Identification of cancer genes using Sleeping Beauty mutagenesis of the *Ptch1* murine tumour model

Hani Mohammad Al-Afghani

Thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Newcastle University

Faculty of Medical Sciences

Institute of Genetic Medicine

November 2013

Abstract

Medulloblastoma (MB) is a paediatric tumour of the cerebellum which is responsible for 15-20% of all childhood brain tumours. Mortality due to this disease is high (~40%) and successful treatment is associated with significant neurological and cognitive consequences, making new therapies desirable. Disruption of the Sonic Hedgehog (SHH) signaling pathway, including mutations in *PTCH1*, define a major subset of human MB. Mice heterozygous for loss of function mutations in the *Ptch1* ortholog develop MBs at low frequency.

To identity genes that co-operate with *Ptch1* in MB development, a Sleeping Beauty insertional mutagenesis screen in this murine model was performed. Mutagenesis significantly increased the frequency of MB formation in *Ptch* heterozygote mice from ~3% to ~25% after 8 months ($p<0.0001$). To identify the genes responsible for this enhanced tumour formation, Splinkerette-PCR and 454-FLX sequencing were applied to define SB insertion sites within 40 tumours. Statistical analysis of these data using Monte Carlo and Kernel Convolution methods identified 18 candidate medulloblastoma genes defined by common insertion sites (CISs), including 11 identified by both methods used. Many of these are genes known to be involved in neuronal development and cell fate determination. Subsequent ARACNe network analysis of the candidate genes in human gene expression datasets has established that seven (*Nfia, Nfib, Tead1, Tgif2, Myt1l, Fgf13 and Crebbp*) lie within a single network which is enriched for both neuronal genes and transcription factors, and includes genes known to interact with the SHH pathway. Bioinformatic analysis confirmed that the tumours generated from SB mutagenised mice were similar in their gene expression pattern to human SHH subgroup tumours, while microarray expression analysis of SB induced tumours identified *Igf2*, a gene previously implicated in MB development, as a key output of network activity.

Finally, analysis of gene expression within granule neuron precursor cells (GNPCs), the cell of origin of SHH subgroup MB, identified *Tgif2* and *Myt1l* as prime candidate genes for downstream functional analysis. Lentiviral based modulation of their gene expression was therefore attempted in primary GNPCs. shRNA based knock-down of *Myt1l* was achieved, and shown to be associated with significantly increased expression of both *Gli1* (the downstream effector of the Shh pathway) and *Math1* (a marker of undifferentiated neuronal cells), consistent with a reduction in *MYT1L* expression playing a role in human MB development. Further analysis of this gene, and others defined here, should both improve our understanding of MB and define potential targets for therapeutic intervention.

Declaration

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

Several Figures are adapted from a publication on which I am co-author [Lastowska et al. (2013) Acta Neurocommunications 1:35 (http://www.actaneurocomms.org/content/1/1/35#B63)], which summarises the major findings of this research.

Hani Al-Afghani

01 November 2013

Dedication

To my great parents who supported me by their health, wealth and prayers. In particular to my beloved mother who passed away at the beginning of master course.

To my beloved and great friend and wife Riham and great daughter Hala and son Mohammad who acclimatized me to the environment for studying.

To my great home country Kingdom of Saudi Arabia.

Acknowledgement

I should like to express my sincere gratitude to my research supervisors, Dr Michael Jackson and Professor Sir John Burn, for their continued encouragement, enthusiasm and support during my PhD studies. I feel extremely privileged to have been a part of their group.

I would like to express my special thanks to Dr Maria Lastowska, for her greate support in terms of technical helps especially at the beginning of this project and for providing her samples to generate some of the results presented here.

I would like give a great thank you for Dr Daniel Williamson for his great support in terms of Cell culture and Viral experiments.

I would like to thanks all my friends at the Institute of Genetic medicine for their help and support during this time of my life, technically and emotionally.
A special thank you is reserved for my close family who offered their endless support via encouragement and kindness.

I also preserve huge thanks and gratitude to my home country for endless support

financially and emotionally to achieve this higher educational level.

Table of Contents

Table of Figures

List of Tables

List of Abbreviations

CNS	Central Nervous System
WHO	World Health Organisation
ENM	Ermed Extraneural Metastasis
PNET	Primitive Neuroectodermal Tumour
MBEN	Medulloblastoma With Extensive Nodularity
MB	Medulloblastoma
CGH	Comparative Genomic Hybridization
SHH	Sonic Hedgehog
MMAPK	P38 Mitogen-Activated Protein Kinase Pathway
DHH	Desert Hedgehog
IHH	Indian Hedgehog
GNPCs	Granule Neuron Precursor Cells
SUFU	Suppressor-Of-Fused
SMO	Smothened
CSC	Cancer Stem Cell
APC	Adenomatous Polyposis Coli
FZD	Frizzled Receptor
EGFR	Epidermal Growth Factor Receptor
BMP	Bone Morphogenetic Protein
MAPK	Mitogen-Activated Protein Kinase
TSA	Trichostatin A
SB	Sleeping Beauty
SA	Splice Acceptor
SD	Splice Donor
PA	Poly-Adenylation
MSCV	Murine Stem Cell Virus
BCN	Basal Cell Nevus
BCC	Basal Cell Carcinomas
Rb	Retinoblastoma
VZ	Ventricular Zone
EGL	External Granular Layer
CPCs	Cancer Propagating Cells

BCPCs	Brain Cancer Propagating Cells
ER	Estrogen Receptor
BrDu	Bromodeoxyuridine
PCR	Polymerase Chain Reaction
TA	Thymine And Adenine Sequence
CT	Connective Tissue
LM-PCR	Linker Mediated-PCR
RMS	Rhabdomyosarcomas
Lym	Lymphoma
MC	Monte Carlo Statistical Method
GKC	Gaussian Kernel Convolution Statistical Method
CISs	Common Insertion Sites
MOA	Inferred Mode Of Action
LA	Liver Adenoma
AML	Acute Myeloid Leukemia
NGS	Next Generation Sequencing
WB	Western Blot
E	Exonic Primers
Q PCR	Quantitative PCR
GO	Gene Ontology
LRLP	Lower Rhombic Lip Progenitors
ARACNe	Algorithm For The Reconstruction Of Accurate Cellular Networks
KM	Kaplan-Meier Plot
RIN	RNA Integrity Number
cRNA	Complimentary RNA
SVM	Support Vector Machine
NMF	Non-Negative Matrix Factorisation
RNA	Ribosomal Nucleic Acid
DNA	Dinucleotide Nucleic Acid
SV	Synaptic Vesicle
TF	Transcription Factors
HPE	Holoprosencephaly
ORF	Open Reading Frame
cDNA	Complimentary DNA
shRNA	Short Hairpin RNA
Ct	Cycle Threshold

IM	Immunochemistry
WT	Wild Type
MOI	Multiplicity Of Infection
Ab	Antibody
KD	Kilo Dalton

Chapter 1

INTRODUCTION

1.1. Cancer

Cancer, which can be defined as uncontrolled cell proliferation which exceeds the body's normal restrictions, is one of the leading causes of death worldwide. It was responsible for approximately 7.9 million deaths in 2008, and if current trends continue this figure is expected to dramatically increase by 2030, reaching a maximal level of about 13.2 Million (World Health Organisation, 2013a). However, in 2008 the estimated cancer incidence (as opposed to mortality) was 12.7 million new cases per annum, and this is likely to rise to approximately 22 million by 2030 worldwide if the present trends in cancer are continued (World Health Organisation, 2013a). In the United Kingdom (UK), it is estimateded that 1 out of every 3 persons will be affected by one or more forms of cancer during their lifetime (Cancer Research UK, 2013a). Almost all body organs can be affected by cancer. The incidence of the most common cancers in the UK, and the numbers of deaths they caused are shown in Figures 1.1. and 1.2 for the year 2008. The 4 most common cancers (breast, lung, colorectal and prostate cancer) are responsible for approximately 47% of cancer deaths (Figure 1.2, Cancer Research UK, 2013b).

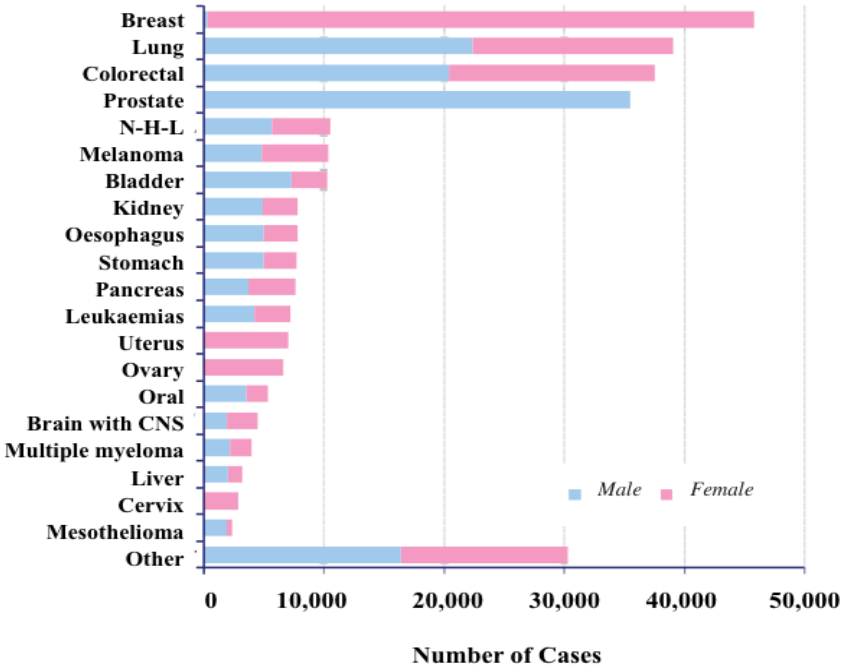

Figure 1.1: New cases of the most common cancers in the UK in 2008, split by gender. N-H-L = Non Hodgkin Lymphoma (adapted from www.cancerresearchuk.org).

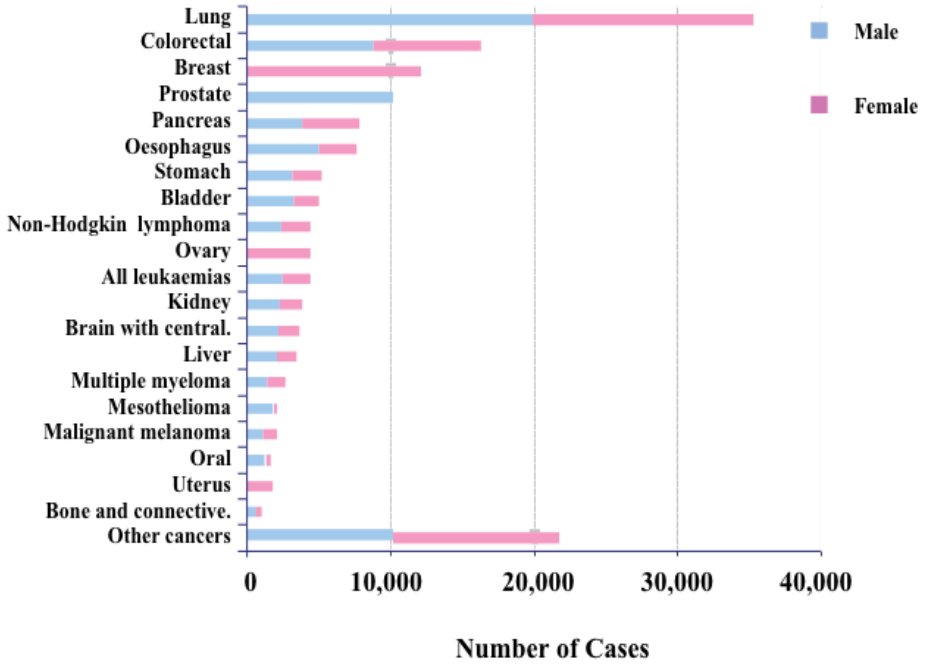

Figure 1.2: Deaths due to the most common cancers in the UK in 2008, split by gender (Adapted from www.info.cancerresearchuk.org).

Carcinogenesis arises as a consequence of mutations affecting single cells. These mutations usually are due to either external (environmental) or internal (inherited) factors, or a combination of both. Major environmental agents include radiation and carcinogens integrated within chemicals, food additives and tobacco smoke. Multiple mutagenic insults are generally required for normal cells to be transformed into malignant invasive cells. This could happen via two main mechanisms. Firstly, stimulation of cellular proliferation, which may be caused by specific mutations that in turn may stimulate target cell expansion favouring further mutations. The second, through a mechanism known as chromosomal instability, that leads to increase further mutation rates (Hanahan and Weinberg, 2011).

A recent review has highlighted six important functional changes as the essential transformative features of all immortalised cancerous cells: 1 - self-sufficiency from growth signals, 2 - insensitivity to anti-growth signals, 3 - Avoidance of apoptosis, 4 - unlimited replication, 5 - continued angiogenesis, and 6 - tissue invasion and metastasis (Hanahan and Weinberg, 2011). Although functional changes involved in cancer can be defined into these six categories, individual genes which are mutated in cancer are often classified as either genes which have the ability to inhibit tumourgenesis (tumour supressors) or genes which promote tumourgenesis (oncogenes). Mutations affecting these two classes of gene are loss of function and gain of function respectively, and can drive healthy cells toward malignancy (Osborne *et al.*, 2004;Zhang *et al.*, 2007).

Oncogenes are the transformed stages of proto-oncogenes (normal genes with oncogenic potential) after a gain of function mutation, and have the capability to promote the cancer phenotype. Individuals may be highly tumour prone when a proto-oncogene within their somatic cells becomes oncogenic as a result of chromosomal instability, gene amplification, or mutation. A large number of proto-oncogenes have been identified and many of them, such as *RAS* (rat sarcoma), *MYC* (myelocytomatosis oncogene cellular homolog), *ERK* (extracellular signal-regulated kinase), and *EGFR* (epidermal growth factor receptor), encode proteins that are responsible for managing cell regulation and differentiation (Todd and Wong, 1999);(Vogelstein and Kinzler, 2004) . Oncogenes can be further subdivided into five main subtypes. These subtypes are intracellular transduction system (cell surface receptors), growth factors, DNA binding nuclear proteins (e.g. *c-Myc*), cyclin-dependent kinases (e.g. *CDK4*) and cyclin-dependent kinase inhibitors (e.g. *CDKN2D*) (Osborne *et al.*, 2004).

Tumour suppressor genes are those genes where loss of function mutations increase the risk of tumourgenesis (Haber and Harlow, 1997)). In the context of cancer,

the normal function of these genes can be considered as being protective against uncontrolled cellular overgrowth. Alterations that lead to down-regulation or ablation of the function of such genes will compromise their protective function as overgrowth inhibitors, leading to conversion of normal cells into malignant invasive cells (Knudson, 1993).

