expression profiling and ChIP-Seq reveal novel biologic targets in T cell anergy". *Mol. Immunol.* **55**(3–4): 283–291.

Zitvogel, L., Galluzzi, L., et al., 2015. "Type I interferons in anticancer immunity". *Nat Rev Immunol* **15**(7): 405–414.

Zitvogel, L., Tesniere, A., et al., 2006. "Cancer despite immunosurveillance: Immunoselection and immunosubversion". *Nat. Rev. Immunol.* **6**(10): 715–727.

Appendix 1: Samples used in biological analyses

Appendix 1, Table 1: Summary of all clinical cases including in any of the biological analyses (RNA-seq and/or DNA methylation and/or IHC).

Study ID	Paired RNA-Seq	Paired Methylation	Paired IHC	RNA-Seq Group	Methylation (Classifier score)		Gender	Location
					P	R1		
Epend003	FFPE	YES	YES	PF1	PFA (1)	PFA (1)	M	PF
Epend018	NO	NO	YES				M	ST
Epend022	NO	NO	YES		PFA (1)		M	PF
Epend029	NO	YES	YES		RELA (1)	RELA (1)	M	ST
Epend030	NO	NO	YES		PFA (1)		M	PF
Epend033	FFPE	NO	YES	PF1		PFA (1)	M	PF
Epend034	NO	NO	YES		PFA (1)		F	PF
Epend050	NO	YES	NO		DNET (1)	RELA (1)	M	ST
Epend051	FFPE	NO	NO	PF1	PFA (1)		F	PF
Epend078	NO	YES	NO		PFA (0.5)	PFA (0.96)	M	PF
Epend085	FFPE	NO	YES	PF2	PFA (0.98)		F	PF
Epend092	FF	NO	YES	PF1			F	ST
Epend095	FFPE	YES	YES	PF1	PFA (0.76)	PFA (0.84)	F	PF
Epend096	NO	NO	YES			SPINE (0.51)	M	ST
Epend097	FF	YES	YES	PF2	PFA (1)	PFA (0.95)	F	PF
Epend098	NO	YES	YES		PFA (0.89)	PFA (0.91)	F	PF
Epend101	NO	YES	NO		PFA (0.98)	PFA (0.57)	M	PF
Epend103	NO	YES	YES		YAP (1)	YAP (1)	F	ST
Epend114	FF	NO	YES	ST			M	ST
Epend115	NO	NO	YES			PFA (1)	F	PF
Epend118	FFPE	YES	YES	ST	RELA (1)	RELA (1)	M	ST
Epend121	FF	YES	NO	PF1	PFA (0.97)	PFA (1)	M	PF
Epend122	FF/FFPE	YES	NO	PF1	PFA (1)	PFA (1)	F	PF
Epend123	FF	YES	NO	PF2	HGNET_MN1 (1)	HGNET_MN1 (1)	M	SP
Epend124	FF/FFPE	YES	YES	PF2	YAP (1)	YAP (1)	F	ST
Epend129	FF	YES	YES	PF2	MPE (1)	MPE (1)	F	SP
Epend136	NO	YES	YES		PFA (1)	PFA (0.99)	F	ST
Epend138	NO	NO	YES			PFA (0.99)	M	PF
Epend140	FF	NO	YES	ST			F	ST
Epend147	NO	NO	YES		PFA (0.71)		F	PF
Epend161	FF	YES	NO	PF2	PFA (1)	PFA (1)	F	PF

Study ID	Paired RNA-Seq	Paired Methylation	Paired IHC	RNA-Seq Group	Methylation (Classifier score)		Gender	Location
Epend162	FF	YES	NO	ST	PFA (1)	PFA (0.99)	F	PF
Epend163	NO	YES	NO		PFA (1)	PFA (1)	M	PF
Epend176	NO	YES	NO		PFA (1)	PFA (0.87)	M	PF
Epend182	NO	YES	NO		RELA (1)	RELA (1)	F	ST
Epend183	NO	NO	YES		PFA (1)		M	PF
Epend185	NO	NO	YES				F	ST
Epend186	NO	YES	YES		PFA (1)	PFA (1)	M	PF
Epend190	NO	NO	YES			PFA (1)	M	PF
Epend193	FF	YES	YES	PF2	PFA (1)	PFA (0.95)	M	PF
Epend195	FFPE	YES	YES	PF1	PFA (1)	PFA (0.99)	M	PF
Epend196	NO	YES	YES		PFA (0.99)	PFA (1)	F	PF
Epend197	FFPE	YES	YES	PF1	PFA (1)	PFA (1)	M	PF
Epend203	FFPE	YES	YES	PF1	PFA (0.98)	PFA (0.97)	F	PF
Epend205	FFPE	YES	YES	PF1	PFA (0.93)	PFA (0.98)	M	PF
Epend207	NO	NO	YES		PFA (0.99)		F	PF
Epend208	FF	NO	NO	PF1			M	PF
Epend209	NO	YES	YES		RELA (1)	RELA (1)	F	ST
Epend211	FFPE	YES	NO	PF1	PFA (0.95)	PFA (0.97)	M	PF
Epend213	FFPE	YES	YES	PF1	PFA (0.97)	PFA (1)	F	PF
Epend224	NO	YES	YES		RELA (1)	RELA (1)	M	ST
Epend230	FFPE	YES	YES	PF1	PFA (1)	PFA (0.94)	M	PF
Epend234	FFPE	YES	YES	PF2	PFA (0.99)	PFA (0.95)	M	PF
Epend244	NO	YES	NO		PFA (1)	PFA (0.88)	M	PF
Epend245	NO	NO	YES		PFA (0.99)		M	PF
Epend256	NO	YES	YES		PFA (0.98)	PFA (0.94)	M	PF
Epend258	FFPE	YES	YES	ST	RELA (1)	RELA (1)	M	ST
Epend277	NO	YES	NO		PFA (1)	PFA (1)	F	PF
Epend279	NO	NO	YES					PF
Epend280	NO	YES	YES		RELA (1)	RELA (1)		ST
Epend281	NO	NO	YES					PF
Epend287	FFPE	NO	YES	PF1	PFA (0.58)		M	PF
Epend289	FF	YES	NO	PF2	PFA (1)	PFA (1)	M	ST
Epend293	FFPE	YES	YES	ST	HGNET_MN1 (1)	HGNET_MN1 (1)	M	PF
Epend300	NO	NO	YES				F	PF
Epend305	FFPE	YES	YES	PF1	PFA (0.84)	PFA (0.9)	M	ST
Epend306	FFPE	YES	YES	ST	RELA (0.98)	RELA (0.84)	F	ST
Epend316	FFPE	YES	YES	PF2	PFB (0.79)	PFB (0.78)	M	PF
Epend330	FFPE	YES	NO	ST	RELA (1)	RELA (1)	F	ST

Study ID	Paired RNA-Seq	Paired Methylation	Paired IHC	RNA-Seq Group	Methylation (Classifier score)		Gender	Location
Epend331	FFPE	YES	NO	PF1	PFA (1)	PFA (1)	F	PF
Epend332	NO	YES	NO		PFA (1)	DNET (0.76)	M	PF
Epend333	FFPE	NO	NO	ST	HGNET_BCOR (1)		M	PF
Epend334	NO	YES	NO		PFA (0.75)	PFA (0.66)	F	PF
Epend335	FF	NO	NO	PF2			M	PF
Epend336	FF	NO	NO	PF1		PFA (0.51)	F	PF
Epend337	FF	YES	NO	PF1	PFA (0.68)	PFA (0.74)	F	PF
Epend338	FF	NO	NO	PF2	PFA (0.74)		M	PF
Epend339	FF	NO	NO	PF1			M	PF
Epend340	FF	YES	NO	PF1	PFA (0.75)	PFA (0.59)	M	PF
Epend341	FF	NO	NO	PF2				PF
Epend342	FF	NO	NO	PF2			M	PF
Epend343	FF	YES	NO	PF1	PFA (0.52)	PFA (0.80)	M	PF
Epend344	FF	NO	NO	PF1			M	PF
Epend345	NO	YES	NO		PFA (0.97)	PFA (1)	F	PF
Epend346	FF	NO	NO	PF2			M	PF
Epend347	FF	YES	NO	PF1	PFA (1)	PFA (1)	M	PF
Epend348	FF	YES	NO	PF2	PFA (0.78)	PFA (0.58)	F	PF
Epend357	NO	NO	YES				F	ST
Epend371	NO	YES	NO		PFA (0.99)	PFA (1)	F	PF
Epend372	NO	YES	NO		RELA (1)	RELA (1)	F	ST
Epend381	NO	YES	YES		RELA (0.99)	RELA (1)		ST
Epend400	NO	YES	YES		PFA (1)	PFA (1)	F	PF
Epend401	NO	NO	YES					
Epend402	NO	NO	YES					

Appendix 1, Table 2, Page 1: Clinical features of the cases included in any of the biological analyses (RNA=Seq and/or DNA Methylation and/or IHC). RT: Radiotherapy. S: Surgery. CT: Chemotherapy. F/U: Follow-up.

Study ID	Relapses	Age	TTP	Status	F/U Duration	Grade		RT		S		CT		Resection		Mets	
						P	R1	P	R1	P	R1	P	R1	P	R1	P	R1
Epend003	2	22	31	A	131	II	III	N	Y	Y	Y	Y		GTR	GTR	Y	
Epend018	1	68	124	A	162	III		Y	Y	Y	Y	N	N	GTR	GTR	N	N
Epend022	5	34	4	A	245	II	II	N	Y	Y	Y	Y		STR	STR		
Epend029	4	115	2	D	68	III	III	Y	Y	Y	Y	N		GTR			Y
Epend030	3	35	27	D	80	III	III	N	Y	Y	Y	Y		GTR	GTR		
Epend033	3	25	2	D	19	III	III	N	Y	Y	Y	Y		STR		Y	
Epend034	2	30	8	A	83	II		N	Y	Y	Y	Y		GTR	STR		
Epend050	1	152	4	A	155	III	III	N	Y	Y	Y	Y	N	STR			
Epend051	2	30	27	D	36	III	III	Y	Y	Y	Y	Y		STR	GTR		
Epend078	3	104	35	D	68	II	II	Y		Y	Y	Y		GTR	STR		
Epend085	2	39	17	A	31	III		Y		Y	Y	N		STR			
Epend092	2	46	26	A	338	II	II	Y		Y		Y		STR		N	
Epend095	1	24	104	D	113	III	III	N	Y	Y	Y	Y		GTR		N	N
Epend096	1	67	9	D	12	III		Y		Y		Y		STR			Y
Epend097	2	145	26	A	158	III		Y		Y	Y	Y		STR	STR	N	N
Epend098	2	68	149	D	173	III	III	N	Y	Y	Y	Y	N	STR	GTR	N	N
Epend101	4	54	17	D	68	II		N	Y	Y	Y	Y		STR	STR	Y	
Epend103	1	17	4	A	260	III	III	N	N	Y	Y	Y		STR			
Epend114	1	50	32	D	41	III	III	N	N	Y	Y	Y		STR			
Epend115	1	72	4	D	17	II		N	N	Y	Y	Y		STR			N
Epend118	3	162	23	D	110	II	II	N	Y	Y	Y	Y	Y	STR			
Epend121	1	13	16	A	84	III	II	N	Y	Y	Y	Y		GTR	STR	N	N
Epend122	1	39	2	A	69	III	II	N	Y	Y	Y	Y		STR	STR	N	N
Epend123	3	50	11	D	63	III	III	Y	Y	Y	Y	N	Y	GTR	STR	N	Y
Epend124	2	3	15	A	62	III	III	N	N	Y	Y	Y	N	STR	STR	N	N
Epend129	3	72	9	D	63	II	II	N	Y	Y	Y	N	Y	GTR		N	N
Epend136	5	17		D	154	II	III	Y	Y	Y	Y	Y		STR			
Epend138	1	65	27	A	51	III	III	Y		Y	Y	N					
Epend140	3	84	17	D	67	III		Y		Y	Y	N		GTR			
Epend147	4	178	17	D	30	III		Y		Y	Y			GTR			
Epend161	3	23	19	A	95	II	II	N	Y	Y	Y	Y	N	GTR	GTR		
Epend162	1	35	8	D	29	II	III	N	Y	Y	Y	Y	Y	GTR	STR	N	N
Epend163	2	42	27	D	48	II	II	Y	N	Y	Y	N	Y	STR	STR	N	N
Epend176	4	21	13	D	43	III		N	Y	Y	Y	Y		STR	GTR	N	N