Tumour suppressor genes can be further classified into three broad groups according to their function: **1. Caretakers:** These encode proteins that protect the genome from mutation accumulation (i.e. indirect protection) such that their mutation creates genomic instability. Examples would be genes involved in DNA-repair, such as *MLH1* and *MSH2* which are frequently mutated in some forms of colorectal cancer. **2. Gatekeepers:** These encode proteins that limit cellular overgrowth by encouraging affected cells to commit to the apoptotic cell death mechanism. The classic example is *p53* which is mutated in a wide variety of cancers (Macleod, 2000). Both of these kinds of genes required two hits in order to affect both alleles and promote malignancy, and it is noteworthy that individuals carrying inherited mutations of such genes are at a greatly increased risk of developing tumours relative to the population as a whole. For example, Lynch syndrome is an inherited predisposition to colorectal cancer defined by inheritance of mutant forms of *MLH1* or *MSH2*, while LiFraumeni syndrome is a familial cancer predisposition caused by inheritance of *p53* loss of function mutations (Macleod, 2000). **3. Landscapers:** These encode proteins that are involved in the tumourigenesis process after mutation, for example by modifying the microenvironment of the tumour cells, such as genes which control the extra cellular matrix (ECM) proteins (Macleod, 2000; Kinzler and Vogelstein, 1997; Kinzler and Vogelstein, 1998).

1.2. Childhood cancers

The overall incidence of cancer within the population is highly age dependent, consistent with an increased mutagenic burden associated with increasing age. As an individual's age increases, the opportunities for their cells to be exposed to carcinogens and gain mutations will also increase. Despite this, children have a high risk of developing specific forms of cancer, mainly leukaemias, lymphomas, central nervous system (CNS) tumours, and bone tumours. Yearly, about 1500 new cancer cases are reported in children in the UK. The mortality rates in children affected by the most common cancers (Haematological and CNS malignancy) under the age of 15 years in the UK between 1996-2005 are shown in Figure 1.3. Leukaemia (blood cancer) accounts for approximately 30% of childhood cancers. Improvement in diagnostic and

therapeutic procedures over the last few decades have led to gradual improvement in survival rates within effected children, although this varies between different tumours. For instance, survival rates have improved to about 80% in children with acute lymphoplastic leukaemia (ALL) – the most common blood malignancy effecting childhood (Tubergen *et al.*, 1993), whereas the brain tumours survival rate has remainded relatively unaltered at approximately 65% (Cancer Research UK, 2013c).

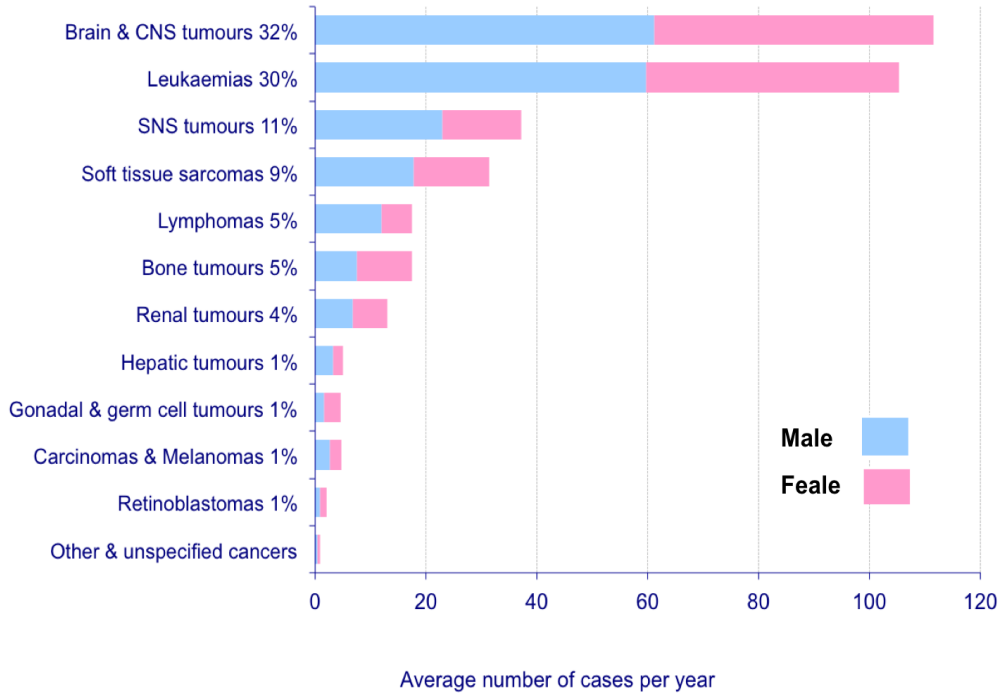

Figure 1.3 The most common cancer causing death in children between 1996-2005 in the UK (Adabted from Cancer research UK, 2013).

1.2.1 Brain tumours

Central nervous system (CNS) tumours are the second most common group of tumours affecting children after blood malignancies, and the incidence of major subgroups in relation to age is shown in Figure 1.4. They account for approximately 27% of all tumours affecting children up to the age of 14 years in the UK over the period from 2008 – 2010 (Cancer Research UK, 2013). Histologically, brain tumours can be subdivided into glial tumours and nonglial (neuronal) tumours. Gliomas, which account for a majority of brain tumours in childhood, are generally classified on the basis of shared similarities with specific lineages of glial cells. In line with this scheme, tumour cells resembling astrocytes are classified as astrocytomas, while those similar to oligodendrocytes are classified as oligodendrogliomas (Rich and Eyler, 2008). During

the period 1996–2005, astrocytomas accounted for about 43% of all childhood brain tumours, with low grade tumours constituting approximately 76% of these. The second most common subgroups of the CNS tumours are the intracranial and intraspinal embryonal tumours and these account for about 19% of all childhood brain tumours. Among these, 73% are medulloblastomas. Medulloblastomas (MBs), are found in the posterior fossa. The least frequent subgroup of CNS tumours is ependymomas, which account for only 10% (Cancer research UK, 2013).

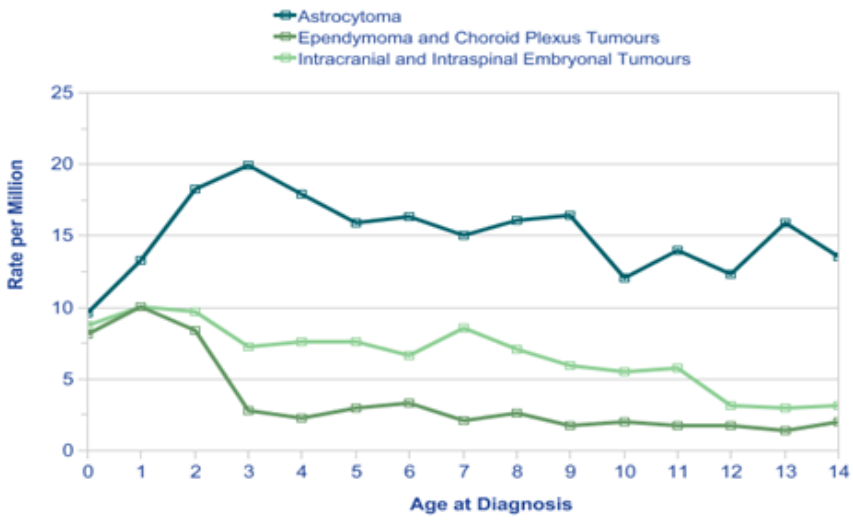

FIGURE 1.4 Frequency of Brain and CNC cancers in children, according to age (Adapted from cancerresearchuk.org).

In general, neurologic tumours present challenges to clinicians as the biology of most of these tumours is poorly defined. Indeed, the fact that the pathogenesis of these tumours is poorly understood has impacted negatively on the nature of treatment options available. While recent advances have considerably improved the treatment options for other forms of cancer such as leukaemias, the median survival for primary brain tumours has remained low (15 months for glioblastoma), although, slight improvement in overall survival for primary brain tumours has been observed over the last 30 years (Rich and Eyler, 2008). However, a recent comparison of survival rates for brain tumours in the UK and US revealed that children and young adults diagnosed with CNS tumours in general in the UK have relativelypoor survival rates (children with MB in particular, suffer from relatively high recurrence rates within 2 years of diagnosis (Mathew et al., 2014).

1.3. Medulloblastoma

1.3.1. Overview of disease

Medulloblastoma (MB) is responsible for between 20 and30% of all childhood brain tumours (Li *et al.*, 2013). In 2007, the World Health Organisation (WHO) classified medulloblastoma as a highly severe (grade IV) form of CNS tumour. This new classification was based on the consensus reached by an international working group consisting of pathologists, geneticists and other international experts. The international classification of disease for oncology (ICD-O) was founded more than 30 years ago and in the fourth edition of this system, publically available in 2007, medulloblastoma was risk stratified into two groups (high risk and average risk), depending on characterised macroscopical morphology, location, age, and biological behaviour (Louis *et al.*,2007). However, with the inclusion of immunohistochemical and recently analysed molecular genetic markers, MB tumour stratification has been revised into three risk groups with highly significantly different outcomes ($P<0.0001$); low-risk, standard-risk and High-risk, (Ellison *et al.*, 2011). Low-risk MBs (13% of total) are β-catenin nucleopositive, with no metastasis, large-cell/anaplasia (LC/A) or *MYC* expression. High-risk MBs (28%) are tumours with metastasis, LC/A appearance and *MYC* expression. Finally, those tumours lacking the above parameters (59%) are defined as standard-risk (Ellison *et al.*, 2011).

Macroscopically, medulloblastoma is defined as a tumour of the midline that arises in the cerebellar vermis or/and hemispheres. Moreover, the classification states that medulloblastoma is a poorly outlined pinkish or greyish soft tissue growth, possibly with foci of haemorrhages, necrosis or calcification (Batra *et al.*, 1995;Ellison *et al.*, 2011). Approximately 7-10% of all cases are associated with the spreading of the disease outside the CNS; termed extraneural metastasis or ENM, (Mazloom *et al.*, 2010). In children, the various morphological phenotypes of medulloblastoma account for 40% of tumours of the cerebellum, and 33% of posterior cerebral fossa tumours. The annual incidence of medulloblastoma is 6.0 per million populations per year, with a significantly higher incidence in boys than girls. In children, the incidence peaks between 5 and 7 years of age. Overall, 10% of cases are diagnosed within the first year of life, while approximately 30% of all cases occur in children above 10 years of age. However, less than 1% of MBs occur in adulthood. Furthermore, research indicates that medulloblastoma is more common in Caucasians than in other races. (Dhall, 2009).

With the recent use of multimodal treatment of MB, the overall five year survival rates for patients with high-risk and average-risk MB are about 70% and 90% respectively (Gajjar *et al.*, 2006). However, the survival rates for disseminated disease, associated with poorer prognostic outcomes and recurrent MB, is still low at approximately 36% (Packer *et al.*, 2006). In addition, patients who undergo treatment and survive the disease often have long term side effects, such as bone marrow suppression, growth and hormonal disruption, impaired fertility, and learning, behavioural, and coordination problems (Massimino *et al.*, 2011).

*1.3.2. Histopathological and Molecular Classification of Medulloblastom*a

Historically, medulloblastomas and other neuro-embryonal tumours have been classified according to their histological appearance and location within the central nervous system (Rossi *et al.*, 2008). In the first classification schema for neurological tumours, Bailey and Cushing grouped medulloblastoma with other embryonal tumours characterised by rapid proliferation and minimal differentiation (Ellison *et al.*, 2003). To classify embryonal tumours that could not be neatly categorised, Hart and Earle in 1973 introduced the term primitive neuroectodermal tumours (PNET). While the term PNET was originally used to classify cerebral high-grade undifferentiated neuroepithelial tumours of childhood such as medulloblastoma, the term has been extended to include all tumours of a similar nature (Gilbertson and Ellison, 2008). More recently, to take account of new molecular genetic findings, WHO recently developed a new classification scheme for neurologic tumours. Within this scheme the WHO define five histopathologic variants of medulloblastoma. These are: classic medulloblastoma 6470/3, desmoplastic/nodular medulloblastoma 9471/3, medulloblastoma with extensive nodularity (MBEN) 9471/3, anaplastic medulloblastoma, 9474/3, and large-cell medulloblastoma 9474/3 (Navajas and Giralt, 2010).

Molecular profiling techniques have been extensively applied to MBs and genetic mutation and gene expression profiles have proved useful for sub-classification of the disease. Indeed, the observation that the different MB subtypes have distinct cytogenetic profiles provided the first clues that different subtypes may develop via distinct molecular processes. However, the first comprehensive scheme for molecular sub-typing of MBs was proposed by (Kool *et al.*, 2008). These researchers combined the mRNA expression profiles of a series of 62 medulloblastomas (using Affymetrix HG-U133 Plus 2.2 Gene Chips) with comparative genomic Hybridization (CGH) analysis. Unsupervised two-way, hierarchical cluster analysis identified 5 molecular

subtypes of medulloblastomas (subgroups A—E), each with distinct clinicopathological features, characteristic gene expression signatures, and distinct mutations. Cluster A exhibited alterations in WNT and TGF-beta signalling – this cluster corresponds to MBs with alterations in the WNT pathway (including *β-catenin, APC, AXIN*); cluster B was characterised by alterations in sonic hedgehog (SHH) signalling (including *PTCH1* and *SUFU* mutations) and low levels of *OTX2* expression; and the closely-related clusters C, D and E were sub-typed on the basis of the expression of genes involved in neuronal differentiation, biogenesis and cytoskeleton organisation (Kool *et al.*, 2008).

A subsequent study carried out on a larger series (103) of primary human MBs using gene expression profiles, DNA copy number aberrations, and the application of unsupervised hierarchical cluster analysis (Northcott *et al.*, 2011), concluded that there was clear evidence for only 4 clinicogenetic subgroups of MB [defined as WNT, SHH, Group C and Group D (Northcott *et al.*, 2011)]. The SHH subgroup showed a preponderance of desmoplastic MBs, with the exclusive and statistically significant expression of known *SHH* target genes *HHIP, SFRP1* and *MYCN*. The WNT group showed an exclusive statistically significant expression of *WIF1, DKK1,* and *DKK2*. According to age prevalence, the authors identified the SHH subgroup as being prominent within either infants <3years, or adults >16 years old. They also noticed that the known MB oncogenes *OTX2* and *FOXG1B* are highly expressed within group C and D. *MYC*, however, was highly expressed within both WNT and C groups. Regarding gender, the SHH subgroup was more common within males more than females; 1.55:1, respectively. Another important outcome established in this study was a new diagnostic classifier based on immunohistochemistry, with exclusive expression of *DKK1* in WNT subgroup, *SFRP1* in SHH, *NPR3* in C subgroup, and *KCNA1* in D subgroup (Northcott *et al.*, 2011).

In 2012, a further metagene analysis advanced the consensus view that MB can be subdivided into four subgroups, and put forward a revised group nomenclature - with Group C and Group D of Northcott *et al.*, (2011) becoming Group 3 and Group 4 respectively (Taylor et al., 2012). Despite these advances in molecular classification, there is no clear correspondence between the new molecular subgroups and established histopathological subtypes. For instance, half of SHH subtypes present with desmoplastic histology and the remaining half of SHH MBs with anaplastic histopathological features (Hatten and Roussel, 2011). This overlap suggests that further genetic characterisation and identification of further genes involved in MB development

will be required to resolve phenotype/genotype relationships. The following sections will discuss the different MB subgroups in more detail.