Study ID	Relapses	Age	TTP	Status	F/U Duration	Grade		RT		S		CT		Resection		Mets	
Epend182	8	74	23	D	61	III		Y	N	Y	Y	N	N	GTR	GTR	N	N
Epend183	1	23	29	A	259	II	III	N	Y	Y	Y	Y		STR	GTR	N	N
Epend185	5	13	46	A	241	II		N	N	Y	Y	Y		STR	STR		
Epend186	4	15	40	D	106	III	III	N	Y	Y	Y	Y		STR	STR		
Epend190	1	22	16	A	225	II	II	N	Y	Y	Y	Y		STR	GTR	N	N
Epend193	2	26	67	D	77	II	III	N	Y	Y	Y	Y	Y	GTR	GTR	N	N
Epend195	2	17	4	D	6	III	III	N	N	Y	Y	Y		GTR			
Epend196	5	24	27	D	71	II	III	N	Y	Y	Y	Y		GTR	GTR	N	N
Epend197	4	16	16	D	34	II	II	N	Y	Y	Y	Y		GTR	STR	N	Y
Epend203	1	62	0	A	152	II		Y	Y	Y	Y	N		STR			
Epend205	3	52	21	D	42	III	III	Y	Y	Y	Y	N	Y	GTR		N	Y
Epend207	4	77	13	D	37	III		Y		Y	Y	N		STR	GTR	N	N
Epend208	1	16	1	A	68												
Epend209	7	75	4	D	152	II		Y		Y	Y	Y		STR	STR		
Epend211	1	18	122	A	134	II		N	Y	Y	Y	Y		GTR			
Epend213	1	26	68	A	126	II		N	Y	Y	Y	Y		GTR			
Epend224	2	40	12	D	22	III	III	N	Y	Y	Y	Y		STR	STR		
Epend230	4	24	17	D	45	III		N	Y	Y	Y	Y	N	GTR	GTR		
Epend234	2	27	109	D	114	II		N	N	Y	Y	Y		GTR	STR		
Epend244	2	16	54	D	71	II		N	Y	Y	Y	Y		GTR			
Epend245	2	18	18	D	28	II		N	Y	Y	Y	Y		GTR			
Epend256	2	24	9	D	28	II		N	Y	Y	Y	Y		STR			
Epend258	3	10	5	A	99	III	III	N	N	Y	Y	Y		STR		Y	
Epend277	4	20	24	A	151			N	N	Y	Y						
Epend279	2	187	17	D	40	III		Y		Y		Y		STR			
Epend280	2	120	15	D	30	III		Y		Y	Y	N		STR			
Epend281	2	106	77	A	110	II		Y		Y	Y	N		GTR			
Epend287	1	30	34	A	54	III	III	Y	Y	Y	Y	N	Y	GTR	STR	N	Y
Epend289	1	28	16														
Epend293	7	101	47	D	153	II	III	Y	Y	Y	Y	N	Y	GTR			
Epend300	4	152	27	D	64	II		Y	N	Y	Y	Y		STR	STR		
Epend305	2	78	31	D	47	III	III	Y	Y	Y	Y	Y		STR	STR		
Epend306	2	123	19	D	28	II		Y	N	Y	Y	N	Y	GTR	STR		
Epend316	1	199	6	A	145	II	II	Y	N	Y	Y	Y	Y	STR	GTR	N	N
Epend330	5	73	16	A	92	III	III	N	N	Y	Y		Y	GTR			
Epend331	1	18	36	D	37	II	II	N	N	Y	Y			GTR	GTR		
Epend332	4	73	16	D	32	II		Y	Y	Y	Y			GTR	GTR	N	Y
Epend333	3	24	17	D	27	III	III	N	N	Y	Y	Y		GTR	STR		

Study ID	Relapses	Age	TTP	Status	F/U Duration	Grade		RT		S		CT		Resection		Mets	
Epend334	1	57	31	A	135	II	II	Y	Y	Y	Y	N	N	GTR	GTR	N	N
Epend335	1	160	35	D	68	II	II	Y	Y	Y	Y	N	N	GTR	GTR	N	N
Epend336	3	79	19	A	117	II	II	Y	Y	Y	Y	N	N	GTR	GTR	N	Y
Epend337	3	20	50	D	82	III	III	Y	Y	Y	Y	Y	Y	STR	STR	N	N
Epend338	2	89	23	D	47	III	III	Y	Y	Y	Y	N	N	GTR	GTR	N	N
Epend339	2	71	5	D	16	III	III	Y		Y	Y	N	Y	GTR	STR	N	Y
Epend340	3	43	8	D	24			Y	Y	Y	Y	N	N			N	Y
Epend341	1	24	51	A	70			Y	Y	Y	Y	N	N				
Epend342	5	46	9	A	54			Y	Y	Y	Y	N	N	STR	STR		
Epend343	2	32	6	A	53	III	III	Y		Y	Y	Y	Y	STR	GTR	N	N
Epend344	2	8	4	D	56				Y	Y	Y	Y	N				
Epend345	1	7	14	D	41			Y	Y	Y	Y	Y	N				Y
Epend346	2	33	20	D	54			Y	Y	Y	Y	Y	N				
Epend347	2	46	44	A	89			Y	Y	Y	Y	Y	N				
Epend348	1	40	9	A	55				Y	Y	Y	Y	N				
Epend357	3	158	7	A	67		III	Y	N	Y	Y	Y		GTR	GTR		
Epend371	1	119	35	A	35	II										N	N
Epend372	1					II											
Epend381	2					III	III										
Epend400	1			D		III				Y	N			STR		N	Y
Epend401																	
Epend402																	

Appendix 2: File formats

A brief summary of the file formats encountered in the RNA sequencing analysis is outlined below.

File formats: Fastq

Biological data generated from high throughput sequencing reads can be stored as a plain text file in the fastq format. Fastq consists of 4 lines for each Illumina sequencing read arranged as follows:

(1) The @ character followed by the Sequence ID or another title (this follows a structured format for Illumina reads);
(2) Raw sequence letters (e.g. AGTGATAGA…);
(3) + character followed optionally by a repeat of line (1);
(4) Symbols which encode the quality of the reads, this must be equal in length to line (2) (Cock et al., 2010).

In addition to the above, the fastq file can contain a sequence ID in the Illumina format which takes on the following form:

```
@SequenceID: @ABCDE_1234:17:3:2:27:851:1832#0/1
```

This encodes the following:

- ABCDE_1234: ID of the sequencer;
- 17: Run number on sequencer;
- 3: Flow cell ID;
- 2: Flow cell lane;
- 27: Tile number within the flow cell lane;
- 851: X co-ordinate of cluster within the tile;
- 1832: Y co-ordinate of cluster within the tile;
- #0: Index number for multiplex sequencing;
- /1: Membership of a pair of reads (/1 for the first read, /2 for the second read) (Illumina, 2014).

File Format: SAM/BAM

The output of aligning high throughput sequencing data to the transcriptome and/or genome is a BAM or SAM file. These abbreviations stand for Binary Alignment/Map format or Sequence Alignment/Map format respectively. These are text files made up of an optional header followed by an alignment section (The SAM/BAM Format Specification Working Group, 2016). The header section can contain information about the format of the file, the reference sequence and

read groups. The alignment section contains 11 mandatory fields as described in the working group documentation:

- QNAME: Query template name – usually the read name;
- FLAG: Encodes specific properties of the read such as whether it is correctly mapped or whether it is part of a pair. This type of annotation uses a combination of bitwise flags;
- RNAME: Name of the reference sequence which may include the chromosome name;
- POS: The left most mapping position of the first matching base. If the sequence is unmapped POS is set as 0. For SAM files the reference starts at 1 and for BAM files the reference starts at 0;
- MAPQ: The mapping quality;
- CIGAR: The CIGAR string indicates how the bases in the aligned sequence correspond with the reference. Base lengths are combined with an operation (M=alignment match, I=insertion to reference, D=deletion from reference). When interpreted in combination with the POS column the exact details of the alignment can be understood;
- RNEXT: The reference sequence of the next alignment (i.e. the pair of the read if paired reads);
- TLEN: Observed template length. If all segments mapped to the same reference the template length is the number of bases from the left-most mapped base to the right-most mapped base;
- SEQ: The sequence itself;
- QUAL: The quality score for each base in SEQ. This uses the same code as for the .fastq format.

The alignment section also contains a number of optional fields with tags encoding specific parameters. These tags may contain information such as the strand details from the library preparation process or quality data form the sequencing machine itself (The SAM/BAM Format Specification Working Group, 2016).

BAM files are compressed using the BGZF compression format which allows for rapid access to the data. They are binary versions of SAM files and are not human readable.

Appendix 3: Scripts used for bioinformatic analyses

Code was executed in either the R statistical environment or via the HPC interface if other programmes used (e.g. Trimmomatic, Tophat 2, Bowtie, Picard tools).

#Adapter and quality trim (Trimmomatic)
java -jar Trimmomatic-0.35.jar PE -threads 16 4589FFPE-xxx.R1.fastq.gz 4589FFPE-xxx.R2.fastq.gz output_forward_pairedxxxFFPE.fq.gz output_forward_unpairedxxxFFPE.fq.gz output_reverse_pairedxxxFFPE.fq.gz output_reverse_unpairedxxxFFPE.fq.gz ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:8:TRUE LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:30

#Filter mapping to rRNA, tRNA & MTrna (Tophat 2 and Bowtie)
tophat -o "./tophatoutxxxFFPErrna19" -g 1 -p 16 -r -45 --no-mixed rRNA4 output_forward_pairedxxxFFPE.fq.gz output_reverse_pairedxxxFFPE.fq.gz

#Convert unmapped BAM file back to fastq file for mapping to reference genome and transcriptome (Tophat 2)
bam2fastx -A -o xxxFFPEfiltered19.fq.gz -P "./tophatoutxxxFFPErrna19/unmapped.bam"

#Map filtered fastq files to reference genome and transcriptome (Tophat 2)
tophat -o "./tophatoutxxxFFPEgenomemap19strandspecific" -g 1 -p 16 -r -45 --library-type fr-firststrand -G gencodev11.gtf hg19 xxxFFPEfiltered19.1.fq.gz xxxFFPEfiltered19.2.fq.gz

#Calculate insert sizes (Picard Tools)
java -jar picard.jar CollectInsertSizeMetrics I=./bamfiles/accepted_hitsFFR001.bam O=./insertsize/insertFFR001.txt H=./insertsize/insertFFR001.pdf M=0.5