1.3.2.1 SHH subgroup tumours

Approximately 25-30% of MB samples are defined by disruption of the SHH signalling pathway (Jones *et al.*, 2012;Kool *et al.*, 2012;Robinson *et al.*, 2012). The most frequent histological feature within the SHH subgroup is desmoplastic histology which is observed in ~89% of tumours from infants (0 < 4 years), and in 25% of tumours from older children (4 – 16 year) (Kool *et al.*, 2012). The age distribution of this subgroup is characterised by a bimodal distribution shape as patients are either infants or adults (>16 years) (Kool *et al.*, 2012). In adults, SHH subgroup tumours account for almost half of all tumours (Remke *et al.*, 2011). In terms of prognosis, patients with SHH supgroup tumours have a better prognosis if they are infants (77% survival) rather than older children (51%) or adults (34%) (Kool *et al.*, 2012).

In regard to the gene expression that characterises this subgroups, several genes known to have roles in both SHH signalling pathway regulation and MB generation are exclusively expressed at a high level in this subgroup, such as *PTCH1*, *SUFU*, *ATOH1*, *PTCH2*, *SERP1*, *SMO* and *GLI1* (Thompson *et al.*, 2006;Kool *et al.*, 2008;Jones *et al.*, 2012). The most frequent chromosomal aberration in SHH tumours is Chr9q deletion, observed in anything from 21–47% of tumours analysed (Kool *et al.*, 2008;Northcott *et al.*, 2011;Kool *et al.*, 2012;Pugh *et al.*, 2012;Li *et al.*, 2013). Genes involved in other pathways, such as P53 and PI3K, have been recently shown to be disrupted in SHH subgroup MBs (*MDM4* and *PPM1D* mutations and focal deletions of *TP53*, and *PIK3C2G* and *PIK3C2B* mutations). This suggests that these pathways could contribute to MB generation beside the SHH pathway (Northcott *et al.*, 2012). Interestingly, (Northcott *et al.*, 2011) established that 55/151 SHH MBs were from infants and were mainly localised in the vermis zone and were commonly metastatic. In contrast, 96/151 SHH subgroup MBs were from adults, arose in hemispheres of the cerebella, and no metastases were observed. This suggests that further clinical and/or genetic subdivisions within the SHH subgroup remain to be defined. In regards to the medulloblastoma created from the *Ptch1* model in this project, there was considerable variation in the size and position of tumours observed. Some were clearly confined to a single hemisphere of the cerebellum, while others affected both hemispheres equally. However, the precise position was not routinely recorded as material was collected on the basis of genotype

and/or the existence of neurological symptoms only. No confirmation of tumour presence was attempted at post-mortem.

1.3.2.2 WNT subgroup tumours

The WNT MB subgroup comprises about 10–15% of all MB samples (Li et al., 2013). The study conducted by (Kool *et al.*, 2012) reported that almost all (97%) of WNT subgroup tumours analysed by immunohistochemistry showed classic histology. WNT supgroup tumours were found to affect children and adults, but seem to be rarely observed in infants (1%). As mentioned above, (Kool *et al.*, 2008) showed that mutation of *CTNNB1* are common in WNT subgroup tumours, and 88.7% of all WNT subgroups tumours analysed by (Northcott *et al.*, 2012) also had mutations of *CTNNB1*. In addition, at the chromosomal level, the complete or partial deletion of chromosome 6 was recorded in 76.3 – 85% of WNT subgroup tumours, making it a hallmark of this group of MBs (Kool *et al.*, 2012;Northcott *et al.*, 2012). In children, this subtype of MB is usually associated with a good prognosis (Northcott et al., 2011). Analysis of the mouse model *(Blbp-Cre, Ctnnb1lox*$^{(ex3)/lox(ex3)}$*, Atoh1-Cre and Tp53*$^{flx/flx}$*),* where the WNT pathway is activated, suggests that the WNT MB subgroup may arise from the lower rhombic lip of the cerebellum (Gibson *et al.*, 2010).

1.3.2.3 Subgroup 3 and 4 tumours

These subgroups of MB have been collectively termed the non-WNT and non-SHH subgroups as they do not map to mutation/activation of known signalling pathways. Both subgroups do have some identifying clinical and molecular features. Most of the tumours in these subgroups are characterised by classical histology. However, at low frequency, subgroups 3 and 4 present with a desmoplastic appearance (Kool *et al.*, 2012). These subgroups rarely present among infants and adults and both subgroups show a 2:1 excess of males to females within all ages analysed (Kool *et al.*, 2012). Furthermore, both subgroups display metastatic features at high frequency at all age groups; 30% for group 3 and 31% for Group 4, rising to 47% for Group 3 and 36% for group 4 tumours when infants only are analysed. Because of the high frequency of metastatic disease in these subgroups, it is not surprising that prognosis for both groups is poor relative to WNT and SHH subgroups.

In terms of cytogenetic alterations, Group 3 is associated with gain of chromosome 1q, loss of 5q 10q, and *MYC* amplification. Both Group 3 and 4 tumours have a high incidence of isochromosome 17q, although this is higher in group 4 (66%)

than Group 3 (26%) (Jones *et al.*, 2012), while Group 4 shows frequent loss of the X chromosome in affected females (Taylor *et al.*, 2012). Although these subgroups show similar expression patterns to each other for most genes, which makes them difficult to distinguish, they both have some distinctive genetic expression patterns. Specifically, high expression of genes involved in the cell cycle and the p38 mitogen-activated protein kinase (*MMAPK*) pathway in Group 3, and high expression of genes involved in neuronal differentiation/development and cytoskeleton organisation in Group 4 (Kool *et al.*, 2008;Northcott *et al.*, 2011).

1.4 Signalling pathways involved in MB

Through molecular analyses, including those outlined above, several known signalling pathways have been implicated in the development of MB, with the pathways known to have the most critical role being the Shh and Wnt pathways which define two of the MB subgroups (Northcott *et al.*, 2011).

1.4.1 The SHH/PTCH signalling pathway

Signalling by the hedgehog (Hh) family of secreted glycoproteins is implicated in the determination of embryonic cell proliferation and migration, specification of organ size, and patterning of several tissues including skin, lung, brain, bone, and blood. The Hedgehog family of proteins includes Sonic Hedgehog (SHH) {7q36}, Desert Hedgehog (DHH) {12q13.1}, and Indian Hedgehog (IHH) {2q33–q35} (Pola *et al.*, 2003). The SHH pathway plays a critical role in histogenesis, and dysfunctions in the pathway have been associated with several human syndromes and neoplasms. Indeed, the elucidation of the molecular pathways involved in medulloblastoma initiation and promotion came from the association between SHH with Gorlin syndrome (Dhall, 2009). SHH is known to stimulate the proliferation of cerebellar granule neuron prescursor cells (GNPCs) by activating the Gli transcription factors (*GLI1*, *GLI2*), and inhibition of SHH signalling *in vivo* leads to a reduced number of granule neurons (Hui *et al.*, 1994). Mutations (resulting in pathological activation) in a subset of SHH-responsive cells result in unrestrained cell division and, eventually, medulloblastoma. Initial mutations were identified in the *PTCH1* gene (Pazzaglia, 2006). *PTCH1* is a trans-membrane protein that acts as a receptor for SHH. In the absence of a signal from *SHH, PTCH1* switches off the pathway by binding to Smoothened (*SMO*). Upon SHH binding to *PTCH1*, the pathway is switched on and the signal is transduced to *SMO*, then into the nucleus where it activates Gli transcription factors (Rossi *et al.*, 2008).

Loss-of-function mutations in *PTCH1* and Suppressor-of-Fused (*SUFU*) (a downstream molecule in SHH signalling cascade) occur in medulloblastomas (Zurawel *et al.*, 2000), and according to (Pazzaglia, 2006) 25% of sporadic human medulloblastomas have *PTCH1*, *SMO* or *SUFU* mutation.

Based on the above findings, researchers have argued that tumours associated with excessive hedgehog signalling might result from dysregulated self-renewal properties of cancer stem cells (CSC) (Sutter *et al.*, 2010); however, many questions remain regarding the repertoire of genes activated by the SHH pathway and the extent of their overlap in different subtypes of MB.

1.4.2 Wnt/Wingless (WNT) signalling pathways in Medulloblastoma

The association between medulloblastoma and the WNT pathway was initially suggested by patients with Turcot syndrome. These patients have germline mutations in the adenomatous polyposis coli (*APC*) gene and have a 92-fold higher risk of developing medulloblastomas and other neuroepithelial tumours relative to the general population (Hamilton *et al.*, 1995). Like SHH, the WNT pathway also plays a key role in embryogenesis. In this pathway, Wnt ligand associates with its receptor Frizzled (*FZD*), which in turn phosphorylates Dishevelled. Dishevelled then triggers the release of β-catenin from an inhibitory complex comprising (*APC*) and the Axin proteins (*AXIN1*, *AXIN2*). Subsequent translocation of β-catenin into the cell nucleus is presumed to mediate tumourigenesis via the aberrant activation of oncogenes such as cyclin D1 (*CCND1*), *MYC*, *AXIN2* and RE1-silencing transcription factor (*REST*) (Huse and Holland, 2010). The association between Wnt pathway mutations and medulloblastoma tumourigenesis has been confirmed by a study which found that approximately 20% of patients with medulloblastomas have mutations in *APC*, *AXIN1*, *AXIN2* or *CTNNB1* (which encodes *β-catenin*) (Huse and Holland, 2010).

1.4.3 Other signalling pathways involved in MB

In addition to the pathways which define specific subgroups, alterations in a number of other known signalling pathways have been implicated in MB. For instance, Fan *et al.*, (2006), has suggested that NOTCH signalling is involved. The Notch signaling pathway is initiated by ligand binding to transmembranous receptors (*Notch 1 – 4*) followed by an intracellular signal transduction cascade initiated by the g-secretase complex at the intracellular portion of Notch. Cleavage of Notch results in the release of the intracellular domain of Notch (ICN) into the cytoplasm and its subsequent

translocation into the nucleus where it promotes the transcription of several target genes (Hairy/Enhancer of split (Hes 1-7); *cyclin D1* and apoptosis related genes, among others). Overexpression of *Notch2* and *Hes1* has been shown to promote medulloblastoma tumorigenesis in mice, while *Notch1* inhibits the process. Furthermore, *NOTCH2* over-expression has been demonstrated in human MB (Fan *et al.*, 2004;Xu *et al.*, 2009). However, precisely how NOTCH signaling affects tumorigenesis has yet to be elucidated (Hatton *et al.*, 2010).

The epidermal growth factor receptor (*EGFR*) family of receptor tyrosine kinases (including *ERBB1- ERBB4*) have also been implicated in MB. Approximately 80% of all medulloblastomas have been shown to overexpress *ERBB2,* while overexpression of *ERBB4* is observed in tumours showing *GLI1* repression (de Bont *et al.*, 2008). The over expression of *ERBB2* and *ERBB4* has also been linked to poor prognosis (Kim *et al.*, 2010). EGFR signalling can lead to sequential phosphorylation of several proteins downstream, including proteins within the Ras/Raf pathway, signal transducers and activators of transcription (*STAT*) pathway, the Akt pathway and mitogen-activated protein kinase (*MAPK*) pathway, which collectively affect cellular functions such as differentiation, proliferation, apoptosis, and survival (Huse and Holland, 2010).

Finally, Bone morphogenetic proteins (BMPs), soluble effectors of differentiation belonging to the Transforming growth factor (*TGF-b*) family of cytokines (Fogarty *et al.*, 2007), have been implicated in MB development by (Rios *et al.*, 2004). They demonstrated that *BMP2* antagonizes SHH dependent proliferation by expressing *TIEG1* which downregulates *MYCN* transcription. Subsequent gene profiling by (Zhao *et al.*, 2008), reported that several genes within the BMP signaling pathway are repressed in medulloblastomas with constitutive SHH pathway activation. The BMP signaling pathway has been implicated in cell cycle exit, migration and differentiation of GNPCs during embryogenesis (Roussel and Hatten, 2011).

1.4.4 Epigenetic and MicroRNA silencing in medulloblastoma

While the role of loss of function mutations in the tumorigenesis of medulloblastoma is well defined, recent evidence indicates that both epigenetic and microRNA silencing also play significant roles in tumorigenesis (Shahi *et al.*, 2011), and in some instances may affect genes involved in the pathways outlined above.

Epigenetic changes linked to cancer include histone modification via acetylation, methylation and phosphorylation, and hypermethylation of CpG motifs in

DNA promoter regions (Vibhakar *et al.*, 2007). Several oncogenes have been shown to be epigenetically silenced in medulloblastomas. These include: SFRPs genes, the S100 gene family, Kruppel-like factor 4 (*KLF4*), *HIC1* and *RB1* (de Bont *et al.*, 2008;Roussel and Hatten, 2011). Furthermore, a recent study by Nakahara *et al.*, (2010), concluded that *KLF4* is repressed either by a genetic or epigenetic mechanisms in medulloblastomas and that it may function as a tumor suppressor. Using whole-genome microarray analysis in combination with trichostatin A (TSA), a histone deacetylase inhibitor (HDAC), Vibhakar *et al.*, 2007, demonstrated that several genes including Dickkopf -1 (*DKK1*), a Wnt antagonists, are upregulated on HDAC inhibition.

MicroRNAs are 18-22 nucleotide long RNA sequences which are transcribed from noncoding parts of the genome and have post transcriptional gene silencing activity (Venkataraman *et al.*, 2010b). Generally, miRNAs repress translation of mRNA by base pairing with the 3'-untranslated region (3' UTR). In some instances, they can mediate mRNA degradation. While the exact role of miRNA is only beginning to emerge, some reports have implicated them in tumor formation and inhibition (Kefas *et al.*, 2009). Overexpression of miRNA 21 is associated with glioblastoma while miRNA 128a is strongly down regulated in medulloblastoma (Venkataraman *et al.*, 2010a). Consistent with these findings, Northcott *et al.*, (2009), demonstrated that miRNAs in the 17/92 cluster function as oncogenes in MBs with an SHH signature. In another study, it has been demonstrated that neuronal miRNA-326 acts in a feedback loop with *NOTCH1* and has prospects as a therapy against neurologic tumors (Kefas *et al.*, 2009).

1.5 High throughput MB cancer gene discovery.

Microarray analyses, and particularly their integration with clinical information, cytogenetics, and details of tumour pathology, have provided insight into MB subgroup structure and their potential importance in prognosis and treatment. However, they have also highlighted the need to dissect tumorigenesis further, to identify additional genes involved in the disease progression in different subgroups, and identify new therapeutics tagets. Critically, while microarray analyses have identified hundreds or thousands of genes which show subgroup specific expression patterns, establishing which are critical to the disease process, as opposed to incidental or downstream correlates, is not straightforward. The development of next generation sequencing technologies has recently enabled an additional layer of information to be cross referenced to such analyses.

The first large scale MB sequence analysis used high-density microarray analyses and sequencing of almost all coding exons of known genes and miRNAs to analyse a cohort of 22 human MB (Parsons *et al.*, 2011). In total, 96% of target amplicons and 95% of target bases were successfully amplified and sequenced. 225 somatic mutations were identified, 199 of them point mutations, with 183 being non-silent mutation (Parsons *et al.*, 2011). Importantly, in addition to the detection of alterations in the SHH and WNT pathways, inactivating mutations of the histone lysine N-methytransferase genes (*MLL2 or MLL3*) were identified in approximately 16% of analysed MB (Parsons *et al.*, 2011).

Although comprehensive, studies such are this are not capable of identifying all mutagenic events. For instance, noncoding mutations, copy number changes, and epigenetic alterations will not be identified. It is noteworthy that the authors of this exome sequence analysis pointed out that the initiating events in over 50% of MBs remained to be identified (Parsons *et al.*, 2011).

1.6. Transposons and their application to cancer gene discovery

There is clearly a need for complementary, experimental, cancer gene identification methods (Parsons *et al.*, 2011), both to identify cancer genes not qualitatively altered during cancer development, and to provide confirmatory evidence of the relevance of genes already implicated in disease by high throughput quantitative analyses such as those described above. The primary aim of the work described in this thesis is to use one such technique, transposon mutagenesis screening, to identify genes involved in medulloblastoma. Such screens are performed in murine models of disease, and comparative analyses of human and mouse datasets can then be employed to identify key genes involved in disease in both species for further investigation, and to shed light on the underlying biology of tumour development.

1.6.1 Transposons

Transposons are mobile segments of DNA which are present in virtually all genomes. In man, they constitute close to 45% of the entire human genome (Feschotte *et al.*, 2002). These mobile segments of DNA use mechanisms that are distinct from recombination to insert themselves at different locations within the host DNA. These segments of DNA can travel not just within a particular genome but also between species. Interspecies transfer of transposons is referred to as horizontal gene transfer (de la Cruz and Davies, 2000).

Structurally, transposons exhibit great variability, with some containing only those sequences which are essential to their own transposition. Other transposons have, in addition, genes which code for many other functions that have no bearing whatsoever on their movement. However, they all have certain features that are common (Grabher and Wittbrodt, 2007): First, most transposable elements are flanked by direct repeats. These repeats are usually 3 to 12 bp long and do not move with the elements as they are not part of the transposon. These repeats are generated during transposition as a result of DNA repair at the insertion points, and the repeat sequences differ depending upon the insertion site. However, each type of transposable element has an unchanging flanking repeat length. These flanking direct repeats are indicative of staggered cuts which are made when insertion takes place. This process is depicted in Figure 1.5. Once the staggered cuts are made, short single stranded DNA sequences are generated on either side of the transposon, and flanking direct repeats created once replication of these sequences takes place ((Zayed *et al.*, 2004).

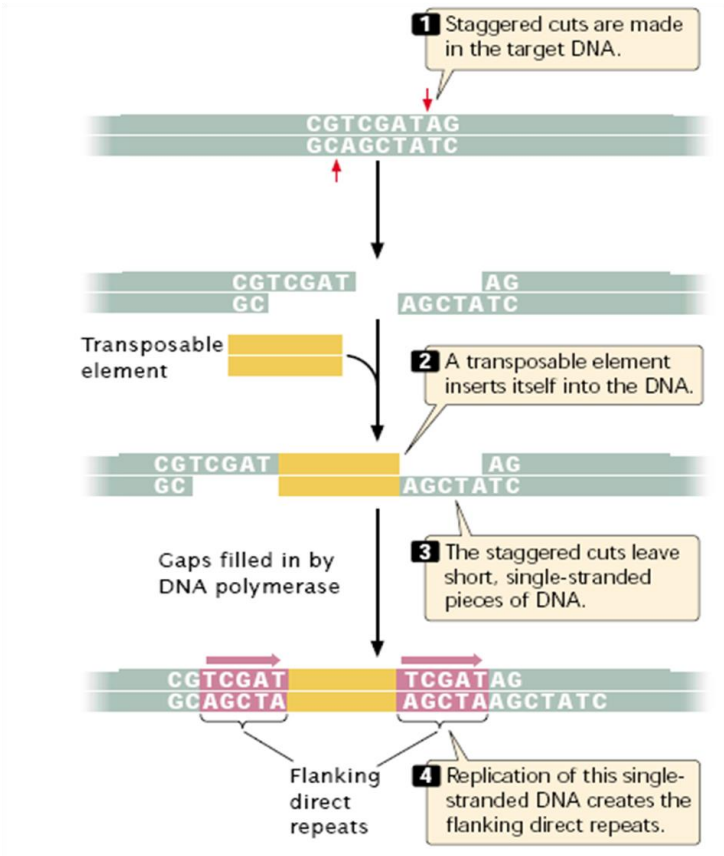

Figure 1.5. Generation of flanking direct repeats during the insertion of a transposable element into DNA. Firstly the DNA will be open via a struggered cut. Then the insertion of the transposable element will take place. This will leave a small footprint sequences in the DNA. The last step is the replication of flanking direct repeats. (Adapted from Pierce 2008).

Second, most transposons have terminal inverted repeats. These are sequences which are seen at the ends of the transposons and which range in size from 9 to 40bp. The terminal inverted repeats are important in the movement of the transposons. Figure 1.6 depicts the features that are commonly found in different transposable elements (Grabher and Wittbrodt, 2007).

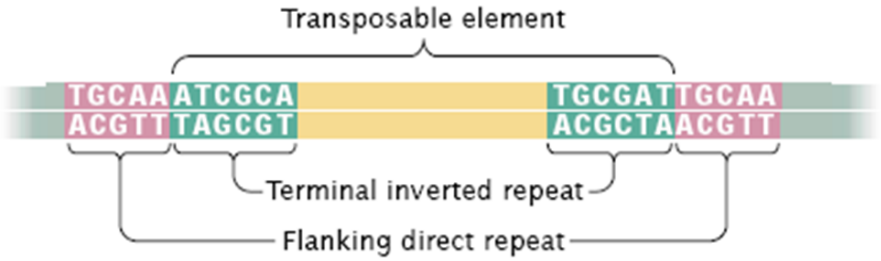

Figure 1.6. Features common to most transposable elements. Many transposable elements create flanking direct repeats upon insertion, and many contain terminal inverted repeats (Adapted from (Pierce, 2008).

Eukaryotic transposons can be split into two subdivisions: The first group replicates using DNA intermediates; the second group consists of retrotransposons which replicate via an RNA intermediate. They do not code for a transposase enzyme and typically replicate like retroviruses (Kalendar *et al.*, 1999).

1.6.2. Transposition

Transposition is the transfer of a short segment of DNA to another position on the DNA strand. The steps involved in transposition are common to eukaryotic and prokaryotic elements. First, staggered cuts are created in the DNA where the transposable element will be inserted. Next, the transposon attaches to the single-stranded ends of the target DNA. Finally, replication of the DNA at the single stranded openings occurs. Transposition may be carried out via DNA or RNA intermediates; when DNA intermediates are used, either replicative or non-replicative transposition can occur (Izsvak *et al.*, 2009).

Replicative transposition involves an increase in the number of transposons. A new copy of the transposon is inserted at a new location even as the old one is left at the original location. This method is also called copy and paste transposition (Kazazian, 2004). This mode of replication can take place between 2 distinct DNA molecules or between 2 segments within the same DNA molecule. Unlike replicative transposition, non-replicative transposition does not involve the replication of the complete transposable element. Rather, the element moves from one location to another location.

This involves only the cleavage of the target DNA and the insertion and joining of the transposon in the opening that has been created by the cleavage. The cleavage is carried out by a transposase which is encoded by the transposon (Tavakoli and Derbyshire, 2001) .

Retrotransposons are transposons which use RNA intermediates. The use of RNA intermediates requires an additional step of conversion to double-stranded DNA with the use of the enzyme reverse transcriptase (Kalendar *et al.*, 1999).

1.6.3 Transposition and Cancer

There is a strong association between transposition and mutagenesis. This is because transposons can interrupt genetic function when inserted into genes. Studies shown that more than 50% of all mutations in the fruit fly *Drosophila melanogaster* that are caused by spontaneous mechanisms are a direct consequence of transposition of elements into, or adjacent to, a functional gene (Chung *et al.*, 2008b). The mutations induced by transposition are usually harmful but some have also been noted to be of considerable benefit to the affected organism. For instance, some transposons carry genes that code for antibiotic resistance in some bacteria (de la Cruz and Davies, 2000). Neurofibromatosis is an example of a human cancer which can occasionally result from knockout of one copy of the *NF1* gene as a result of transposition. This has been observed to be caused by an Alu sequence inserting into an intron of the *NF1* gene (Wallace *et al.*, 1991). Insertion of the Alu sequence into this gene induces an error in RNA splicing and the net result is that one of the exons of the gene is omitted in the resultant mature mRNA. The omission of this exon causes the reading frame to shift. As a result, translation of the mRNA yields abnormal proteins that precipitate the disease.

Rearrangements can also occur due to the transfer of genetic material and recombination during transposition, and these can induce mutations that are carcinogenic. Inversions, deletions, and duplications can result due to homologous recombination (Ivics *et al.*, 1997;Izsvak *et al.*, 1997;Izsvak *et al.*, 2009). Non-replicative transposition can also lead to rearrangements which can then cause cancers. This can happen if repair of the broken DNA is not carried out effectively. If the repair fails to take place, a deletion will occur since the acentric fragment will get lost. Other diseases which have been linked to transposition events include muscular dystrophy and haemophilia (Li *et al.*, 2001;Ostertag *et al.*, 2007).

1.7 Transposons as a cancer gene discovery tool

A wide variety of techniques and strategies have been developed to identify genes responsible for the promotion of cancerous cells. In addition to genetic mapping and mutation detection techniques that can identify germline and somatic mutational events within human tumours (Mattison *et al.*, 2009), a variety of mutagens have been used in forward genetic screening to induce mutations within model organisms (Starr and Largaespada, 2005). Until relatively recently, matagenesis screens were usually performed using radiation or chemicals, and have been associated with a high mutation frequency (Starr and Largaespada, 2005). However, the task of identifying the genes responsible for any resultant phenotype was usually time consuming and challenging, due to the fact that large-scale mutations affecting multiple genes can be induced, and because mutations induced by these methods need to be mapped genetically (Collier and Largaespada, 2007b;Grabher and Wittbrodt, 2007).

Retroviruses and transposon mutagenesis have become popular alternatives to the above techniques to identify the genes that predispose an organism to cancer, and make identification of the mutagenic events more straightforward as the mutations have the molecular tag of insertion. However, retroviruses are active in a limited number of cell types, and these are invariably actively dividing cells (Pajer *et al.*, 2006). Recently, several transposon systems have been developed which circumvent this limitation. One system, known as Sleeping Beauty (SB), has the ability to perform in a wide variety of somatic cells of the target organism, thereby allowing for a wide range of cancer types to be analysed (Collier *et al.*, 2005). SB is a deployed system, consisting of a defective transposon (T2/Onc) or its derivatives (T2/Onc2 and T2/Onc3), and the active enzyme, the transposase (SB10 or SB11) (Collier *et al.*, 2005;Dupuy *et al.*, 2009). This system is based on a "cut-and-paste" mechanism, in which the transposase excises T2/Onc from donor DNA and inserts it into the target DNA (Ivics *et al.*, 1997). The inserted transposons can cause pseudo-random insertional mutations. Because the T2/Onc transposon has been engineered to contain splice acceptor (SA) and splice donor (SD) signlas, poly-adenylation (PA) signals, and an enhancer (from the murine stem cell virus, MSCV), it can induce both loss and gain of function mutations (Collier *et al.*, 2005).

The SB system of murine cancer mutagenesis has become a useful tool for identification of cancer genes since the proof of principle was published in 2005. This landmark work (Collier *et al.*, 2005), used SB to increase the frequency of sarcomas generated in *Arf*-deficient mice. A series of 45 common insertion sites were identified

and mutations in *Braf* were detected in 22 tumours. SB mutagenesis has since been used successfully as a tool to identify candidate cancer genes in adult tumours, such as *Zmiz1* in squamous-cell carcinoma and *Rian* in hepatocellular carcinoma, *Smad4*, *Pten* and *ApcI* in colorectal cancer, *Pde4d* in prostate cancer, and *Csf1* in high-grade astrocytoma (Dupuy *et al.*, 2009;Rahrmann *et al.*, 2009;Starr *et al.*, 2009;Bender *et al.*, 2010). In addition, several SB studies have generated MBs at low frequency in both tumour models and wild-type mice (Dupuy *et al.*, 2005;Su *et al.*, 2008;Collier *et al.*, 2009). This suggests that this system is capable of causing mutations within one or more MB cell of origin, and suggests that it might prove effective if applied to a MB specific cancer model.

1.8 MB murine models available for transposon screens.

The aim of the work described here is to identify cancer genes relevant to MB. Screens which analyse a specific cancer type have mutagenised existing cancer models, aiming to identify genes which interact with the cancer predisposition (when mutated by the SB transposon) by altering tumour penetrance or latency. The choice of cancer model used in the mutation screen is therefore likely to have a major impact upon the spectrum and relevance of the genes identified. A model which clearly recapitulates the human disease, and where the predisposition is understood, would be preferred to a non-specific model.

At the outset of this work there were several MB mouse models which could be used for mutagenesis studies, all modelling the SHH subgroup of tumours and each created by targeting components of the Shh pathway. One is a *Sufu+/-/Tp53-/-* model (Lee *et al.*, 2007) where a gene trap approach was applied to obtain the *Sufu* mutation. The *Sufu* null mice are embryonic lethal, with arrest at embryonic day 10 (E10). The median survival of *Sufu+/-;p53-/-* mice is 4 months with a tumour penetrance of 58%. The *Sufu+/-;p53-/-* tumours express genes which are hallmarks of Shh signalling pathway activation such as *N-myc, Sfrp1, Ptch2,* and *cyclinD1*. These results indicate that this model could be used for mutagenesis, although *Tp53* mutations are not a hallmark of this tumour type.

Another model was created by expressing a constitutively active form of Smoothened (*SmoA1* and *SmoA2*) – a down stream target gene of Shh signalling – under the control of *NeuroD2* (ND2) (Hallahan *et al.*, 2004). This ND2;SmoA1 line developed MB in about 48% of animals with a median age of 25.7 weeks. The tumours showed significant proliferation of GNPCs and the expression of *Gli1* and *N-myc* were

also observed at high levels, indicating the stimulation of the Shh pathway. Notch signalling was also investigated in these tumours via down stream targets (*Notch2* and *Hes5*) and their expression was enhanced, confirming the efficient induction of the Notch pathway by Shh pathway activity (Hallahan *et al.*, 2004). The authors concluded that both the Shh and Notch pathways contribute to the development of MB tumours in this mouse model and in humans as well. The same group of researchers have more recently generated a further model with homozygous knock out of Smoothened (Smo/Smo), (Hatton *et al.*, 2008). With this double mutation, MB incidence was significantly elevated up to 94% within 4-8 weeks. It is noteworthy that tumours in this model present with leptomeningial metastases. The elevated tumour incidence, early mode of onset, and metastases make it useful for analysis of the SHH subgroup of MB.

However, by far the most extensively studied model is the *Ptch1+/-* mouse model (Goodrich *et al.*, 1997), and this will be described in the following section.