#Count reads mapped to the genome (FeatureCounts/RSubRead)
```
library("Rsubread")

setwd("~/files")

#Import files and set up list of BAM files for processing:
csvfile <- "Sample Sheet"
(sampletable <- read.csv("ffpesamplesheet.csv",row.names=1))
dir <- "."
filenames <- file.path(dir, sampletable$bam_file)
file.exists(filenames)

#Generate counts matrix using Rsubread:
exonfeatures <- featureCounts(files=filenames, isPairedEnd=TRUE,
requireBothEndsMapped=TRUE, countChimericFragments=FALSE, minMQS=10,
strandSpecific=2, annot.ext="gencodev11.gtf", isGTFAnnotationFile=TRUE,
GTF.featureType="exon", GTF.attrType="gene_id", nthreads=16)

#View parameters of count assay:
sum(exonfeatures$counts)
write.csv(exonfeatures$counts, "hg19exoncountstrandFFPE.csv")
summary(exonfeatures)
exonstats <- exonfeatures$stat
write.csv(exonstats, file="exonstatsstrandFFPE.csv")
```

```
#Generate counts matrix using Rsubread:
genefeatures <- featureCounts(files=filenames, isPairedEnd=TRUE,
requireBothEndsMapped=TRUE, countChimericFragments=FALSE, minMQS=10,
strandSpecific=2, annot.ext="gencodev11.gtf", isGTFAnnotationFile=TRUE,
GTF.featureType="gene", GTF.attrType="gene_id", nthreads=16)

#View parameters of count assay:
sum(genefeatures$counts)
write.csv(genefeatures$counts, "hg19genecountstrandFFPE.csv")
summary(genefeatures)
genestats <- genefeatures$stat
write.csv(genestats, file="genestatsstrandFFPE.csv")
```

#Extract gene lengths for FPKM calculations

```
library(GenomicFeatures)
txdb <- makeTxDbFromGFF("gencodev11.gtf", format="gtf")
exons.list.per.gene <- exonsBy(txdb,by="gene")
exonic.gene.sizes <-
lapply(exons.list.per.gene,function(x){sum(width(reduce(x)))})
```

#Calculate TPM Matrix:

```
library(GenomicFeatures)
library("org.Hs.eg.db")
library("AnnotationDbi")

rawdata <- read.csv("~/Documents/Analysis/RNA-seq Analysis/2016 Re-
analysis/FeatureCounts Output/hg19exoncountstrandredo.csv")
X2 <- gsub("\\.[0-9]*$", '', rawdata$X)
rawdata$X <- NULL
row.names(rawdata) <- X2
rawdata$X <- NULL

coldata <- read.csv("~/Documents/Analysis/RNA-seq Analysis/Data
outputs/Sample Sheets/ffsamplesheet.csv")
row.names(coldata) <- coldata$rnaseq_id

b <- as.character(coldata$rnaseq_id)
colnames(rawdata) <- b

genelength <- read.csv("~/Documents/Analysis/RNA-seq Analysis/Data
outputs/genelength.csv")
ma_fpkm <- data.frame(rawdata)
ma_fpkm$gene <- row.names(ma_fpkm)
genelength$kblength <- genelength$V1/1000
ma_fpkm <- ma_fpkm[with(ma_fpkm, order(gene)), ]
ma_fpkm$gene <- NULL
ma_fpkm <- (ma_fpkm/genelength$kblength)
rawdata <- as.data.frame(ma_fpkm)

fpm <- function(x) {
  result <- c(x/((sum(x))/1000000))
  return(result)
}

rawdata$`sample name/column name` <- fpm(rawdata$`sample name/column
name`)
```

```
ma_tpm <- rawdata
tpmlog <- (ma_tpm+1)
tpmlog <- log2(tpmlog)

tpmlog$entrez <- mapIds(org.Hs.eg.db,
                  keys=row.names(tpmlog),
                  column="ENTREZID",
                  keytype="ENSEMBL",
                  multiVals="first")
tpmlog$symbol <- mapIds(org.Hs.eg.db,
                  keys=row.names(tpmlog),
                  column="SYMBOL",
                  keytype="ENSEMBL",
                  multiVals="first")


col_idx <- grep("entrez", names(tpmlog))
tpmlog <- tpmlog[, c(col_idx, (1:ncol(tpmlog))[-col_idx])]

col_idx <- grep("symbol", names(tpmlog))
tpmlog <- tpmlog[, c(col_idx, (1:ncol(tpmlog))[-col_idx])]
```

#EdgeR for Differential Expression between primary and recurrence

```
library("org.Hs.eg.db")
library("AnnotationDbi")
library("GenomicFeatures")
library("edgeR")

countdata <- read.csv("Raw counts matrix")
coldata <- read.csv("Sample sheet")


b <- as.character(coldata$rnaseq_id)
colnames(countdata) <- b

countdata <- cbind(X2, countdata)


rownames(countdata) <- countdata$X2
countdata$X2 <- NULL
countdata$entrez <- mapIds(org.Hs.eg.db,
                    keys=row.names(countdata),
                    column="ENTREZID",
                    keytype="ENSEMBL",
                    multiVals="first")
countdata$symbol <- mapIds(org.Hs.eg.db,
                    keys=row.names(countdata),
                    column="SYMBOL",
                    keytype="ENSEMBL",
                    multiVals="first")

col_idx <- grep("entrez", names(countdata))
countdata <- countdata[, c(col_idx, (1:ncol(countdata))[-col_idx])]

col_idx <- grep("symbol", names(countdata))
countdata <- countdata[, c(col_idx, (1:ncol(countdata))[-col_idx])]
```

```
countdata <- na.omit(countdata)


y <- DGEList(counts=countdata[,3:34], genes=countdata[,1:2])
y$samples

#Remove duplicated symbols
o <- order(rowSums(y$counts), decreasing=TRUE)
y <- y[o,]
d <- duplicated(y$genes$symbol)
y <- y[!d,]
nrow(y)

#Remove lowly expressed transcripts
keep <- rowSums(cpm(y)>1) >(50% of samples)

y <- y[keep, , keep.lib.sizes=FALSE]
nrow(y)

#Recompute library sizes
y$samples$lib.size <- colSums(y$counts)
y$samples
#Use entrez gene IDs as row names
rownames(y$counts) <- rownames(y$genes) <- y$genes$X

y <- calcNormFactors(y)
y$samples

patient <- factor(coldata$paired)
recurrence <- factor(coldata$primrecc)
location <- factor(coldata$location)
data.frame(Sample=colnames(y),patient,recurrence)
design1 <- model.matrix(~patient+recurrence)
rownames(design1) <- colnames(y)
design1

y <- estimateGLMCommonDisp(y, design=design1)
y <- estimateGLMTrendedDisp(y, design=design1)
y <- estimateGLMTagwiseDisp(y, design=design1)

plotBCV(y)
fit <- glmFit(y, design=design1)
lrt <- glmLRT(fit, coef=(position in dataframe for comparison))
```

#Code for ChAMP DNA Methylation Data

```
testDir <- "Directory containing DNA methylation .IDAT files"
myLoad <- champ.load(testDir, arraytype = "450k")
champ.QC()
Normalisation <- champ.norm(resultsDir = "results directory")
mdsPlot(Normalisation, numpositions=1000
```

Appendix 4: Basic RNA sequencing outcomes

Appendix 4, Table 1: Basic RNA sequencing outcomes for the FFPE tissue. Age refers to the age of the tissue block at the time of sequencing. Nanodrop (spectrophotometer) concentration is measured in ng/μl. 260/280 and 260/230 are the purity scores from the spectrophotometer. Genes indicates the number of genes detected at >1 count in the sample.

RNA-Seq ID	ID	Raw reads	% Trimmed	% rRNA	% Human Alignment	% Bacterial Alignment	Age	Nanodrop Conc	260/280	260/230	Tissue	Genes
11A	Epend003.P	58050650	2.9	2.1	70.6	10.4	15.4	104.5	1.98	1.05	Core	22917
11B	Epend003.R1	60613568	3.0	3.1	71.8	6.4	12.7	136.6	1.99	0.32	Core	21974
5B	Epend022.R1	45085860	3.3	7.5	74.0	5.8	18.4	116.5	1.99	1.06	Core	27111
5C	Epend022.R2	44823694	2.6	1.5	82.1	4.7	18.1	65.6	2.01	0.1	Core	25861
36B	Epend029.R1	50716789	2.8	14.9	70.7	1.0	11.3	91.1	1.95	0.61	Core	27251
14A	Epend030.P	55774570	2.0	3.6	72.5	3.5	13.8	42.1	2.02	0.34	Core	27093
14B	Epend030.R1	43329931	4.6	0.5	4.4	82.6	11.5	69.4	1.44	0.91	Core	5293
3A	Epend033.P	52190563	8.8	2.2	59.1	11.2	19.3	44.9	1.61	0.58	Core	20489
3B	Epend033.R1	54941100	2.8	2.6	77.5	2.8	19.0	215.7	1.98	1.03	Core	24864
10A	Epend034.P	2201032	8.5	8.5	1.3	77.8	17.2	36.4	1.73	0.73	Core	1651
10B	Epend034.R1	27984209	3.5	71.7	0.8	21.2	12.8	27.4	2.06	0.72	Core	6379
35A	Epend051.P	42779742	3.2	8.2	41.6	35.5	18.4	192.4	1.91	1.33	Core	22852
35B	Epend051.R1	30225114	6.1	3.2	2.2	76.3	16.1	100.5	1.71	0.8	Core	6372
20A	Epend085.P	47849106	2.3	2.3	85.0	1.4	9.0	40.1	1.87	0.73	Core	27958
20B	Epend085.R1	59121375	2.6	2.3	83.0	1.2	7.6	29.8	1.96	0.54	Core	29649
1A	Epend092.P	65231524	2.6	3.9	76.8	3.5	27.9	72.3	0.94	0.51	Core	26864
1C	Epend092.R2	49473715	2.3	4.1	78.0	4.8	24.8	110.8	1.89	0.43	Core	25988
6A	Epend095.P	53661997	2.7	2.5	61.4	21.1	24.2	117.4	1.87	0.47	Core	18292
6B	Epend095.R1	28126061	2.4	6.0	77.2	5.3	15.5	72.3	1.88	0.52	Core	25173
18B	Epend097.R1	38586050	6.0	2.6	39.9	41.4	9.4	45.3	1.86	0.14	Core	21933
32A	Epend098.P	55086293	2.8	1.9	78.3	2.7	20.4	131.3	1.99	0.29	Core	27018
32B	Epend098.R1	30381194	22.1	1.2	2.9	67.9	7.9	34	1.91	0.05	Core	7570
4B	Epend114.R1	41515099	1.9	2.7	78.1	5.9	18.3	310.4	1.94	1.81	Core	24943
2A	Epend118.P	42851384	3.2	2.0	37.2	49.4	22.1	75.9	1.64	0.56	Core	16424
2B	Epend118.R1	55213594	2.4	3.2	78.3	3.4	20.1	78.4	2.25	0.36	Core	24986
2C	Epend118.R2	42161486	6.1	6.7	38.3	22.0	15.5	54.4	0.73	0.89	Core	20888
2D	Epend118.R3	39230674	10.0	2.9	47.3	30.9	13.2	50.1	1.77	1.09	Core	22673
23A	Epend121.P	43235763	2.4	4.4	83.5	0.9	5.4	517.6	1.94	0.77	Core	27106
26A	Epend122.P	46208949	2.6	7.8	77.3	1.9	4.1	142.1	1.97	0.26	Core	26589

RNA-Seq ID	ID	Raw reads	% Trimmed	% rRNA	% Human Alignment	% Bacterial Alignment	Age	Nanodrop Conc	260/280	260/230	Tissue	Genes
26B	Epend122.R1	47356510	2.3	4.3	79.9	1.4	3.9	40.8	1.77	0.39	Core	26706
24A	Epend123.P	56059694	2.3	1.8	81.3	0.5	4.1	90.5	1.78	0.66	Core	28129
25A	Epend124.P	45256088	3.1	13.5	71.2	2.5	3.6	194.7	2.01	0.38	Core	24491
25B	Epend124.R1	4673675	5.4	9.1	54.3	25.5	2.3	26.3	2.03	0.11	Core	14905
39A	Epend129.P	51479763	2.5	4.9	81.0	0.6	4.0	175.5	2.01	0.36	Core	27052
54A	Epend183.P	50845798	5.4	4.1	30.1	49.8	20.8	55.1	1.67	0.17	Scroll	14145
54B	Epend183.R1	37928669	4.8	2.8	0.5	86.4	18.3	229.8	1.89	1.4	Scroll	4568
55C	Epend185.R2	61268582	6.5	5.1	49.0	26.1	14.9	54	1.8	0.65	Scroll	21417
55D	Epend185.R3	51359965	4.8	1.4	1.3	88.3	12.8	68.5	1.78	0.85	Scroll	8463
55E	Epend185.R4	42073846	2.6	7.5	50.7	31.1	11.7	251.4	1.93	1.74	Scroll	23063
53A	Epend186.P	49510679	22.4	0.9	11.1	60.0	19.2	34	1.72	0.21	Scroll	11525
53B	Epend186.R1	36602606	7.9	4.6	15.5	58.6	15.8	55.4	2.03	0.15	Scroll	14088
53C	Epend186.R2	20248984	11.6	8.1	11.6	63.4	12.1	80.2	1.87	0.32	Scroll	12987
53D	Epend186.R3	29629888	5.8	3.1	10.1	69.9	10.9	53.1	1.67	0.35	Scroll	13075
56B	Epend190.R1	43750296	10.7	2.9	0.4	80.1	16.6	216.6	1.94	0.49	Scroll	1847
16B	Epend193.R1	45878755	3.6	8.0	71.7	5.7	9.7	397.8	2.02	0.61	Scroll	27308
16C	Epend193.R2	37715959	2.5	2.9	10.5	78.9	9.4	23	1.96	0.07	Scroll	13986
57A	Epend195.P	40159692	3.7	8.9	45.5	29.3	13.4	212.2	1.93	0.93	Scroll	21581
57B	Epend195.R1	93162604	4.0	3.4	58.8	20.3	13.0	94.3	1.94	0.24	Scroll	23778
52B	Epend196.R1	49460284	8.4	9.3	35.7	33.1	9.8	247.2	1.99	0.46	Scroll	21857
52C	Epend196.R2	55546665	1.9	22.7	62.9	0.7	8.2	84.5	1.8	0.44	Scroll	28425
58A	Epend197.P	51880018	4.3	6.2	73.9	8.1	11.4	228.3	1.95	1.02	Scroll	26859
58B	Epend197.R1	36378455	8.7	10.8	33.3	36.3	10.0	256.5	1.95	1.48	Scroll	20469
62A	Epend203.P	68577710	3.0	3.6	81.2	0.8	11.9	371.1	2	1.14	Scroll	29798
62B	Epend203.R1	28787935	2.2	1.3	50.9	33.9	11.8	105.2	1.9	0.87	Scroll	20310
61A	Epend205.P	62385098	2.0	2.6	84.0	0.9	11.5	129.1	1.9	1.3	Scroll	28383
61B	Epend205.R1	56117115	4.4	10.7	38.0	35.7	9.7	168.2	1.99	1.43	Scroll	18963
61C	Epend205.R2	55199448	2.7	1.7	83.3	0.6	8.6	566.2	2.01	1.72	Scroll	28306
17A	Epend207.P	47155929	2.0	4.4	33.2	50.5	10.8	150.1	1.94	0.54	Scroll	15999
48A	Epend211.P	45077666	2.1	6.3	68.4	12.0	21.9	230.4	2	0.61	Core	21369
48B	Epend211.R1	39972715	2.0	2.7	77.1	7.7	11.7	46.9	1.77	0.29	Core	25664
47A	Epend213.P	50975838	10.0	6.9	12.3	58.7	16.8	59.5	1.81	0.11	Core	16839
47B	Epend213.R1	44876648	3.1	5.5	80.6	1.1	11.1	481.7	1.99	1.2	Core	28116
43A	Epend230.P	45545750	6.8	15.4	1.6	69.7	19.6	104.1	1.88	0.58	Core	7750
43B	Epend230.R1	39969723	5.3	2.7	65.3	14.4	18.1	142.8	1.94	0.31	Core	24460
43C	Epend230.R2	54464174	7.6	7.9	33.1	40.0	16.3	169	1.97	0.39	Core	15855

RNA-Seq ID	ID	Raw reads	% Trimmed	% rRNA	% Human Alignment	% Bacterial Alignment	Age	Nanodrop Conc	260/280	260/230	Tissue	Genes
59A	Epend234.P	57635915	1.9	6.8	78.3	1.5	19.6	110.1	2.01	0.3	Scroll	28617
59B	Epend234.R1	40816548	2.7	4.6	74.8	6.7	10.5	148	1.92	0.59	Scroll	24586
59C	Epend234.R2	46342992	3.8	4.9	73.6	8.4	10.2	55.5	1.92	0.52	Scroll	27145
51A	Epend256.P	47198389	2.0	4.6	78.9	2.2	13.3	192.2	1.96	1.25	Core	27223
51C	Epend256.R2	57076886	2.0	5.5	79.7	1.1	11.6	74.2	1.98	1.2	Core	26249
29A	Epend258.P	42812480	2.5	5.9	77.4	1.0	12.7	488	1.95	1.65	Core	25187
29B	Epend258.R1	58373414	2.9	2.3	77.3	3.2	12.3	216.3	1.99	1.65	Core	24770
30A	Epend280.P	50325106	2.2	4.4	83.2	1.6	11.0	158.6	1.74	0.36	Core	27546
30B	Epend280.R1	11354416	4.1	63.0	4.2	25.8	9.2	21.3	2.12	0.09	Core	9584
46A	Epend287.P	47857070	2.6	6.3	78.7	1.8	3.5	251.2	1.94	1.05	Core	28219
46B	Epend287.R1	45116953	2.2	19.5	54.8	11.9	0.5	58.6	1.93	0.9	Core	23259
65A	Epend288.P	45599658	3.1	6.0	77.6	3.3	7.0	201.4	2.03	0.03	Scroll	27584
65D	Epend288.R3	81189859	3.4	2.4	75.8	6.0	5.0	80.4	1.91	0.26	Scroll	25738
65E	Epend288.R4	47478221	1.9	1.3	82.6	1.9	4.6	156.7	1.93	0.88	Scroll	27132
28A	Epend293.P	35646871	3.7	3.8	81.9	1.9	15.3	287.9	1.94	1.63	Core	26985
28B	Epend293.R1	48154168	2.3	1.4	82.0	2.9	11.3	78.9	1.79	1.11	Core	26018
28E	Epend293.R4	55614306	2.5	4.4	81.6	0.8		234.2	2.03	0.71	Core	27424
28F	Epend293.R5	48844876	1.9	6.7	82.3	1.1		83.4	1.97	0.28	Core	27786
28G	Epend293.R6	25632770	3.2	8.8	66.9	12.7		215.9	1.99	1.17	Core	21274
28H	Epend293.R7	40792024	3.6	2.7	82.0	2.1		70.5	1.96	0.33	Core	27079
27A	Epend305.P	34495433	31.4	1.9	34.3	45.5	13.6	53.6	1.97	0.11	Core	17708
27B	Epend305.R1	15146602	2.4	2.9	83.0	1.2	11.0	140.7	1.94	0.78	Core	24434
44A	Epend306.P	68494371	2.4	2.3	77.1	6.3	13.3	182.2	1.92	0.47	Core	26262
44B	Epend306.R1	58156554	3.0	3.5	79.9	3.1	11.6	138.4	1.93	0.87	Core	28024
45A	Epend316.P	48730352	2.2	2.4	85.5	2.1	11.6	136.2	1.93	0.83	Core	26808
45B	Epend316.R1	39011090	2.1	1.7	85.9	1.9	11.0	107.9	1.95	0.4	Core	27030
38A	Epend329.P	78967281	2.6	8.9	80.2	0.4	2.7	319.8	1.54	0.18	Core	31269
63A	Epend330.P	10613332	5.3	1.6	8.7	77.8	6.9	74	1.72	0.29	Scroll	10791
63B	Epend330.R1	31841825	14.1	3.3	8.7	64.3	5.5	37.7	1.84	0.43	Scroll	15305
63C	Epend330.R2	10126681	4.6	11.4	0.5	79.7	5.3	38.7	1.76	0.4	Scroll	2492
63E	Epend330.R4	49844622	2.7	6.6	60.1	22.7	2.3	95.3	1.94	0.24	Scroll	24067
64A	Epend331.P	53254189	2.4	7.0	76.2	3.2	7.6	39.4	2	1.25	Scroll	26775
64B	Epend331.R1	73119157	2.0	13.1	73.3	0.9	4.6	134.6	1.91	0.26	Scroll	28228
67A	Epend333.P	39293572	7.0	11.9	38.3	17.4	14.8	877.5	2.04	1.84	Scroll	20933
67B	Epend333.R1	53541840	3.7	4.3	72.2	4.2	13.3	24.2	2.01	0.28	Scroll	23866
83B	Epend349.R1	37298100	4.5	1.0	0.6	86.0		35.9	1.67	0.14	Scroll	1835

RNA-Seq ID	ID	Raw reads	% Trimmed	% rRNA	% Human Alignment	% Bacterial Alignment	Age	Nanodrop Conc	260/280	260/230	Tissue	Genes
83C	Epend349.R2	18498813	6.2	0.8	1.2	85.6		59.8	1.81	0.06	Scroll	3336
84A	Epend371.P	42095259	2.9	2.0	5.0	85.0	5.8	89.5	1.83	0.86	Scroll	10118
84B	Epend371.R1	93187157	1.8	6.2	83.2	0.6		151.9	2	1.46	Scroll	29944
85A	Epend372.P	39646334	2.3	0.9	1.0	91.4		26.2	1.86	0.33	Scroll	5061
85B	Epend372.R1	48135029	3.2	3.7	21.5	58.8		42.4	1.75	0.92	Scroll	15709

Appendix 4, Table 2: Basic RNA sequencing outcomes for the FF tissue.

RNA-Seq ID	Raw reads	% Trimmed	%rRNA	% Human Alignment
1A	38011142	7.2	3.9	78.0
1B	50740785	5.1	3.6	83.2
1C	52748458	2.1	2.3	89.5
4A	51488585	5.5	47.9	40.5
4B	60754916	4.7	1.4	85.4
12A	49232540	3.2	4.7	86.4
12C	68920610	4.5	3.2	83.8
15A	53359467	3.3	4.1	86.7
15B	49033158	5.7	1.6	84.9
16A	37422006	3.5	4.7	85.6
16B	40520519	3.3	1.1	88.9
17A	54173568	4.1	1.5	86.8
18A	54211377	5.1	3.2	83.5
18B	50265356	7.6	0.9	79.9
18C	43108985	6.6	1.8	81.6
20A	52575726	3.3	0.8	89.6
20C	72302943	4.8	1.1	88.1
21A	60844655	6.6	0.7	82.9
21B	64208203	4.7	28.4	59.9
23A	58463112	5.4	0.9	85.2
23B	50293092	5.2	4.2	80.8
24A	57208160	3.7	6.4	84.5
24B	57115152	5.0	2.5	85.8
25A	61900222	4.7	7.5	80.9
25B	62055449	3.3	0.7	89.9
25C	66678612	3.7	12.0	77.3
26A	52015796	5.7	0.7	86.1
26B	61519898	2.9	1.8	88.2
34A	42802524	2.7	0.8	89.4
34B	49183057	7.1	2.9	82.1
39A	58093772	5.0	6.8	81.0
39B	78527974	3.6	61.6	29.7
40A	47664372	4.0	7.1	80.3
40B	47682890	6.9	3.4	80.0
41A	57855608	4.6	4.5	83.2
41B	48069317	6.1	2.8	81.0

RNA-Seq ID	Raw reads	% Trimmed	%rRNA	% Human Alignment
69A	58502970	2.3	1.7	90.0
69B	55011425	4.5	8.3	81.1
70A	49651561	6.8	5.9	79.5
70B	22035684	5.6	1.5	83.9
71A	56465234	3.0	1.6	88.5
71B	63250400	3.8	4.2	84.8
72A	55770800	5.2	8.4	79.1
72B	66149159	4.4	5.0	83.0
73A	46803002	6.9	68.1	16.3
73B	63635666	3.7	5.4	84.0
74A	58567648	3.2	6.2	84.0
74B	45609693	4.7	2.8	85.4
74C	52999899	4.4	1.0	87.2
75A	52329878	4.8	3.9	83.9
75B	39869320	7.7	3.1	78.6
76A	51814823	5.5	2.9	83.9
76B	57030618	2.1	2.7	88.8
76C	71129009	3.4	1.9	88.4
76D	54412895	4.8	31.2	56.1
76E	58318558	3.9	4.4	85.1
77A	42510039	5.2	3.4	83.8
77B	54017210	5.2	5.3	84.2
78A	51256068	4.4	1.2	86.8
78B	59336838	5.4	6.7	81.8
78C	49572507	8.0	2.2	78.4
80A	46134028	5.1	1.7	85.4
80B	62867765	4.1	16.9	72.1
81A	59738410	4.9	3.2	84.3
81B	49351093	3.9	7.9	81.7
82A	49812270	6.6	6.0	77.8
82B	49480393	5.8	2.0	85.5

Appendix 5: Lists of enriched ontologies derived from primary and recurrent pair comparisons

Gene Ontology Results (GOrilla) for primary and recurrent comparisons. Contains lists of ontologies up and downregulated at first recurrence in both FFPE and FF datasets. The only exception is the ST and EPN_RELA datasets where only FFPE comparisons were made.