1.8.1 Ptch1 Murine Model of Medulloblastoma

The *PTCH1* gene is the human homolog of the Drosophila patched gene and is localized on chromosome 9q22.3 (Gailani *et al.*, 1992;Johnson *et al.*, 1996). This gene codes for the Patched 1 (*PTCH1*) receptor which plays an important role in embryogenesis. This receptor is an attachment site for the Sonic hedgehog-signalling molecule (Shh) that is vital for the development of several tissues and organs. The activation of the signalling pathway occurs when Shh binds to the Ptch1 receptor (Figure 1.7). In the absence of Shh, another protein called Smoothened or Smo is inhibited. This inhibition is removed when Shh attaches to Ptch1, causing the Gli transcription factors to stimulate the expression of downstream target genes which regulate the differentiation and proliferation of cells (Pazzaglia, 2006). However, artificial mutations of *Ptch1*, *SMO*, and *SuFu* have been shown to stimulate the Shh pathway continuously, leading to major disruption in cellular microenvironmental conditions within the affected organs, and this may leads to tumourigenesis (Goodrich *et al.*, 1997; Hallahan *et al.*, 2004; Lee *et al.*, 2007)

Mutations of the *PTCH1* gene have been associated with the basal cell nevus (BCN) syndrome. Studies also show that neoplasias such as MB and the basal cell carcinomas (BCC) can harbour mutations on this gene. These are autosomal dominant conditions (Dhall, 2009). Several other studies have linked prostate cancer, breast cancer, human small-cell lung cancers, human pancreatic carcinomas and digestive tract tumors to dysregulation of the Shh pathway and consequently to mutation of the

PTCH1 gene (Lewis *et al.*, 2001;Berman *et al.*, 2003;Thayer *et al.*, 2003;Karhadkar *et al.*, 2004;Kubo *et al.*, 2004;Sanchez *et al.*, 2004;Sheng *et al.*, 2004).

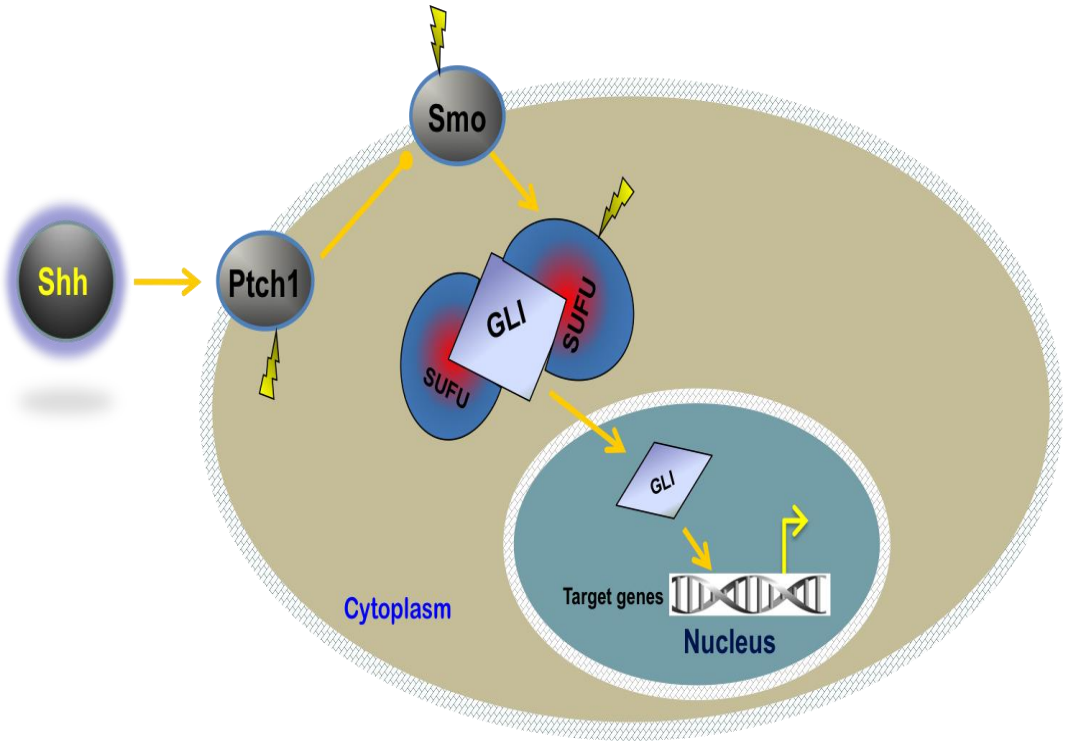

Figure1.7: Summary of the sonic Hedgehog (Shh) pathway: When Shh ligand is absent the pathway is inactivated via the inhibition of Smoothen (Smo) by Ptch1, where both are located on the cell wall. Activation of the signalling pathway occurs when Shh binds to the Ptch1 receptor, causing a protein called suppressor of Fused (Sufu) to release the Gli transcription factors into the nucleus. Consequently, Gli factors stimulate the expression of downstream target genes which regulate the differentiation and proliferation of cells. Mutations in Ptch1, Smo or Sufu, can make Shh pathway continuously active, which may leads to tumourgenesis.

In the *Ptch1* +/- model developed by Goodrich *et al.*, 1997, part of exon 1 and all of exon 2 of *Ptch1*, the murine ortholog of the human *PTCH1* gene, were replaced with *LacZ*. Homozygous *Ptch1-/-* mice died during embryogenesis, and this was concluded to be due to heart malformation. Medulloblastoma frequency in heterozygous mice was originally recorded as 14%, with MB development reaching the peak between 12 – 25 weeks, although the genetic background does affect this (Wu *et al.*, 2012). Histologically, the tumours generated by this model have similar features to those from human medulloblastomas, being characterised by small undifferentiated tumour cells consisting of dark carrot-like nuclei and small cytoplasmic volume.

The *Pcth*$^{+/-}$ model has since been crossed to *Tp53*$^{+/-}$ mice to investigate the incidence of MB formation (Wetmore *et al.*, 2001). This increased the incidence of MB

to more that 95% and reduced the latency to 10-12 weeks postnatally. On the other hand, Tp53-/- mice develop MB like tumours in ~5% of experimental animals. To investigate the role of the retinoblastoma (Rb) pathway in MB incidence, the model was also placed in an *Ink4c*-/- background (*Ink4c* is a downstream target of *Rb*). The incidence of MB was appreciably increased in *Ptch*+/-;*Ink4c*+/- or *Ink4c*-/- mice from 7%-30% with a very short latency. However, the *Ink4c*-/- mice did not show tumours production (Uziel *et al.*, 2005).

Because of the extensive characterisation of this model to date, the strong similarities between the tumours it generates and human tumours, and the relatively low penetrance of the original model (potentially allowing any interaction with a mutagen to be easily observed), it was chosen for the mutagenesis study reported here.

1.9 Cerebellar structure and development

Medulloblastoma, like many childhood tumours, is considered a disease of development. As a result, the impact of aberrant SHH signalling upon tumour development can only be fully understood within the context of its role in normal cerebellar structure. The word cerebellum was originally coined from latin, and means small brain. The cerebellum is located at the base of the skull, above the pons and brain stem, and below the cerebral cortex. It consists of a massive number of neurons, principally granule cells. The cerebellum has several functions, the most important and well characterised relating to fine motor control (processing neuronal outputs of activity to coordinate muscular movement), cognitive functions (including learning), balance, and muscle tone (Martin, 2003).

The adult cerebellum is composed of two structures; the cortex (outer layer) and the medulla. The cerebellar cortex itself consists of 3 layers. The first is the outer molecular layer, which contains 2 types of neurons, known as the stellate and basket cells, in addition to specialised structures called parallel fibres. These fibres are present in large numbers and are the extended axons of granule cells, the most abundant neuron type in the cerebellum. The next layer, the intermediate/ganglionic or Purkinje layer, is composed of purkinje cells, neurons which can be stimulated by the parallel fibres present in the molecular layer. Their axons have the ability to stimulate the deep cerebellar nuclei. The last layer, the cortex or inner granular layer, consists primarily of the cell bodies of granule cells and Golgi cells. Five types of neurons are, however, found in this layer and most are considered inhibitory, as they secrete the inhibitory neurotransmitter GABA. In addition to these structures, 2 types of sensory fibers are found in the cerebellum; climbing fibers (derived from the inferior olivary nucleus) which synapse with purkinje cells, and mossy fibers. The basic structure of all 3 layers is shown in Figure 1.7 (Voogd and Glickstein, 1998). In humans, granule neuron precursor cells (GNPCs) account for ~40% of all neurons, and approximately 45 billion out 110 billion neurons in the brain are GNPCs.

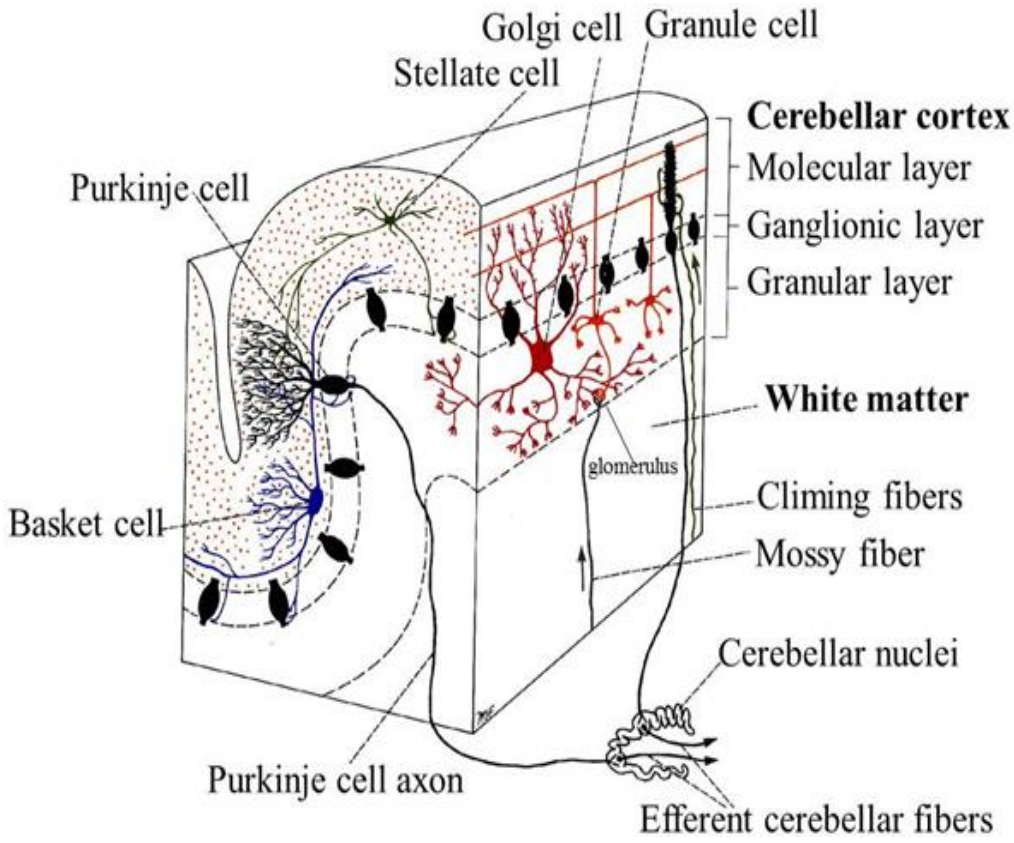

Figure 1.8: Cerebellar anatomic structure: The two main cerebellar layers, the cerebellar cortex and the white matter are shown. The cerebellar cortex consists of molecular, ganglionic (Purkinje) and granular layers. Stellate and basket cell are within the molecular layer. Purkinje cells are within the Purkinje layer. Granule and Golgi cells are within the granule layer. Climbing fibers and mossy fibers are shown. Purkinje cells axon projected towards cerebellar nuclei are shown. (Adapted from University of Tartu, http://sisu.ut.ee/histology/cerebellum)

At early stages of embryonic development (~E10.5 in mouse) brain tissue is organised into three different segments known as the prosencephalon (forebrain), mesencephalon (midbrain) and rhombencephalon (hindbrain) (Figure 1.8A). The cerebellum arises from the most caudal of these, the rhombencephalon. The rhombencephalon is composed of 8 loops, known as rhombomeres, and the cerebellum develops from rhombomere 1, the most proximal of these. The neurons of the cerebellum develop from two main germinal zones - the ventricular zone and the rhombic lip, a structure which skirts the roofplate of the 4th ventricle of the brain (Figure 1.8B). The cerebellar cortex and cerebellar nuclei are generated by a complex pattern of cell movement and signalling (Fink *et al.*, 2006). In this process, GABAergic neurons of the cerebellar cortex (the Purkinje cells and neurons of the cerebellar nuclei, among others) arise in the ventricular zone (VZ) and migrate to the core of the developing cerebellum (Lin and Cepko, 1999;Morales and Hatten, 2006). After this, at

approximately E12, the granule neuron precursor cells (GNPCs) start to proliferate and migrate from the secondary germinal zone, the rhombic lip, forming the external granular layer (EGL). These then exit the cell cycle, begin to differentiate and migrate internally to form the internal granular layer below the Purkinje cells with which they have synapsed. Loss of the EGL is complete after approximately 2 weeks of postnatal development in the mouse to leave the three layers of the adult cerebellar cortex.

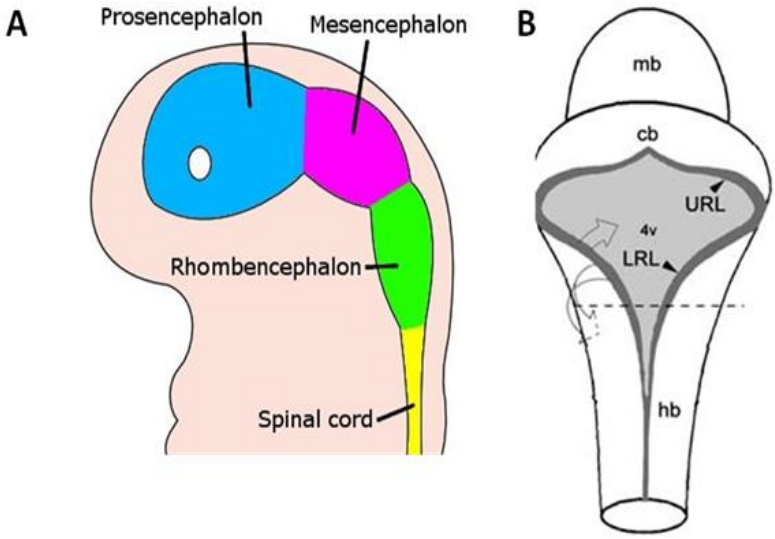

Figure 1.9: Schematic showing different parts of the embryonic developing brain, and position of the rhombic Lip relative to hindbrain. (A) The three different segments of the developing brain. (B) Dorsal appearance of the neural tube and the neuroepithelium of the rhombic lip (dark gray) and the roof plate (light grey). The empty grey arrows show the divergent movement behavior of the LRL. Figures adapted from (Landsberg *et al.*, 2005) and Early brain development (http://mdc06kb.webnode.com).

1.10 Medulloblastoma cells of origin

Although the precise origins of the various subtypes of medulloblastoma remain to be elucidated, recent evidence point to the existence of cancer stem cells (CSC) or cancer propagating cells (CPCs) which maintain the neoplastic cells (Uchida *et al.*, 2000;Reya *et al.*, 2001). These cells are thought to be similar to normal stem cells which are involved in the development of cerebellar tissue; however it is presumed that their replication and differentiation pattern is aberrant (Gilbertson and Ellison, 2008). Currently, different models, including the stochastic model, queen model, hierarchical model and plasticity model, have been proposed to account for the clonal evolution of brain cancer propagating cells (BCPCs) from CSC (Hadjipanayis and Van Meir, 2009). In medulloblastoma, the first BCPC candidate was isolated by (Singh *et al.*, 2004). The cells were identified by virtue of expressing prominin 1 (CD 133) – a marker of normal neural stem cells. Subsequent studies demonstrated that between 1-21% of freshly isolated medulloblastoma cells express CD 133 (Singh *et al.*, 2004;Calabrese *et al.*, 2007). Other markers which have been linked to BCPCs include *Musashi1*, an RNA-binding proteins (Kaneko *et al.*, 2000); the oncogene *Bmi1* (Leung *et al.*, 2004;Bao *et al.*, 2006); the transcription factor *Sox2* (Bao *et al.*, 2006) identified in Glioma stem cells; the cell cycle regulator *MELK* (Nakano *et al.*, 2008); the signaling pathway protein *NOTCH* (Fan *et al.*, 2006) and A2B5, a ganglioside antigen (Read *et al.*, 2009). However, the existence of BCPCs is still controversial (Rich and Eyler, 2008).

Related to the controversial notion of the existence of CPCs, is the suggestion that the large population of proliferating GNPCs includes subpopulations of pre-cursors with distinct genetic profiles which may give rise to one or more of the different subtypes of medulloblastoma (Gong *et al.*, 2003;Hatten and Roussel, 2011). Support for this hypothesis has come from a large-scale CNS gene expression atlas (GENSAT Project), which established that some genes implicated in MB are differentially expressed in particular subsets of GNPCs at various developmental stages (Gong *et al.*, 2003). Indeed, research has shown that cells isolated from MB samples express GNPCs markers such as *p75NTR*, *TrkC*, *Zic1* and *Math1* (Pomeroy *et al.*, 1997;Salsano *et al.*, 2004).

Recently, (Yang *et al.*, 2008) has provided evidence that medulloblastomas originate from GNPCs. Furthermore, they also implicate stem cells in medulloblastoma generation. They conditionally knocked out *Ptch1* through *Math1*-Cre to create a *Math1*-Cre/*Ptch1*$^{C/C}$ mouse. These mice should generate *Ptch1* knockout in GNPCs only. This was confirmed by the expression of a GFP reporter in the cerebellum only of

P8 neonatal mice. The deletion of *Ptch1* was confirmed by detection of Shh signalling activation and by increased expression of *Gli1, Cyclin D1* and *N-Myc* as assayed by qPCR, and increased proliferative activity compare to wild-type cells. The authors also showed that the E.14.5 *Math1*-CreER/*Ptch1*$^{C/C}$ mice (a stage when *Math1* is only expressed in GNPCs) when treated with tamoxifen - a substance that activates estrogen receptor (ER) - developed tumours at a frequency of 100%. Furthermore, postnatal examination of tamoxifen treated mice showed the development of cerebellar tumours by the age of 19 weeks giving further evidence that GNPCs are the cells of origin of MB. The effect of *Ptch1* deletion upon neuronal stem cells (NSCs) was investigated in terms of tumours formation. The authors used Cre controlled by human glial fibrillary acidic protein (*hGFAP-cre*) mice to delete *Ptch1* in NSCs. Tumours formation was observed in the *hGFAP*-cre/*Ptch1*$^{C/C}$ mice during development and by 4 weeks of age these mice showed symptoms of neurological disease and tumour formation in their cerebella. Thus, while the precise cellular origin of MB subgroups remains to be established, in terms of whether these are true cancer stem cells or a pluripotent progenitor cell (Sutter *et al.*, 2010), current evidence suggests that lineage commitment is required for MB development, and that the cells of origin of the SHH MB subgroup are GNPCs.

1.11 The role of SHH in Cerebellar Development

The role of the Shh signalling pathway as a mitogenic factor upon normal cerebellar GNPCs isolated from P4-7 mice has been examined. The proliferation rate was significantly higher within Shh+/+ GNPCs compared to Shh-/- GNPCs. This is consistent with the known proliferative role of Shh pathway upon GNPCs. Culturing purified GNPCs without recombinant Shh protein showed modest cellular proliferation. However, upon the addition of recombinant Shh protein t the proliferation rate increased to ~2.5X on average (Dahmane and Ruiz i Altaba, 1999).

These results are supported by an independent study where GNPCs treated with recombinant Shh protein were analysed by flowcytometric analysis. The Shh GNPCs were abundantly distributed in G_0/G_1 phase, and ~16% were in S phase, compared to 2.5% cells without Shh, being in S phase (after 36 hours experimental life). Interestingly, about 11.5% of all tested GNPCs were in S phase at 0 time (Kenney and Rowitch, 2000). The ectopic induction of the Shh signalling pathway was assessed by analysis of *Gli1* expression. *Gli1* is an Shh downstream target gene, the primary effector of the signalling pathway, and was utilised to give a direct sign of Shh pathway

induction. The *Gli1* expression was higher in Shh+/+ GNPCs compared to Shh-/- GNPCs. This confirms that Shh pathway induction was obtained and that it has a functional effect on GNPCs in culture.

These analyses are also in agreement with a previous study where GNPCs were treated with recombinant Shh protein and showed higher expression of *Gli1* (Wechsler-Reya and Scott, 1999). Importantly, the Shh pathway is known to maintain the treated cells in an undifferentiated state. *Math1* expression was used as an indicator for undifferentiated neuronal cells (Wechsler-Reya and Scott, 1999). *Math1* is a TF that is primarily expressed in the developing mammalian centeral nervous system (Helms and Johnson, 1998). The expression of this gene was found predominantly in the dorsal region of the neural tube near to the roof plate (Alder *et al.*, 1996;Zhang and Goldman, 1996). Helms and Johnson, (1998), found that *Math1* was exclusively expressed within proliferating GNPCs, as $Math1^{+}$ cells were localised only withing the proliferating zone of the external germinal layer (EGL). The observation that *Math1* expression was higher in Shh+/+GNPCs, relative to untreated, is indicative of maintenance of an undifferentiated state.

Because of the early link between SHH signalling and Gorlin Syndrome, the impact of Shh signalling upon cerebellar development has been extensively analysed. Wallace, 1999, established in mice that Purkinje cells have the ability to control the proliferation processes of GNPCs by releasing Shh postnatally, creating a morphogenic gradient within the developing cerebellum. Furthermore, the authors demonstrated that key proteins mediating Shh signalling, such as, *Ptch1*, *Gli1* and *Gli2*, are primarily expressed within the EGL during the postnatal period (Figure 1.9). The hypothesis of a morphogenic gradient was supported by the application of Shh proteins *in vitro* to sections of murine postnatal EGL which resulted in increased expression of *Gli1*, reduced differentiation, and a 20X increase in mitotic activity compared to other known mitogens (*IGF1*, *EGF* and *bFGF*) as assayed by the presence of bromodeoxyuridine (BrDu) (Wechsler-Reya and Scott, 1999). Conversely, *in vivo* experiments using Shh blocking antibodies by hybridoma cell implantation directly influenced the size and foliation of the developing cerebellum, indicating that Shh supports the definition of cerebellar size and foliation (Wallace, 1999).

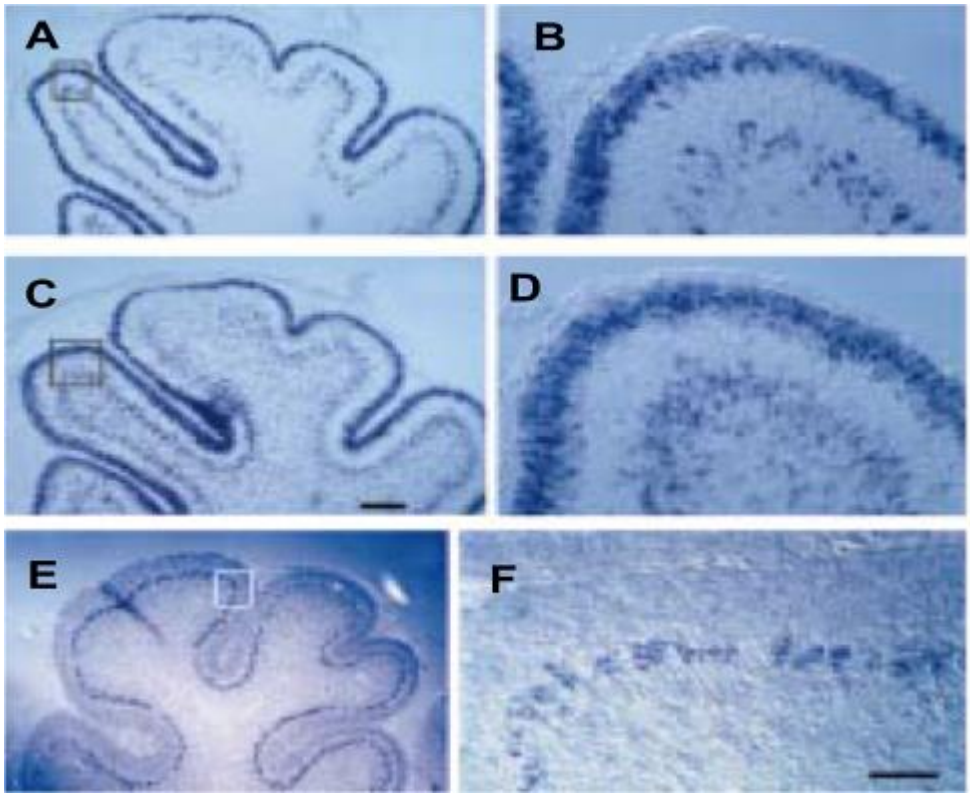

Figure 1.10: The expression of downstream target genes for Shh signalling and its impact in/on the developing cerebellum. (A and B) Show expression of Gli1 in the cortex of the cerebellum; C and D, show expression of Ptch in the cerebellar cortex; E and F show that expression of *Shh* is confined to the Purkinje cells. cerebellar cortex. Expression appears as blue staining in both low and high magnification images. (Adopted from Wallace 1999)

1.12 Project Aim and outline of results chapters

The analyses outlined above indicate that the primary defects of many MBs remain to be elucidated despite recent genome wide mutation analyses, that the *Ptch1*+/- murine model of MB produces tumours with a similar gross morphology to human tumours, and that the basic role of SHH signalling in cerebellar formation is understood. The primary aim of this project was to identify novel genes involved in SHH subtype MB using murine mutagenesis. As there is some evidence that the SB mutagenesis system is capable of generating MBs, this system was chosen as the mutagen and applied to the *Ptch*+/- murine model.

In the first results chapter (Chapter 3), I will outline the screen and the initial results, which establish that SB mutagenesis impacts significantly upon MB frequency in the *Ptch*+/- model, and also generates additional tumour types including liver adenomas. I will also describe the subsequent mapping and identification of candidate cancer genes.

In the second chapter (Chapter 4), I outline the results of gene specific PCR analyses which I applied to three genes; *Tgif2*, *Nfia* and *EGFR*, to confirm the validity of the insertion site data, together with gene ontology analyses which identified a significant excess of transcription factors and neuronal genes among the candidate genes identified.

In the third results chapter (Chapter 5), bioinformatic analyses are presented which establish that the candidate genes not only share common ontologies, but are linked within a single gene expression network, the activity of which differs significantly among tumour subgroups and is associated with survival in human tumours. Additional evidence for the role of this network in tumour formation is also presented in the form of a comparative analysis with mouse gene expression I have generated from SB induced tumours.

Finally, in Chapter 6, I prioritise genes for downstream analysis using existing gene expression data from murine GNPC cells, and identify the genes *Myt1L* and *Tgif2* as prime candidate for experimental validation. I also successfully knock down *Myt1L* expression in GNPC cells and show that the expression of genes involved in both SHH signalling and neuronal differentiation are significantly altered, further validating the importance of the network in SHH subgroup tumour development and neuronal development.

Chapter 2

Materials and Methods

2.1 Mouse Husbandry

Mouse strains used in this project are as follows: B6 *Ptch*tm1mps/J mice (developed by targeted mutation of exons 1-2 of the *Ptch1* gene, (Goodrich *et al.*, 1997); CBA wild type (Charles River Laboratories, Margate, UK); T2/Onc Line 76 (Collier *et al.*, 2005); and SB11 Rosa26 (Dupuy *et al.*, 2005). All mouse work was conducted under, and in accordance with, UK Home Office standards and was completed under Project licence PLL/60/3621. All transgenic strains were first back crossed for a minimum of 3 generations to wild type CBA mice to reduce any influence of genetic background. Then, double transgenic *Ptch1*+/-;T2/Onc+/- mice were mated to heterozygous SB11+/- mice to produce triple transgenic experimental mice, and double transgenic control genotypes (see Figure 3.2). All animals were checked minimally 3 times a week, and those with symptoms of illness were humanely culled. Due to the different tumours arising within the cohort, symptoms of illness were varied. These included weight loss, disinterest in environment, abdominal distention, and behaviour changes indicative of balance problems such as leaning to one side, and head shaking or spinning.

2.2 PCR genotyping

Identification of the various genotypes from young transgenic animals was performed as follows:

DNA templates were prepared by heating ear punch material in 100µl of 25mM NaOH/0.2 mM EDTA at 95°C for 30 minutes. Then, 100 µl of 40 mM TrisHCl was added. Genotyping was carried out in a total volume of 20µl, containing 2µl of genomic DNA (~200ng), 18µl master mix containing 2µl of 6µM forward primer (SB11 5' –ATGGGAAAATCAAAAGAAATCAGCCAAG–3', or *Ptch1* 5' -GTTGCAGTGCACGGCAGATACACTTGCTGA-3'), 2µl of 6µM reverse primer (SB11 5' –GCCAAACAGTTCTATTTTTGTTTCATCAGACCA–3', or *Ptch1* 5' -GCCACTGGTGTGGGCCATAATTCAATTCGC-3'), 2µl dNTPs mix, 4µl 5X buffer and 7.7µl distiled water. The amplicon size for *SB11* is 483bp and *Ptch* is 482bp. Cycle conditions were as follows: Initial denaturation at 94°C for 5 minutes, followed by 34

cycles of 1 minute at 94°C (denaturation), 1 minutes at 55°C or 60°C for SB11 or Ptch respectively (primer annealing), and 1 minute at 72°C (for extension), followed by a final extension at 72°C for 5 minutes. Then 5µl of PCR products were electrophoresed on 1.5% agarose gel at 90V for 40 minutes, with 5µl of a 100pb ladder as size standard.

2.3 Insertion Site Amplification using Linker-mediated PCR

2.3.1 DNA template preparation for insertion site amplification

The samples from both control and tumour were treated as follows: Approximately 3mm cubed sample was transferred to an eppendorf tube. 500µl of lyses buffer and 50µl (10mg/ml) of proteinase-K (Bioline Reagents Limited, London, UK) were added, and incubated at 50°C overnight. The mixture was centrifuged at 10,000rpm for 5 minutes. DNA was then extracted using the Qiagen Bio Robot EZ1 machine (Qiagen, Hilden, Germany) according to the manufacturer recommendations and resuspended in 200ul of water. The final concentration of the DNA was measured using a NanoDrop ND-1000 spectrophotometer (Thermo, Wilmington, Delaware, USA), and each sample was adjusted to a final concentration of ~100ng/µl.

2.3.2 First Digestion

For each LM-PCR reaction (Figure 2.1), 2µg of genomic DNA was digested with *BfaI* (left hand side of the transposon) or *NlaIII* (right hand side of the transposon) in a total volume of 50µl containing 20U of enzyme and 5µl of 10X reaction buffer. Reactions were incubated at 37°C overnight, deactivated by heating at 80°C (*BfaI*) or 65°C (*NlaIII*) for 20 minutes, and then purified using a QIA quick PCR purification column (Qiagen, Hiden, Germany). All digestions were resuspended in 35µl distilled H_2O, with 15µl of the resuspension being used for quality control on a 1.5% agarose gel as required.