ALL TUMOURS: UPREGULATED AT FIRST RECURRENCE (All Terms)

GO term	Description	FFPE FDR	FF FDR
GO:0043062	extracellular structure organization	2.19E-04	1.81E-15
GO:0030198	extracellular matrix organization	3.29E-04	1.77E-15
GO:0032101	regulation of response to external stimulus	6.44E-04	6.00E-11
GO:0006955	immune response	2.08E-03	2.27E-11
GO:0006954	inflammatory response	2.69E-03	4.10E-11
GO:0032501	multicellular organismal process	2.58E-03	1.24E-10
GO:0044243	multicellular organismal catabolic process	9.49E-05	6.51E-08
GO:0030574	collagen catabolic process	2.48E-04	3.50E-08
GO:0044236	multicellular organism metabolic process	4.06E-03	4.94E-07
GO:0032963	collagen metabolic process	3.82E-03	1.85E-06
GO:0044259	multicellular organismal macromolecule metabolic process	2.86E-03	5.57E-06
GO:0006811	ion transport	6.39E-03	2.73E-06
GO:0033993	response to lipid	2.82E-02	4.10E-06
GO:0032944	regulation of mononuclear cell proliferation	3.44E-02	5.52E-06
GO:0050670	regulation of lymphocyte proliferation	2.96E-02	1.17E-05
GO:0070663	regulation of leukocyte proliferation	2.97E-02	1.32E-05
GO:0022617	extracellular matrix disassembly	3.16E-02	1.30E-04
GO:0015671	oxygen transport	2.92E-03	2.48E-03
GO:0050710	negative regulation of cytokine secretion	4.03E-02	6.85E-03
GO:0015669	gas transport	1.82E-02	1.53E-02
GO:0003013	circulatory system process	4.08E-02	7.73E-03
GO:0050954	sensory perception of mechanical stimulus	4.67E-02	2.98E-02

ALL TUMOURS: DOWNREGULATED AT FIRST RECURRENCE (Top 100 Terms Only)

GO term	Description	FFPE FDR	FF FDR
GO:0048731	system development	3.06E-10	2.34E-09
GO:0023052	signaling	7.99E-08	2.34E-11
GO:0007267	cell-cell signaling	6.61E-08	7.50E-10
GO:0032502	developmental process	3.87E-05	1.52E-12
GO:0003008	system process	2.10E-05	5.10E-12
GO:0051960	regulation of nervous system development	9.00E-07	1.38E-10
GO:0048856	anatomical structure development	8.19E-06	2.11E-11
GO:0050877	nervous system process	6.47E-07	3.31E-09
GO:0060284	regulation of cell development	9.03E-07	2.99E-09
GO:0007610	behavior	6.69E-08	1.24E-07
GO:0050767	regulation of neurogenesis	9.73E-06	8.77E-10
GO:0050804	modulation of chemical synaptic transmission	1.67E-07	1.50E-07
GO:0007186	G-protein coupled receptor signaling pathway	8.89E-06	5.18E-09
GO:0051239	regulation of multicellular organismal process	1.06E-03	6.16E-11
GO:0007399	nervous system development	1.74E-07	5.05E-07
GO:0045664	regulation of neuron differentiation	5.30E-05	2.32E-09
GO:0099537	trans-synaptic signaling	1.02E-04	1.99E-09
GO:0099536	synaptic signaling	1.05E-04	2.11E-09
GO:0032501	multicellular organismal process	2.49E-04	9.06E-10
GO:0030001	metal ion transport	2.50E-04	2.59E-09
GO:0006811	ion transport	1.46E-02	1.30E-10
GO:0043269	regulation of ion transport	4.11E-03	8.25E-10
GO:0048869	cellular developmental process	1.16E-02	3.32E-10
GO:0042391	regulation of membrane potential	2.32E-03	1.70E-09
GO:2000026	regulation of multicellular organismal development	1.07E-03	4.71E-09
GO:0044057	regulation of system process	4.39E-02	1.18E-10
GO:0007154	cell communication	8.92E-06	1.08E-06
GO:0007268	chemical synaptic transmission	5.42E-04	2.57E-08
GO:0098916	anterograde trans-synaptic signaling	5.57E-04	2.65E-08
GO:0007611	learning or memory	2.33E-06	1.39E-05
GO:0034220	ion transmembrane transport	3.61E-02	9.30E-10
GO:1903522	regulation of blood circulation	1.28E-02	3.39E-09
GO:0006812	cation transport	2.24E-02	5.25E-09
GO:0050890	cognition	2.12E-06	1.10E-04
GO:0045665	negative regulation of neuron differentiation	3.89E-04	1.06E-06
GO:0050793	regulation of developmental process	8.23E-03	6.96E-08
GO:0007155	cell adhesion	2.29E-06	2.99E-04
GO:0009653	anatomical structure morphogenesis	2.14E-02	3.59E-08
GO:0008016	regulation of heart contraction	3.52E-02	2.43E-08
GO:0022610	biological adhesion	2.78E-06	3.65E-04
GO:0050768	negative regulation of neurogenesis	3.69E-04	2.82E-06
GO:0051094	positive regulation of developmental process	5.36E-05	2.56E-05
GO:0051962	positive regulation of nervous system development	2.16E-04	7.10E-06
GO:0060078	regulation of postsynaptic membrane potential	1.03E-04	1.86E-05
GO:0007626	locomotory behavior	2.88E-05	6.89E-05
GO:0034762	regulation of transmembrane transport	1.92E-02	1.28E-07
GO:0007612	Learning	3.00E-05	1.10E-04
GO:0098660	inorganic ion transmembrane transport	1.52E-02	2.77E-07
GO:0034765	regulation of ion transmembrane transport	1.57E-02	3.93E-07
GO:0051961	negative regulation of nervous system development	7.51E-04	8.97E-06
GO:0010469	regulation of receptor activity	2.37E-03	7.50E-06
GO:0010721	negative regulation of cell development	1.20E-03	1.64E-05
GO:0051241	negative regulation of multicellular organismal process	4.93E-02	4.13E-07
GO:0098662	inorganic cation transmembrane transport	1.90E-02	1.10E-06
GO:0010975	regulation of neuron projection development	2.02E-03	1.09E-05
GO:0045595	regulation of cell differentiation	2.39E-03	1.10E-05
GO:0006814	sodium ion transport	2.70E-03	3.51E-05
GO:0030182	neuron differentiation	3.74E-03	3.09E-05
GO:0050770	regulation of axonogenesis	2.63E-02	5.72E-06
GO:0050808	synapse organization	5.60E-05	3.33E-03
GO:0050806	positive regulation of synaptic transmission	3.45E-05	5.46E-03
GO:0072503	cellular divalent inorganic cation homeostasis	1.73E-03	1.09E-04

GO term	Description	FFPE FDR	FF FDR
GO:0010769	regulation of cell morphogenesis involved in differentiation	2.08E-02	9.96E-06
GO:0007215	glutamate receptor signaling pathway	1.87E-02	1.35E-05
GO:0060079	excitatory postsynaptic potential	8.81E-03	6.29E-05
GO:0072507	divalent inorganic cation homeostasis	2.05E-03	2.71E-04
GO:0045596	negative regulation of cell differentiation	3.11E-02	1.89E-05
GO:0051240	positive regulation of multicellular organismal process	8.23E-03	9.24E-05
GO:0045165	cell fate commitment	2.35E-02	3.78E-05
GO:0035725	sodium ion transmembrane transport	3.54E-03	3.02E-04
GO:2001257	regulation of cation channel activity	2.35E-03	5.38E-04
GO:0051480	regulation of cytosolic calcium ion concentration	2.01E-02	6.66E-05
GO:0006874	cellular calcium ion homeostasis	2.88E-03	4.91E-04
GO:0055074	calcium ion homeostasis	4.10E-03	3.75E-04
GO:0007187	G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	1.22E-04	1.39E-02
GO:0023051	regulation of signaling	4.06E-03	4.95E-04
GO:0022898	regulation of transmembrane transporter activity	2.06E-02	1.10E-04
GO:0035235	ionotropic glutamate receptor signaling pathway	4.54E-03	5.42E-04
GO:0032412	regulation of ion transmembrane transporter activity	1.91E-02	1.73E-04
GO:0048812	neuron projection morphogenesis	8.24E-03	6.78E-04
GO:0048167	regulation of synaptic plasticity	2.79E-02	2.13E-04
GO:0051966	regulation of synaptic transmission, glutamatergic	3.50E-03	1.93E-03
GO:0120039	plasma membrane bounded cell projection morphogenesis	1.39E-02	4.91E-04
GO:0007188	adenylate cyclase-modulating G-protein coupled receptor signaling pathway	2.56E-04	3.14E-02
GO:0050771	negative regulation of axonogenesis	3.52E-02	2.37E-04
GO:0032409	regulation of transporter activity	3.12E-02	2.78E-04
GO:1900449	regulation of glutamate receptor signaling pathway	1.55E-02	7.52E-04
GO:0048858	cell projection morphogenesis	1.74E-02	6.73E-04
GO:0051965	positive regulation of synapse assembly	2.09E-03	8.94E-03
GO:2000463	positive regulation of excitatory postsynaptic potential	2.85E-03	7.82E-03
GO:0055065	metal ion homeostasis	4.34E-03	5.29E-03
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	8.43E-04	2.77E-02
GO:0007204	positive regulation of cytosolic calcium ion concentration	3.98E-02	6.23E-04
GO:0050807	regulation of synapse organization	2.42E-02	1.05E-03
GO:0120035	regulation of plasma membrane bounded cell projection organization	1.35E-02	1.93E-03
GO:0010646	regulation of cell communication	1.01E-02	2.59E-03
GO:0010720	positive regulation of cell development	3.69E-02	8.62E-04
GO:0031344	regulation of cell projection organization	1.85E-02	1.72E-03
GO:0031345	negative regulation of cell projection organization	1.84E-02	1.88E-03
GO:0006875	cellular metal ion homeostasis	1.32E-02	2.64E-03

ST TUMOURS: UPREGULATED AT FIRST RECURRENCE (Top 100 Terms Only)

GO Term	Description	FFPE FDR
GO:0006955	immune response	1.62E-26
GO:0002376	immune system process	3.03E-21
GO:0006952	defense response	3.70E-21
GO:0002682	regulation of immune system process	3.20E-15
GO:0050776	regulation of immune response	1.10E-13
GO:0002684	positive regulation of immune system process	3.46E-13
GO:0070665	positive regulation of leukocyte proliferation	6.80E-13
GO:0042102	positive regulation of T cell proliferation	3.37E-12
GO:0001817	regulation of cytokine production	3.61E-12
GO:0043207	response to external biotic stimulus	3.90E-12
GO:0050671	positive regulation of lymphocyte proliferation	4.89E-12
GO:0006954	inflammatory response	6.17E-12
GO:0032946	positive regulation of mononuclear cell proliferation	6.37E-12
GO:0098542	defense response to other organism	6.39E-12
GO:0070663	regulation of leukocyte proliferation	7.39E-12
GO:0009607	response to biotic stimulus	1.60E-11
GO:0050863	regulation of T cell activation	1.83E-11
GO:0007165	signal transduction	2.00E-11
GO:0050867	positive regulation of cell activation	2.38E-11
GO:0051251	positive regulation of lymphocyte activation	2.43E-11
GO:0002694	regulation of leukocyte activation	2.57E-11
GO:0050670	regulation of lymphocyte proliferation	2.60E-11
GO:0007166	cell surface receptor signaling pathway	3.09E-11
GO:0050865	regulation of cell activation	3.11E-11
GO:0045785	positive regulation of cell adhesion	3.13E-11
GO:0002696	positive regulation of leukocyte activation	3.17E-11
GO:0032944	regulation of mononuclear cell proliferation	3.18E-11
GO:0050870	positive regulation of T cell activation	3.18E-11
GO:1903039	positive regulation of leukocyte cell-cell adhesion	3.19E-11
GO:0001775	cell activation	3.19E-11
GO:0045321	leukocyte activation	3.19E-11
GO:1903037	regulation of leukocyte cell-cell adhesion	4.85E-11
GO:0051707	response to other organism	4.86E-11
GO:0050778	positive regulation of immune response	4.97E-11
GO:0031347	regulation of defense response	5.51E-11
GO:0002253	activation of immune response	6.27E-11
GO:0022407	regulation of cell-cell adhesion	6.70E-11
GO:0050707	regulation of cytokine secretion	7.84E-11
GO:0001819	positive regulation of cytokine production	8.05E-11
GO:0051249	regulation of lymphocyte activation	9.01E-11
GO:0048584	positive regulation of response to stimulus	1.90E-10
GO:0022409	positive regulation of cell-cell adhesion	3.12E-10
GO:0032101	regulation of response to external stimulus	3.12E-10
GO:0002757	immune response-activating signal transduction	3.17E-10
GO:0050896	response to stimulus	3.54E-10
GO:0045087	innate immune response	3.64E-10
GO:0051239	regulation of multicellular organismal process	3.84E-10
GO:0002250	adaptive immune response	4.27E-10
GO:0002764	immune response-regulating signaling pathway	4.31E-10
GO:0050900	leukocyte migration	5.65E-10
GO:0030155	regulation of cell adhesion	1.51E-09
GO:0048583	regulation of response to stimulus	1.71E-09
GO:0009605	response to external stimulus	1.82E-09
GO:0002252	immune effector process	3.38E-09
GO:0050727	regulation of inflammatory response	3.57E-09
GO:0002697	regulation of immune effector process	5.72E-09
GO:0009617	response to bacterium	6.23E-09
GO:0042742	defense response to bacterium	1.20E-08
GO:0042129	regulation of T cell proliferation	2.65E-08
GO:0032501	multicellular organismal process	4.51E-08
GO:0046641	positive regulation of alpha-beta T cell proliferation	1.63E-07
GO:0050715	positive regulation of cytokine secretion	2.44E-07

GO Term	Description	FFPE FDR
GO:0002237	response to molecule of bacterial origin	3.18E-07
GO:0046634	regulation of alpha-beta T cell activation	7.55E-07
GO:0048518	positive regulation of biological process	9.54E-07
GO:0002791	regulation of peptide secretion	1.10E-06
GO:0019221	cytokine-mediated signaling pathway	1.21E-06
GO:0001818	negative regulation of cytokine production	1.27E-06
GO:0030334	regulation of cell migration	1.41E-06
GO:0051240	positive regulation of multicellular organismal process	1.57E-06
GO:0007155	cell adhesion	2.37E-06
GO:0050777	negative regulation of immune response	2.68E-06
GO:0022610	biological adhesion	2.69E-06
GO:0002768	immune response-regulating cell surface receptor signaling pathway	2.76E-06
GO:0050708	regulation of protein secretion	2.79E-06
GO:0046635	positive regulation of alpha-beta T cell activation	3.19E-06
GO:0046649	lymphocyte activation	4.40E-06
GO:0045088	regulation of innate immune response	5.00E-06
GO:0002793	positive regulation of peptide secretion	5.60E-06
GO:0032496	response to lipopolysaccharide	6.03E-06
GO:2000145	regulation of cell motility	1.08E-05
GO:0002429	immune response-activating cell surface receptor signaling pathway	1.14E-05
GO:0002274	myeloid leukocyte activation	1.20E-05
GO:0042108	positive regulation of cytokine biosynthetic process	1.50E-05
GO:0032675	regulation of interleukin-6 production	1.91E-05
GO:0042127	regulation of cell proliferation	2.18E-05
GO:0050710	negative regulation of cytokine secretion	2.24E-05
GO:0002683	negative regulation of immune system process	2.25E-05
GO:0016477	cell migration	2.38E-05
GO:0032940	secretion by cell	2.66E-05
GO:0050714	positive regulation of protein secretion	2.69E-05
GO:0033993	response to lipid	2.78E-05
GO:0070372	regulation of ERK1 and ERK2 cascade	3.02E-05
GO:0046903	secretion	3.10E-05
GO:0002703	regulation of leukocyte mediated immunity	3.23E-05
GO:0046640	regulation of alpha-beta T cell proliferation	3.25E-05
GO:0002263	cell activation involved in immune response	3.33E-05
GO:0042035	regulation of cytokine biosynthetic process	3.40E-05
GO:0051270	regulation of cellular component movement	3.46E-05
GO:0006887	exocytosis	3.47E-05

ST TUMOURS: DOWNREGULATED AT FIRST RECURRENCE (All Terms)

GO Term	Description	FFPE FDR
GO:0048856	anatomical structure development	6.82E-10
GO:0032502	developmental process	1.66E-06
GO:0007610	behavior	3.70E-04
GO:0051960	regulation of nervous system development	3.46E-03
GO:0045664	regulation of neuron differentiation	3.81E-03
GO:0007399	nervous system development	3.87E-03
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	3.92E-03
GO:0061621	canonical glycolysis	4.44E-03
GO:0061718	glucose catabolic process to pyruvate	4.93E-03
GO:0007420	brain development	5.07E-03
GO:0048731	system development	5.23E-03
GO:0061615	glycolytic process through fructose-6-phosphate	5.36E-03
GO:0006735	NADH regeneration	5.54E-03
GO:0061620	glycolytic process through glucose-6-phosphate	5.77E-03
GO:0098609	cell-cell adhesion	6.47E-03
GO:0023052	signaling	6.56E-03
GO:0050767	regulation of neurogenesis	6.59E-03
GO:0009887	animal organ morphogenesis	6.83E-03
GO:0030182	neuron differentiation	1.19E-02
GO:0048513	animal organ development	1.22E-02
GO:0006007	glucose catabolic process	1.45E-02
GO:0007154	cell communication	1.69E-02
GO:0006096	glycolytic process	1.79E-02
GO:0006757	ATP generation from ADP	2.43E-02
GO:0007267	cell-cell signaling	2.59E-02
GO:0032501	multicellular organismal process	3.03E-02
GO:0009653	anatomical structure morphogenesis	3.73E-02
GO:0051966	regulation of synaptic transmission, glutamatergic	4.08E-02
GO:0099643	signal release from synapse	4.38E-02
GO:0019320	hexose catabolic process	4.41E-02

EPN_RELA TUMOURS: UPREGULATED AT FIRST RECURRENCE (Top 100 Terms Only)

GO Term	Description	FFPE FDR
GO:0051239	regulation of multicellular organismal process	1.34E-07
GO:0022610	biological adhesion	7.25E-07
GO:0007155	cell adhesion	8.08E-07
GO:0032501	multicellular organismal process	2.01E-06
GO:0032101	regulation of response to external stimulus	5.42E-06
GO:0043547	positive regulation of GTPase activity	6.13E-06
GO:0006952	defense response	6.53E-06
GO:0006955	immune response	9.55E-06
GO:0007165	signal transduction	3.30E-05
GO:0043087	regulation of GTPase activity	5.30E-05
GO:0001817	regulation of cytokine production	1.13E-04
GO:0051345	positive regulation of hydrolase activity	1.34E-04
GO:0030198	extracellular matrix organization	3.83E-04
GO:0043062	extracellular structure organization	3.84E-04
GO:0002376	immune system process	4.03E-04
GO:0050727	regulation of inflammatory response	4.20E-04
GO:0002730	regulation of dendritic cell cytokine production	4.65E-04
GO:0045765	regulation of angiogenesis	5.21E-04
GO:1901342	regulation of vasculature development	5.83E-04
GO:0007166	cell surface receptor signaling pathway	5.85E-04
GO:0009605	response to external stimulus	7.83E-04
GO:0022603	regulation of anatomical structure morphogenesis	8.26E-04
GO:0050793	regulation of developmental process	8.83E-04
GO:0051336	regulation of hydrolase activity	9.36E-04
GO:0050878	regulation of body fluid levels	9.42E-04
GO:1904018	positive regulation of vasculature development	1.15E-03
GO:0030155	regulation of cell adhesion	1.31E-03
GO:0008360	regulation of cell shape	1.38E-03
GO:0070372	regulation of ERK1 and ERK2 cascade	1.42E-03
GO:0006954	inflammatory response	1.77E-03
GO:0030334	regulation of cell migration	1.79E-03
GO:0051240	positive regulation of multicellular organismal process	1.79E-03
GO:0045766	positive regulation of angiogenesis	1.81E-03
GO:0032651	regulation of interleukin-1 beta production	1.84E-03
GO:0009653	anatomical structure morphogenesis	1.84E-03
GO:0001775	cell activation	1.84E-03
GO:0050707	regulation of cytokine secretion	1.85E-03
GO:0070374	positive regulation of ERK1 and ERK2 cascade	1.87E-03
GO:0002697	regulation of immune effector process	1.92E-03
GO:0098542	defense response to other organism	2.03E-03
GO:0050715	positive regulation of cytokine secretion	2.06E-03
GO:0050706	regulation of interleukin-1 beta secretion	2.09E-03
GO:0050718	positive regulation of interleukin-1 beta secretion	2.32E-03
GO:0050716	positive regulation of interleukin-1 secretion	2.37E-03
GO:0002250	adaptive immune response	2.38E-03
GO:0002252	immune effector process	2.38E-03
GO:0033993	response to lipid	2.39E-03
GO:2000026	regulation of multicellular organismal development	2.40E-03
GO:0048583	regulation of response to stimulus	2.43E-03
GO:0006869	lipid transport	2.59E-03
GO:0072376	protein activation cascade	2.62E-03
GO:0040011	locomotion	2.81E-03
GO:0050704	regulation of interleukin-1 secretion	4.00E-03
GO:0031347	regulation of defense response	4.00E-03
GO:1901700	response to oxygen-containing compound	4.73E-03
GO:0044093	positive regulation of molecular function	4.82E-03
GO:0038001	paracrine signaling	4.87E-03
GO:0045785	positive regulation of cell adhesion	4.90E-03
GO:0032652	regulation of interleukin-1 production	4.91E-03
GO:0032675	regulation of interleukin-6 production	4.94E-03