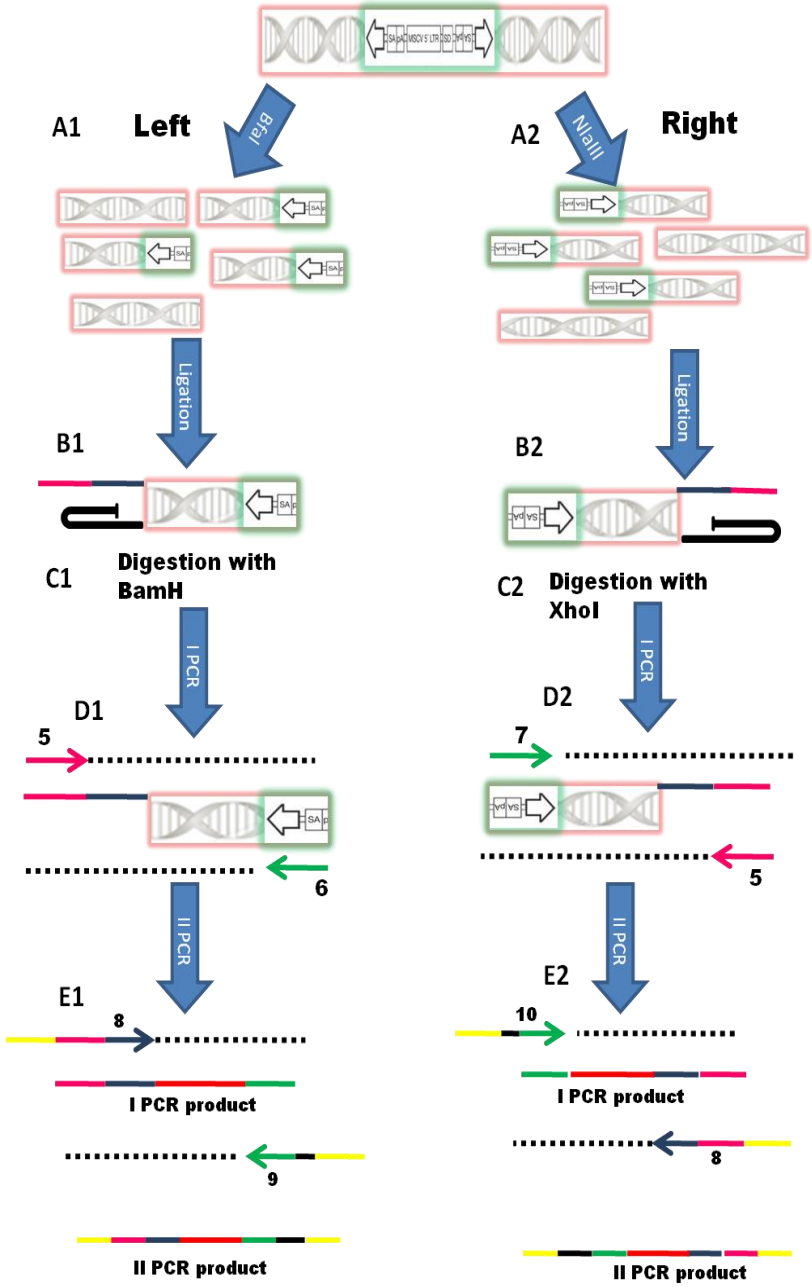

Figure2.1: LM-PCR protocol. (A1 and A2) digestion of sample DNA with Bfa1 (Left hand side of the T2/Onc "Green"), or NlaIII (right hand side of the T2/Onc "Green") enymes. (B1 and B2) Ligation with linkers (pink indicates unique single stranded sequence, black indicates phosphatase splink with hairpin complementarity). (C1 and C2) Further digestion with BamH1 (Lift handside of T2/Onc) or Xhol (Right handside of T2/Onc) restriction enzyms to eliminate the amplification of T2/Onc elementswhich remain in the transgenic array.(D1 and D2) Primary PCR with a common forward primer "5" and a primer complementary to the unique single stranded sequence in the lift had side of T2/Onc "6" and the right "7" linkers. (E1 and E2) Secondary PCR using a common nested primer"8" with end specific nested primers "9" for the Lift hand side of T2/Onc or "10" for the lift, which contain barcoded FLX sequencing primers (see methods). **Figure 1.8: Cerebellar anatomic structure**: The two main cerebellar layers, the cerebellar cortex and the white matter are

2.3.3 Splinkerette linker Ligation

For proper splinkerette linker attachment, two *Bfa*I and *Nla*II linkers were used:

Top *Bfa*I oligonucleotide sequence:

5' – GTAATACGACTCACTATAG**GGCTCCGCTTAAGGGAC** –3'

Bottom *Bfa*I oligonucleotide sequence:

3' –AAAAAAACTTTTTTTTTTAAT**CCGAGGCGAATTCCCTG***AT*-Phos 5'

Top *Nla*III oligonucleotide sequence:

5' -GTAATACGACTCACTATAG**GGCTCCGCTTAAGGGAC***CATG*-3'

Bottom *Nla*III oligonucleotide sequence:

3' -AAAAAAACTTTTTTTTTTAAT**CCGAGGCGAATTCCCTG**-Phos 5'

In both cases, the top and bottom primers were annealed (final concentration 100mM in 200mM NaCl) to generate single stranded overhangs complementary to the endonuclease sites (shown in italics). Primers complementary to each linker primer are subsequently used in the primary PCR reaction (sequence underlined). The ligation reaction conditions were as follows: 95°C for 5 minutes as first denaturation, followed by 95°C for 2 minutes, then primer annealing at 52°C for 10 minutes. Linker ligations were performed in a total volume of 20µl per reaction containing 10µl of either *Bfa*I or *Nla*III digested DNA, 2µl of 10X buffer, 6µl of annealed *Bfa*I or *Nla*III linkers, and 2µl of 800U T4 DNA ligase, and were incubated at room temperature overnight. The ligation reactions were then purified using a QIAquick PCR purification column (Quiagen, Hiden, Germany), and resuspended in 30µl H_2O. Where necessary, ligation efficiency was checked using a digest only sample to enable visualisation of ligation products.

2.3.4 Second digestion

Ligated DNA from 2.2.4 was digested with either *Bam*Hl or *Xho*l, for the left and right reactions, respectively. The total volume in both cases was 50µl, containing 28µl DNA, 1µl one of the appropriate enzyme (20U), 5µl buffer, 0.5µl of BSA and 15.5µl of H_2O. Both reactions were incubated at 37°C overnight. *Xho*l digests were deactivated by heating at 65°C for 20 minutes, and digested DNA was then purified

using a QIAquick PCR purification column (Quiagen, Hiden, Germany), and resuspended in 30µl H_2O.

2.3.5 Primary PCR

Primary PCR was carried out using a splinkerette primer '5' (5' -GTAATACGACTCACTATAGGGC-3'), and either a left IR/DR primer (*Bfa*I) ' 6' , (5' -CTGGAATTTTCCAAGCTGTTTAAAGGCACAGTCAAC-3'), or right IR/DR primer (*Nla*III) '7' (5' -GCTTGTGGAAGGCTACTCGAAATGTTTGACCC-3') in a total volume of 50µl, containing 3µl of *BamH*l or *XhoI* digested DNA, and 47µl of a master mix containing: 0.5µl of 2.5mM '5' as a common primer, 0.5µl of either 2.5mM '6' primer for *BfaI* or 2.5nM '7' primer for *NlaIII*, 5µl 10X buffer, 2µl of 50mM $MgCl_2$, 1µl of 10mM dNTPs (2.5mM each) and 0.25µl high fidelity Platinum Taq (Invitrogen, Madison, WI, USA). PCR was performed with an initial denaturation step of 94°C for 2 minutes, followed by 25 cycles of 15 seconds at 94°C (denaturation), 30 seconds at 60°C (primer annealing), and 90 second at 72°C (for extension), followed by a final extension of 72°C for 5 minutes. The PCR products were subsequently diluted 1:75 in H_2O to be used in the secondary PCR.

2.3.6 Secondary PCR

Secondary PCR was carried out using diluted primary PCR product as a template, and used nested versions of the primary PCR primers carrying both the required fusion sequences for GS20 FLX Pyrosequencing (Fusion L left and right and Fusion B) and a unique 10bp barcode recognition sequence for each sample. Primers were designed such that the nested transposon primer had the Fusion A and barcode attached (Fusion A (Italic) – barcode (bold) – nested SB primer (underlined) in the sequences that follow):

Primer ' 9' ; 5' *CGTATCGCCTCCCTCGCGCCATCAG***TCGGAGTCAA**<u>GACTTG</u>

<u>TGTCATGCACAAAGTAGATGTCC</u>3' and

Primer ' 10' 5' *CGTATCGCCTCCCTCGCGCCATCAG***TCGGAGTCAA**<u>TAAGGT</u>

<u>GTATGTAAACTTC</u>3' for left and right, respectively.

The nested linker primer has the Fusion B sequence attached (linker nested (Italic) – Fusion B (underline)):Primer ' 8' ;

5' -*CTATGCGCCTTGCCAGCCCGCTCAG*AGGGCTCCGCTTAAGGGAC-3' .

The reactions were performed in a final volume of 50µl containing 3µl of diluted primary PCR product (1:75) and 47µl of a master mix containing: 0.5µl of 2.5mM primer '8' as a common primer, 0.5µl of either 2.5mM primer '9' for *BfaI* or 2.5nM primer '10' for *NlaIII*, 5µl 10X buffer, 2µl of 50mM $MgCl_2$, 1µl of 10mM dNTPs (2.5mM each) and 0.25µl high fidelity Platinum Taq (Invitrogen, Madison, WI, USA). PCR was performed with an initial denaturation step at 94°C for 2 minutes, followed by 25 cycles of 15 second at 94°C (denaturation), 30 seconds at 60°C (primer annealing), and 90 second at 72°C (for extension), followed by a final extension at 72°C for 5 minutes. 20µl of Secondary PCR products were checked on a 1.5% agarose gel electrophoresed at 90 Volts for 40 min.

2.3.7 Sample preparation for FLX sequencing

Each secondary PCR sample was purified using a QIAquick PCR purification column (Qiagen, Hilden, Germany), and eluted in a final volume of 30µl H_2O. DNA quality and quantity was checked using the Nanodrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, Delware, USA). Approximately, 100ng/µl DNA of each sample was pooled for FLX sequencing as appropriate, with adjustment to take account of average product size. Pooling was performed to generate on average 500 reads per sample. Samples with less than 100 reads were re-sequenced prior to data analysis. After pooling, a final purification step using a Qiaquick PCR purification column (Quiagen, Hiden, UK) was performed. The samples were diluted to a final concentration of 2 Â 10s molecules/ml for pyrosequencing using the FLX LS20 or Titanium protocols (Roche, Penzberg, Germany). This final dilution, and all FLX sequencing, was performed by Dr Jonathan Coxhead (NewGene Ltd, Newcastle University), according to manufacturer's recommended protocols.

2.4 Sequence analysis

To generate sequence files specific for each PCR reaction, GSMapper software (Roche 454 Life Science, Branford, CT, USA) was used to map all sequence reads to a virtual genome consisting of all barcoded FLX sequencing primers. As a preliminary check of data quality, the left and right sequence data from one spleen tumour sequenced in the first FLX run (PTS11-21) were manually mapped to the mouse genome (build July 2007 [NCBI37/mm9]) using the BLAT algorithm (Kent 2002). In regards to the entire FLX sequencing data set, After trimming off all reads from the

linker fragments and FLX barcodes, all sequencing reads were mapped to the mouse genome using the methods described in March *et al.,* (2011). These genomic sequences were then aligned to the mouse genome reference assembly (build July 2007 [NCBI37/mm9]). The output was then further filtered utilising the following criteria: the best match cut off has to be at least 90% identical, including a perfect match to the "TA" at the bigininng of the alignment.

2.5 Gel preparation and electrophoresis

1.5gm of Saekem Agarose gel (Lonza, Rockland, USA) was dissolved in 100ml of 1% tris-acetate-EDTA (TBE) Buffer. 100µl of (1:5) Gel red (Biotium Inc., Hayward, CA, USA) was added to the gel for visualization purposes. Electrophoresis was performed as 90 volts for 40 minutes.

2.6 RNA Preparation

All reagents utilized in RNA preparations were prepared using Diethylene Pyrocarbonate (DEPC) treated water. All surfaces and required materials for RNA isolation and preparation were cleaned with RNAseZap (Ambion Inc., Austin, Texas, USA).

2.6.1 RNA prepared from mouse brain tissue

Total RNA was isolated from ~10mg normal cerebellar tissue or MB tumours obtained by Dr Maria Lastowska or Dr. Mike Jackson (Institute of Genetic Medicine, Newcastle University) which were stored at -80°C. RNA was isolated using the RNeasy Lipid Tissue Mini Kit (Qiagen, Hilden, Germany) in accordance with the manufacturer's recommended protocol. The Isolated RNA was eluted with 50µl RNAse free water.

2.6.2 RNA prepared from primary and established cell lines

Total RNA was isolated from granule neuron precursor cells (GNPCs) or 3T3 cell line pellets using the RNeasy Mini Kit (Qiagen, Hilden, Germany) in accordance with the manufacturer's protocol. The Isolated RNA was eluted with 50µl RNAse free water.

2.6.3 DNase treatment of isolated RNA samples

All isolated RNA samples were treated with RNase-Free DNase according to the manufacturer's recommended protocol (Qiagen, Hilden, German) to remove remaining plasmid or genomic DNA.

2.6.4 RNA quantitation and integrity ratio determination

Following elution, RNA concentration was determined using a Nanodrop (Labtech International, Uckfield, East Sussex, UK). All samples with an A260/A280 ratio of less that 1.7 or greater that 2.2 were disposed of. The RNA integrity number (RIN) (8/10) and RNA quality were determined by using an Agilent 2100 Bioanalyser (Agilent Technology, Waldbronn, Germany).

2.6.5 cDNA synthesis

cDNAs were synthesised using a High Capacity cDNA Reverse Transcriptase kit (Applied Biosystem, Warrington, UK). A total of 50-100ng of RNA was used as template, with 2µl of 10X RT Buffer, a final concentration of 80mMdNTP mix, 20U of RNase inhibitor, 2µl of 10X RT random primers, and 50U reverse transcriptase, the volume being made up to a total volume of 20ul with distilled water. cDNA quantity and quality were determined by a using NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, Delaware, USA).