GO Term	Description	FFPE FDR
GO:0032715	negative regulation of interleukin-6 production	4.99E-03
GO:0045087	innate immune response	5.00E-03
GO:0002793	positive regulation of peptide secretion	5.00E-03
GO:0048646	anatomical structure formation involved in morphogenesis	5.07E-03
GO:0001525	angiogenesis	5.07E-03
GO:0043085	positive regulation of catalytic activity	5.24E-03
GO:0050790	regulation of catalytic activity	5.29E-03
GO:2000145	regulation of cell motility	5.32E-03
GO:0065008	regulation of biological quality	5.46E-03
GO:0003073	regulation of systemic arterial blood pressure	5.50E-03
GO:0032731	positive regulation of interleukin-1 beta production	5.54E-03
GO:0065009	regulation of molecular function	5.70E-03
GO:0043207	response to external biotic stimulus	5.72E-03
GO:0014070	response to organic cyclic compound	5.82E-03
GO:0045321	leukocyte activation	5.97E-03
GO:0051056	regulation of small GTPase mediated signal transduction	6.47E-03
GO:0016477	cell migration	6.82E-03
GO:0015711	organic anion transport	7.06E-03
GO:0051270	regulation of cellular component movement	7.43E-03
GO:0023052	signaling	8.19E-03
GO:0009607	response to biotic stimulus	8.92E-03
GO:0001818	negative regulation of cytokine production	8.92E-03
GO:0014910	regulation of smooth muscle cell migration	8.95E-03
GO:0050714	positive regulation of protein secretion	9.21E-03
GO:0006820	anion transport	9.26E-03
GO:0032732	positive regulation of interleukin-1 production	9.35E-03
GO:0048870	cell motility	1.09E-02
GO:0002366	leukocyte activation involved in immune response	1.10E-02
GO:0032502	developmental process	1.11E-02
GO:0035023	regulation of Rho protein signal transduction	1.17E-02
GO:0032655	regulation of interleukin-12 production	1.17E-02
GO:0098609	cell-cell adhesion	1.17E-02
GO:0003008	system process	1.20E-02
GO:0031348	negative regulation of defense response	1.24E-02
GO:0006811	ion transport	1.29E-02
GO:0051707	response to other organism	1.29E-02
GO:0007015	actin filament organization	1.30E-02
GO:0046903	secretion	1.30E-02
GO:0002263	cell activation involved in immune response	1.30E-02
GO:0040012	regulation of locomotion	1.31E-02

PF LOCATION TUMOURS: OVERLAPPING FFPE DOWNREGULATED AND FF UPREGULATED TERMS (All Terms)

GO Term	Description	FFPE (DOWN) FDR	FF (UP) FDR
GO:0002376	immune system process	1.02E-05	2.88E-28
GO:0006952	defense response	1.49E-09	4.65E-23
GO:0006955	immune response	1.10E-08	3.30E-23
GO:0032101	regulation of response to external stimulus	5.58E-08	3.76E-14
GO:0019221	cytokine-mediated signaling pathway	2.69E-10	1.66E-11
GO:0002682	regulation of immune system process	1.34E-03	2.00E-17
GO:0002684	positive regulation of immune system process	2.99E-03	5.03E-17
GO:0001775	cell activation	1.88E-02	2.24E-17
GO:0007166	cell surface receptor signaling pathway	2.79E-08	8.70E-11
GO:0002252	immune effector process	1.39E-02	2.09E-16
GO:0048584	positive regulation of response to stimulus	2.49E-07	1.85E-11
GO:0006954	inflammatory response	3.37E-05	9.14E-13
GO:0007165	signal transduction	1.18E-04	5.12E-12
GO:0051239	regulation of multicellular organismal process	8.63E-06	1.12E-10
GO:0045321	leukocyte activation	3.82E-02	3.82E-14
GO:0050776	regulation of immune response	4.13E-03	4.34E-13
GO:0048583	regulation of response to stimulus	1.05E-05	4.40E-10
GO:0001817	regulation of cytokine production	3.87E-02	1.73E-13
GO:0045087	innate immune response	2.20E-04	3.85E-11
GO:0040011	locomotion	1.21E-04	2.18E-10
GO:0060337	type I interferon signaling pathway	9.85E-03	7.48E-12
GO:0030155	regulation of cell adhesion	1.36E-03	1.10E-10
GO:0032103	positive regulation of response to external stimulus	4.44E-05	4.02E-09
GO:0050900	leukocyte migration	3.15E-02	1.03E-11
GO:0050778	positive regulation of immune response	3.52E-02	2.47E-11
GO:0045785	positive regulation of cell adhesion	1.87E-02	7.54E-11
GO:0016477	cell migration	2.61E-05	8.84E-08
GO:0050727	regulation of inflammatory response	3.20E-05	8.84E-08
GO:0042330	taxis	1.89E-02	2.83E-10
GO:0006935	chemotaxis	1.90E-02	2.89E-10
GO:0031347	regulation of defense response	2.14E-02	3.68E-10
GO:0048870	cell motility	4.87E-04	1.75E-08
GO:0046649	lymphocyte activation	2.36E-03	9.09E-09
GO:0007155	cell adhesion	3.48E-03	8.81E-09
GO:0022610	biological adhesion	4.20E-03	1.08E-08
GO:0051240	positive regulation of multicellular organismal process	2.35E-03	3.23E-08
GO:2000145	regulation of cell motility	1.38E-02	6.05E-09
GO:0030334	regulation of cell migration	8.10E-03	1.05E-08
GO:0002253	activation of immune response	1.97E-02	7.08E-09
GO:0051241	negative regulation of multicellular organismal process	5.21E-06	2.88E-05
GO:0002683	negative regulation of immune system process	2.12E-02	2.82E-08
GO:0040012	regulation of locomotion	2.32E-02	2.79E-08
GO:0050896	response to stimulus	1.98E-02	3.49E-08
GO:0023052	signaling	1.13E-02	7.95E-08
GO:0007267	cell-cell signaling	5.74E-03	2.87E-07
GO:0051270	regulation of cellular component movement	1.14E-02	2.24E-07
GO:0030335	positive regulation of cell migration	1.13E-02	2.94E-07
GO:0060333	interferon-gamma-mediated signaling pathway	5.37E-03	7.08E-07
GO:2000147	positive regulation of cell motility	1.48E-02	4.98E-07
GO:0051272	positive regulation of cellular component movement	1.04E-02	1.36E-06
GO:0040017	positive regulation of locomotion	2.28E-02	8.22E-07
GO:0023051	regulation of signaling	9.52E-04	2.59E-05
GO:0002521	leukocyte differentiation	3.83E-02	7.14E-07
GO:0050920	regulation of chemotaxis	4.71E-02	6.61E-07
GO:0010646	regulation of cell communication	5.00E-04	8.84E-05
GO:0042127	regulation of cell proliferation	3.51E-04	1.68E-04
GO:0048518	positive regulation of biological process	4.91E-03	1.99E-05

GO Term	Description	FFPE (DOWN) FDR	FF (UP) FDR
GO:0032501	multicellular organismal process	1.12E-02	1.04E-05
GO:0010466	negative regulation of peptidase activity	3.54E-04	4.00E-04
GO:0060326	cell chemotaxis	1.28E-03	1.15E-04
GO:0048522	positive regulation of cellular process	6.44E-03	3.14E-05
GO:0030595	leukocyte chemotaxis	2.02E-03	1.06E-04
GO:0030593	neutrophil chemotaxis	1.32E-04	2.43E-03
GO:2000026	regulation of multicellular organismal development	7.09E-05	5.78E-03
GO:0006928	movement of cell or subcellular component	3.55E-02	1.85E-05
GO:0034341	response to interferon-gamma	2.98E-02	2.25E-05
GO:0009966	regulation of signal transduction	1.13E-03	7.26E-04
GO:0050793	regulation of developmental process	7.30E-05	1.38E-02
GO:0010647	positive regulation of cell communication	9.87E-03	1.16E-04
GO:0010951	negative regulation of endopeptidase activity	1.36E-03	8.94E-04
GO:0071621	granulocyte chemotaxis	1.54E-04	8.48E-03
GO:0023056	positive regulation of signaling	9.56E-03	1.59E-04
GO:0050921	positive regulation of chemotaxis	3.08E-03	6.13E-04
GO:0065009	regulation of molecular function	3.48E-02	7.48E-05
GO:0009967	positive regulation of signal transduction	5.75E-03	4.97E-04
GO:0030198	extracellular matrix organization	2.54E-02	1.32E-04
GO:0002673	regulation of acute inflammatory response	3.91E-03	8.91E-04
GO:0043062	extracellular structure organization	2.90E-02	1.31E-04
GO:0008284	positive regulation of cell proliferation	1.89E-02	2.10E-04
GO:1902531	regulation of intracellular signal transduction	3.90E-02	1.08E-04
GO:1902533	positive regulation of intracellular signal transduction	1.40E-02	3.24E-04
GO:0052547	regulation of peptidase activity	7.29E-03	6.27E-04
GO:0052548	regulation of endopeptidase activity	1.14E-02	5.91E-04
GO:1990266	neutrophil migration	8.35E-04	8.45E-03
GO:0018212	peptidyl-tyrosine modification	4.18E-02	2.95E-04
GO:0070098	chemokine-mediated signaling pathway	4.25E-03	3.69E-03
GO:0097530	granulocyte migration	7.94E-04	2.18E-02
GO:0050731	positive regulation of peptidyl-tyrosine phosphorylation	4.60E-02	1.41E-03
GO:0045937	positive regulation of phosphate metabolic process	2.30E-02	3.11E-03
GO:0010562	positive regulation of phosphorus metabolic process	2.32E-02	3.12E-03
GO:0002920	regulation of humoral immune response	5.75E-03	1.53E-02
GO:0048585	negative regulation of response to stimulus	4.09E-03	2.54E-02
GO:0042327	positive regulation of phosphorylation	1.92E-02	6.13E-03
GO:0001934	positive regulation of protein phosphorylation	2.17E-02	5.86E-03
GO:0045861	negative regulation of proteolysis	7.35E-03	2.02E-02
GO:0043410	positive regulation of MAPK cascade	1.36E-02	1.33E-02
GO:0032102	negative regulation of response to external stimulus	2.35E-02	7.96E-03
GO:0022603	regulation of anatomical structure morphogenesis	2.30E-02	2.11E-02

PF1 DOWNREGULATED TERMS IN BOTH FFPE AND FF COHORTS (Top 100 Terms Only)

GO term	Description	FFPE FDR	FF FDR
GO:0006952	defense response	1.83E-22	1.74E-08
GO:0032501	multicellular organismal process	3.85E-05	8.19E-23
GO:0006954	inflammatory response	6.99E-14	7.64E-11
GO:0006955	immune response	7.39E-18	4.98E-04
GO:0032502	developmental process	4.12E-06	1.21E-15
GO:0019221	cytokine-mediated signaling pathway	2.98E-17	4.33E-04
GO:0051239	regulation of multicellular organismal process	2.56E-10	8.31E-11
GO:0007186	G-protein coupled receptor signaling pathway	5.34E-13	1.65E-07
GO:0032101	regulation of response to external stimulus	3.20E-11	1.66E-07
GO:0048584	positive regulation of response to stimulus	2.70E-10	6.95E-08
GO:0048856	anatomical structure development	2.68E-03	2.54E-14
GO:0007166	cell surface receptor signaling pathway	1.01E-10	3.19E-06
GO:0051240	positive regulation of multicellular organismal process	8.76E-11	4.59E-06
GO:0002376	immune system process	5.63E-13	2.32E-03
GO:0050896	response to stimulus	7.59E-06	1.88E-10
GO:0048869	cellular developmental process	4.93E-05	3.28E-11
GO:0048583	regulation of response to stimulus	1.06E-09	1.61E-06
GO:0023052	signaling	1.49E-10	2.19E-05
GO:0050727	regulation of inflammatory response	3.42E-09	1.39E-06
GO:0045937	positive regulation of phosphate metabolic process	3.80E-06	1.26E-09
GO:0010562	positive regulation of phosphorus metabolic process	3.86E-06	1.29E-09
GO:0007267	cell-cell signaling	3.19E-11	1.78E-04
GO:0003008	system process	4.16E-05	1.38E-10
GO:0010646	regulation of cell communication	3.85E-06	2.72E-09
GO:0042330	taxis	1.39E-07	8.32E-08
GO:0006935	chemotaxis	1.44E-07	8.43E-08
GO:0060326	cell chemotaxis	1.49E-07	1.08E-07
GO:0023051	regulation of signaling	9.47E-06	2.73E-09
GO:0043410	positive regulation of MAPK cascade	2.22E-07	1.47E-07
GO:0042127	regulation of cell proliferation	3.16E-04	1.34E-10
GO:0001934	positive regulation of protein phosphorylation	2.93E-06	3.34E-08
GO:0032103	positive regulation of response to external stimulus	2.57E-08	4.32E-06
GO:0043408	regulation of MAPK cascade	1.29E-06	1.07E-07
GO:0048518	positive regulation of biological process	1.94E-03	8.58E-11
GO:0006959	humoral immune response	3.47E-09	5.37E-05
GO:0002526	acute inflammatory response	2.23E-05	1.19E-08
GO:0050877	nervous system process	2.33E-08	1.25E-05
GO:0051174	regulation of phosphorus metabolic process	2.89E-05	1.76E-08
GO:0030154	cell differentiation	3.47E-03	1.65E-10
GO:0019220	regulation of phosphate metabolic process	2.62E-05	2.56E-08
GO:0007165	signal transduction	1.31E-06	5.59E-07
GO:0030198	extracellular matrix organization	2.78E-05	3.38E-08
GO:0043062	extracellular structure organization	2.75E-05	3.84E-08
GO:0042327	positive regulation of phosphorylation	4.48E-06	2.55E-07
GO:2000026	regulation of multicellular organismal development	1.55E-06	8.03E-07
GO:0030595	leukocyte chemotaxis	1.43E-06	1.05E-06
GO:0007154	cell communication	1.88E-07	8.07E-06
GO:0001932	regulation of protein phosphorylation	9.04E-06	1.69E-07
GO:0023056	positive regulation of signaling	3.24E-05	5.02E-08
GO:0050900	leukocyte migration	4.89E-06	4.31E-07
GO:0010647	positive regulation of cell communication	2.27E-05	9.50E-08
GO:0051241	negative regulation of multicellular organismal process	6.83E-08	5.56E-05
GO:0001525	angiogenesis	1.67E-02	2.36E-10
GO:0048646	anatomical structure formation involved in morphogenesis	7.37E-03	8.76E-10
GO:0046903	secretion	5.02E-07	1.30E-05
GO:0008284	positive regulation of cell proliferation	1.25E-03	8.77E-09
GO:0009653	anatomical structure morphogenesis	1.05E-03	1.34E-08
GO:0048522	positive regulation of cellular process	6.44E-03	2.47E-09
GO:0051094	positive regulation of developmental process	1.85E-07	9.80E-05
GO:0009967	positive regulation of signal transduction	1.54E-04	1.46E-07

GO term	Description	FFPE FDR	FF FDR
GO:0050793	regulation of developmental process	1.20E-04	3.03E-07
GO:0040011	locomotion	1.69E-06	2.55E-05
GO:0006953	acute-phase response	8.86E-05	5.40E-07
GO:0033993	response to lipid	3.74E-03	2.47E-08
GO:1902533	positive regulation of intracellular signal transduction	3.46E-04	3.11E-07
GO:0065008	regulation of biological quality	8.05E-03	1.37E-08
GO:0030593	neutrophil chemotaxis	4.60E-06	2.80E-05
GO:0071621	granulocyte chemotaxis	5.78E-06	2.39E-05
GO:2000147	positive regulation of cell motility	2.72E-03	5.17E-08
GO:0045765	regulation of angiogenesis	5.67E-05	2.90E-06
GO:0030335	positive regulation of cell migration	2.04E-03	8.76E-08
GO:0070098	chemokine-mediated signaling pathway	4.10E-07	6.08E-04
GO:0007610	behavior	1.53E-08	1.64E-02
GO:0009966	regulation of signal transduction	2.30E-04	1.10E-06
GO:0009888	tissue development	1.07E-02	2.44E-08
GO:0051272	positive regulation of cellular component movement	2.63E-03	1.19E-07
GO:0016477	cell migration	2.79E-06	1.24E-04
GO:0009605	response to external stimulus	2.82E-04	1.76E-06
GO:0042325	regulation of phosphorylation	6.43E-05	7.78E-06
GO:0097529	myeloid leukocyte migration	7.68E-06	7.03E-05
GO:1990266	neutrophil migration	5.84E-06	1.04E-04
GO:0001817	regulation of cytokine production	6.47E-08	1.25E-02
GO:0097530	granulocyte migration	1.04E-05	8.38E-05
GO:0002687	positive regulation of leukocyte migration	8.54E-05	1.04E-05
GO:0002682	regulation of immune system process	8.84E-07	1.20E-03
GO:0022603	regulation of anatomical structure morphogenesis	5.67E-03	2.74E-07
GO:0043207	response to external biotic stimulus	6.18E-06	3.79E-04
GO:0050920	regulation of chemotaxis	2.64E-03	1.17E-06
GO:1901342	regulation of vasculature development	5.99E-04	5.18E-06
GO:0002690	positive regulation of leukocyte chemotaxis	1.43E-04	2.35E-05
GO:0001819	positive regulation of cytokine production	6.47E-06	6.74E-04
GO:0040017	positive regulation of locomotion	1.06E-02	4.20E-07
GO:0030334	regulation of cell migration	6.02E-04	9.80E-06
GO:0009607	response to biotic stimulus	1.34E-05	6.54E-04
GO:0050921	positive regulation of chemotaxis	3.87E-04	2.56E-05
GO:0048870	cell motility	2.11E-05	5.03E-04
GO:0009617	response to bacterium	4.21E-04	2.91E-05
GO:0030155	regulation of cell adhesion	2.99E-06	5.35E-03
GO:0050804	modulation of chemical synaptic transmission	1.02E-06	1.73E-02
GO:0050729	positive regulation of inflammatory response	3.84E-05	5.01E-04

PF2 UPREGULATED TERMS IN BOTH FFPE AND FF COHORTS (All Terms)

GO Term	Description	FFPE FDR	FF FDR
GO:0002376	immune system process	1.36E-02	3.10E-37
GO:0006952	defense response	6.41E-05	4.02E-24
GO:0030198	extracellular matrix organization	1.01E-02	5.60E-26
GO:0043062	extracellular structure organization	1.07E-02	5.77E-26
GO:0007165	signal transduction	1.03E-02	1.02E-25
GO:0007155	cell adhesion	4.18E-09	9.40E-16
GO:0022610	biological adhesion	4.93E-09	1.46E-15
GO:0006955	immune response	6.44E-07	1.38E-17
GO:0001775	cell activation	1.87E-02	2.43E-21
GO:0032940	secretion by cell	3.48E-05	1.95E-15
GO:0046903	secretion	1.73E-04	1.66E-15
GO:0006954	inflammatory response	1.13E-02	5.82E-17
GO:0006887	exocytosis	1.67E-03	4.89E-16
GO:0045055	regulated exocytosis	2.23E-03	1.95E-15
GO:0030334	regulation of cell migration	3.65E-02	1.50E-14
GO:0050900	leukocyte migration	1.78E-03	3.39E-13
GO:0030155	regulation of cell adhesion	9.18E-03	2.24E-13
GO:0032101	regulation of response to external stimulus	7.39E-04	4.11E-12
GO:2000145	regulation of cell motility	1.51E-02	2.26E-13
GO:0051270	regulation of cellular component movement	2.97E-02	1.29E-13
GO:0051239	regulation of multicellular organismal process	1.69E-03	3.01E-12
GO:0098609	cell-cell adhesion	1.12E-08	1.15E-06
GO:0032501	multicellular organismal process	6.58E-04	3.40E-11
GO:0040012	regulation of locomotion	3.66E-02	1.08E-12
GO:0032502	developmental process	2.89E-02	1.50E-12
GO:0051240	positive regulation of multicellular organismal process	3.88E-03	1.93E-11
GO:0032879	regulation of localization	1.07E-05	2.12E-08
GO:0040011	locomotion	3.39E-02	3.15E-10
GO:0007154	cell communication	2.66E-09	6.94E-03
GO:0050865	regulation of cell activation	2.90E-02	2.29E-09
GO:0042330	taxis	1.67E-03	5.34E-08
GO:0006935	chemotaxis	1.70E-03	5.39E-08
GO:0001817	regulation of cytokine production	3.52E-03	2.79E-08
GO:0003008	system process	5.19E-06	4.86E-05
GO:0001819	positive regulation of cytokine production	1.13E-03	3.69E-06
GO:0045087	innate immune response	1.55E-02	4.37E-07
GO:0043270	positive regulation of ion transport	8.23E-06	8.26E-04
GO:0002697	regulation of immune effector process	4.96E-02	7.25E-07
GO:0060326	cell chemotaxis	3.48E-02	1.72E-05
GO:0010959	regulation of metal ion transport	2.18E-04	2.75E-03
GO:0006898	receptor-mediated endocytosis	1.96E-02	2.75E-04
GO:0051050	positive regulation of transport	3.31E-03	1.69E-03
GO:1903522	regulation of blood circulation	9.00E-03	9.14E-04
GO:0010035	response to inorganic substance	3.67E-02	6.37E-04
GO:0051234	establishment of localization	8.94E-03	3.59E-03
GO:0006022	aminoglycan metabolic process	2.42E-02	1.46E-03
GO:0050707	regulation of cytokine secretion	4.76E-02	9.35E-04
GO:0007162	negative regulation of cell adhesion	4.94E-02	1.01E-03
GO:0030203	glycosaminoglycan metabolic process	3.40E-02	1.95E-03
GO:0008016	regulation of heart contraction	7.59E-03	1.90E-02
GO:1903510	mucopolysaccharide metabolic process	2.26E-02	8.97E-03

Appendix 6: Publications

Published abstracts

Ritzmann, T., Rogers, H., et al., 2016. "EPN-22 Recurrent paediatric ependymoma: a retrospective review of 194 cases". *Neuro-Oncology*. **18**(Suppl3): iii35.

Ritzmann, T., Rogers, H., et al., 2017. "EPND-03 RNA-Seq gene expression analysis of 106 formalin-fixed paraffin-embedded paediatric ependymomas". *Neuro-Oncology*. **19**(Suppl_4): iv15.

Published editorials

Ritzmann, T., Grundy, RG., et. al., 2016. "The Molecular Era of Brain Tumour Research is upon us: But what now for our Patients?". *J Brain Tumors Neurooncol*. **1**:e104.

Articles in press

Ritzmann, T., Grundy, RG., 2018. "Translating childhood brain tumour research into clinical practice: The experience of molecular classification and diagnostics". *Paediatrics and Child Health*.

Articles submitted for peer review

Ritzmann, T., Rogers, H., et al., "A retrospective analysis of recurrent paediatric ependymoma reveals extremely poor survival and ineffectiveness of current treatments across paediatric molecular subgroups".

Articles in preparation

Ritzmann, T., Rogers, H., et al., "RNA sequencing of 106 FFPE paediatric ependymomas with storage times of up to 30 years"

Ritzmann, T., Rogers, H., et al., "Transcriptomic patterns of paediatric ependymoma recurrence