2.7 Quantitative real time PCR (qPCR)

All PCR products were designed to be in the size range of 90 – 150bp in length. qPCR was performed using the ABI 7900 HT Fast real-time PCR system (Applied Biosystems, Warrington, UK) in accordance with the manufacturer's protocols. All qPCR probes were designed to be annealed at ~70°C; and included a FAM fluorochrome at the 5' end of the probe, while the 3' end was conjugated to the quencher TAMRA (Metabion, Martinsried, Germany). The primer concentration was optimized by assessing reaction efficiency at 3 different concentrations (100, 200 and 300pM) with 3 different cDNA concentrations (1, 10 and 100ng). In a 15µl rection,

100pM TaqMan probe was used, with 10ng cDNA, and 7.5µl of TaqMan Universal PCR Master Mix (Applied Biosystem, Warrington, UK). All PCR reactions were carried out in triplicate using the following amplification conditions: 50°C for 2min to activate the polymerase, an initial denaturing step of 95°C for 10min, followed by 45 cycles of 95°C for 15sec for denaturing, and 60°C for 1 min for the annealing and elongation steps**.** Noteworthy, all primer were designed by Primer3 software (http://flypush.imgen.bcm.tmc.edu/primer/primer3_www.cgi) with accemption of *Gli1 and Math1* being adabted from (Romer *et al.*, 2004)

Gene expression was normalised using a ubiquitously expressed gene as a calibrator (β-Actin). ΔCt was then determined as the difference between the average Ct values of the target gene and the average Ct values of the calibrator gene (ΔCT = CT (target) - CT (calibrator)). Relative expression was then calculated using the ΔΔCt method, as the difference between experimental samples and the calibration reference control sampls (either 0-time or normal cerebella cDNA, where applicable) (Comparative expression level = ΔCt target – ΔCt reference). The absolute comparative gene expression was calculated as $2^{-\Delta\Delta Ct}$ (Livak and Schmittgen 2001).

Table 2.1: List of primers used in qPCR

Primer	Sequence
Tead1_real_T_F	CATCCACAAGCTCAAACACC
Tead1_real_T_R	CAAATACACAGGCCATGCAG
Tead1_Probe	TGGTGGTAACAAACAGGGATACACAAGAAA
Myt1l_real_T_F	AACCTGAGCCAGTCCCTGAT
Myt1l_real_T_R	CATTTCCGTCAAAGTAGTCACG
Tead1_Probe	CGCCAACATCCAGCTGCCTC
Tgif2_real_T_F	GGCTCTATCTGCACCGCTAC
Tgif2_real_T_R	GAAGCATGTCAGGGAGAAGG
Tgif2 _Probe	CCAACCTCTCGGTGCTGCAGA
Igf2_real_T_F	CGGACTGTCTCCAGGTGTC
Igf2_real_T_R	GTCCGAGAGGGACGTGTCTA
Igf2_Probe	GGAAGAACTTGCCCACGGGG
Gli1_Real_T_F	GCTTGGATGAAGGACCTTGTG
Gli1_Real_T_R	GCTGATCCAGCCTAAGGTTCTC
Gli1_Probe	CCGGACTCTCCACGCTTCGCC
Math1_Real_T_F	ATGCACGGGCTGAACCA
Math1_Real_T_R	TCGTTGTTGAAGGACGGGATA
Math1_Probe	CCTTCGACCAGCTGCGCA ACG

2.8 Granule Neuron Precusor cells (GNPCs) isolation

GNPCs isolation was performed according to a protocol provided by Dr Beth Hatten (St Jude's Hospital, USA). All chemicals were obtained from Sigma, Aldrich, Missouri, USA, unless otherwise stated. Cerebella were dissected from Post natal day 4-6 (P4–6) mice, when most proliferation is occurring in the EGL. All animals were killed according to a schedule 1 method (dislocation). The head was cut off and skin was peeled off. The skull was then cut from the spine on each side of the head toward the eyes. The brain was removed from the skull in one piece and transferred into a petri dish containing ice cold CMF-PBS [consisting of NaCl 8g, KCl 1.2g , glucose 8g, $NaH_2PO_4.H_2O$ 2g, KH2PO4 1g, NaHCO3 8ml of 2% stock , EDTA 10ml of 1M EDTA, to a final volume of 1 litter H2O, PH 7.4). Pia and meninges were carefully teased off completely from the cerebellum, and the cerebellum was transferred into a 15ml falcon tube containing fresh ice cooled CMF-PBS to await GNPC isolation.

Once all dissections were completed, the CMF-PBS was removed, and 1ml Trypsin-DNAse [CMF-PBS 14.91ml, Trypsin 150mg (Worthington Biochem. Corp., NJ, USA), DNAse 15mg (Worthington Biochem. Corp., NJ, USA), 1N NaOH 0.09ml and MgSO4.7H2O 22.5mg to final volume 15ml) was added to a maximum of 6 cerebella, and incubated for 5-15 min at 37°C.

The trypsin-DNAse solution was then removed by pipetting, 1ml of DNAse solution was added [DNAse 7.5mg (Worthington Biochem. Corp., NJ, USA) dissolved in 14.83ml BME media (Gibco, Paisley, UK) and 0.113ml of Glucose] and the cells were dissociated by regular pressure using decreasing gauge needles (19 and 21 gauge). The cells were centrifuged for 5 min at 800g and 4°C (with break ON). After removal of the supernatant, the cell pellet was resuspended in 2ml fresh DNAse solution and 4ml CMF-PBS solution. The cells solution was then split into two 3ml aliquots and in 15ml falcon tube layered on to a Percoll gradient consisting of 5ml of 35% percoll solution [100 ml prepared by adding 25ml CMF-PBS-EDTA to 35ml Percoll and 40mldH_2O), and 5ml of 60% percoll [100 ml prepared bt adding 25ml CMF-PBS-EDTA to 60ml Percoll 60ml, 15ml dH_2O, and 0.3ml 0.4% trypan blue). The tubes were then centrifuged at 2000g for 10 min at 4 °C (Break ON).

After centrifugation, the cell were gently and carefully harvested from the intermediate layer situated on top of the 60% percoll solution and transferred into new 15ml falcon tube filled with ice cooled CMF-PBS solution to dilute the percoll. Cells were then centrifuged at 700g for 5 min at 4°C. The CMF-PBS was then removed and the cell pellets were resuspended in 50µl of DNAse and 2ml GNPCs media [in 50ml

Neuro basal media (Gibco, Paisley, UK) add 0.5ml 100X Pen/Strep, 0.5ml 100X l-Glutamin, 0.5ml of D-Glocose 45%, 1ml SITE and 1ml of 50X B27 (Gibco, Paisley, UK), 50µl of 1000X N-Acetyl Cystein] and centrifuged at 700g for a further 5 min at 4°C. The next step was to remove the supernatant and resuspend the cell pellet in 50µl DNAse and 2ml GNPCs media. The cells were plated in a 35 mm cell culture plate, and incubated at 35°C with 5% CO_2. After 1 hour of incubation the cells were transferred to a new 35mm culture plate and reincubated for 1 hour. The cells were plated at the desired density, ranging from 1–5X10^5/cm^2, on a precoated culture plated with 100µg/ml of poly-D-lycin (Millipore; Billerica, MA, USA) and matrigel (200µl in 15mlCMF-PBS). Recombinant Shh protein at a concentration of 3µg/ml was added on the day of culturing where applicable. The media was changed every 3 days, and Accutase solution (Innovative Cell Technologies Inc., San Diego, CA, USA) was used for cells detachment purposes.

2.9 Bacterial transformation and constructs

2.9.1 Lentiviral constructs

Five pLKO.1-puro based lentiviral plasmids (Mission-RNAi, Sigma-Aldrich, USA) containing shRNA-Myt1l were used for the *Myt1L* knockdown analysis: shRNA-Myt1l-8 (TRCN0000012108), shRNA-Myt1l-9 (TRCN0000012109), shRNA-Myt1l-10 (TRCN0000012110), shRNA-Myt1l-11 (TRCN0000012111), shRNa-Myt1l-12 (TRCN0000012112), together with one pLKO.1-puro empty vector (Mission-RNAi, Sigma-Aldrich, USA). A pReceiver-Lv102-FLAG-epitope Tag plasmid, together with a puromycin based lentiviral vector containing the full length cDNA for *Tgif2* (ORF-Tgif2) with an N-terminal FLAG-epitope Tag driven by the CMV promoter (Ex-Mm15242-Lv102, Genecopoeia, USA, http://www.genecopoeia.com), were used for the *Tgif2* over-expression. Of these, only shRNA-Myt1l-8, shRNA-Myt1l-9, and ORF-Tgif2 were used in transduction analyses.

2.9.2 Transformation

For each transformation, 50µl of competent E.Coli cells (DH5α, Gibco-BRL, Life Technologies, Rockville, MD) were thawed on ice, then 1µl (~50ng) of the appropriate Lentiviral vector was added and incubated for 20 minutes on ice. The cells were then heat-shocked at 42°C for 45 seconds, followed by a further incubation on ice for 5 minutes. Following this, 950µl of LB broth (with no antibiotic) was added to each

transformation, which were then incubated at 37°C for 1 hour. Each transformation mix was then spread out in different quantities onto agar plates containing ampicillin and the plates were incubated overnight at 37ºC. Individual white colonies were resuspended in 5ml of media with 5µl of ampicillin (50µg/ml), to harvest the plasmids and 100µl of the above mixture was added to sterile water for colony PCR to check if the colony contained the vector. The DNA plasmids were then purified using mini prep columns (Qiagen, Hilden, Germany).

2.9.3 Virus Production and titration

The virus production was performed utilising Calcium phosphate transfection technology as previously described with lentiviral packaging vectors (Tarkkonen *et al.*, 2012), with slightly modification. The procedure was performed using a CalPhos Transfection Kit, according to the manufacturer's protocol (cat. 631312, Caltech, Pasadena, CA, USA), by co-transforming 10µg of the appropriate lentiviral vector with 5µg of pMD2.G and 10µg of psPAX2 into ~80% confluent of Human embryonic Kidney (HEK293T) cells (provided by Dr Dan Williamson, Newcastle University) cultured in DMEM complete media per 10-cm plate. After 48hrs the virus containing media was collected and filtered (0.45µm filter) and the lentiviral titer was established using 293T cells (Tarkkonen *et al.*, 2012) using a 6-well plate format: 2µl of the viral supernatant was added to the second well, and a serial dilution was performed by transferring 200µl of viral containing media into the third well and repeating the dilution for wells 4-6. This resulted in viral concentrations of $1X10^3$ (well 2), $1X10^4$, $1X10^5$, $1X10^6$ and $1X10^7$ (well 6) cfu/ml (colony forming unit). Puromycin (1µg/ml) was added 24hr post infection. After 8 days, the cells were fixed and stained with crystal violet and the colonies were counted.

2.10 Generation of stable *Myt1l* Knockdown GNPCs

The 2 shRNA-Myt1ls with the highest titration were used in the transduction procedure with GNPCs. shRNA-Myt1l-8 ($4X10^6$), shRNA-Myt1l-9 ($1.5X10^6$), PLKO.1 empty vector ($2.5X10^6$) were each used to infect $1.5X10^5$ GNPCs, seeded post isolation in 96-well plates with the presence of 3µg/ml recombinant Shh protein (Cell Guidance Systems, Cambridge, UK) where applicable. After 24hr of GNPC culture, 3 multiplicities of infection (MOI) of the viral supernatants were added. Puromycin (1µg/ml) was added 24hr post transduction, then cells were harvested after 3 days for gene expression analysis.

2.11 Generation of stable *Tgif2* expressing cells

Mouse 3T3 cells were obtained from Prof. Linda Lako (Institute of Genetic Medicine, Newcastle, UK). Around 187,500 cell/T25 flask were seeded in complete DMEM media. Twenty four hours post culturing, the lentivirus carrying ORF-Tgif2 ($2X10^6$ /ml titration) was added for transduction at 3 different MOIs (1MOI, 5MOI and 10MOI; 93.75µl, 468.75µl and 937.5µl, respectively). After 24hr of transduction 2µg/ml of puromycin was added to select the vector containing cells. The cells were maintained according to standard cell culture protocols and harvested 3 days post transduction. Control cells were established in parallel to these experiments.

2.12 Protein isolation from whole cell lysates

Protein was extracted from whole cell lysates using urea buffer. Briefly, 50µl of urea buffer, consisting of 90% urea (8.8M), 8% 1ml Tris and 2% sodium phosphate (5M), was added to cell pellets and vortexed for 30 sec. Protease inhibitor (1µl) was then added and the lysates were incubated for 30 min on ice. The lysates were then centrifuged at 4°C for 30 min, and the supernatant was collected and stored at -80°C for future use.

2.13 Protein quantification

The proteins extracted from GNPCs, 3T3 cells, and whole brain were quantified using Micro BCA Protein Assay Kit (Pierce, Rockford IL, USA). All protocols and steps were conducted and completed in accordance with manufacturer's recommended protocols. A standard curve was constructed using bovine serum albumin (provided

with the above kit between the concentration 0 - 200μg/ml). Absorption was measured at 562nm and all samples were quantified in duplicate.

2.14 Wester Blot assay

Protein extracted from cells or tissues was mixed with 2X SDS sample buffer (1.52g Tris base, 20ml glycerol, 2.0g SDS 2.0ml β-mercaptoethanol, 1mg Bromophenol blue, pH 6.8 to a volume of 100ml with H_2O) and incubated at 95°C for 5 min. For each sample, 30-50μg of whole cell lysate was loaded on 4–20% Amersham ECL Gel (Amersham International, Amersham, UK). Samples were run on the gel with 1X running buffer [consists of 100ml of 10X running buffer (10X buffer consists of 30.3g Tris base, 144g Glycine and 10g SDS and 1L of H_2O) and 900ml H2O] for 45min to 1 hour depending on protein size. The samples were then blotted using 1X blotting buffer [consists of 100ml of 10X blotting buffer (10X blotting buffer consists of 30.3g Tris base, 144g Glycin and 1L of H_2O) diluted with 200ml absolute Ethanol and 700ml H_2O] for 1–2 hours depending on protein size. Protein transfer was checked by staining the gel with Ponceau Red for 5 min. Non-specific binding was blocked by incubating the filter in a 10% milk solution [5g milk in 50ml 1X TBST (10X 24.23g Trizma HCl, 80.06g NaCl) at 4°C overnight. After washing the membrane, primary antibody (Anti-*Myt1l* Ab (Millipore, Germany) or AntiD Ab for ORF-Tgif2 protein (CGAB-DDK-0050, Genecopoeia, USA) was diluted in 1% milk (0.5g milk in 50ml TBST) and applied for 1 hour with continuous agitation. The membrane was then washed and secondary Ab (BD-Pharmagene, Sandiago, CA) diluted in 1% milk, was applied for 1 hour. Pictures of protein bands were developed by Syngene BOX-Gel Documentation and Analysis System (Syngene, Cambridge, UK).

2.15 Illumina expression analysis

Approximately 200ng of total RNA was amplified and biotin labeled from each of 30 triple transgenic experimental mice tumours (*Ptch1*+/-;T2/Onc+/-;SB11+/-), 6 tumours from *Ptch1*+/- predisposition only mice, and 6 normal cerebella, utilizing the Illumina TotalPrep RNA amplification Kit (Applied Biosystem, Foster City, USA) in accordence wth the manufacturer's recommendations. The RNA Integrity number (RIN) was higher than 8.0 for all samples, with the exception of 2 samples where the RINs were 7.7 and 6.7. The size distribution of generated cRNA was checked using an Agilent 2100 Bioanalyser (Agilent Technologies, Santa Clara, CA, USA. Around 750ng

of each cRNA was then hybridised to Illumina Mouse8 reference Arrays (Illumina, Essex, UK) in accordance with the manufacturer's protocols by the Wellcome Trust Clinical Research Services (Edinburgh, UK). Microarray gene expression data generated was deposited in the Gene Expression Omnibus as submission GSE43994.

2.16 Identification of common insertion sites within tumours and controls

In all, 41 MBs from the experimental genotype, and 30 cerebellum controls from the transposition-only genotype, were analysed by linker-mediated PCR designed to amplify across insert/genomic DNA junctions. In addition, 19 hepatocellular neoplasias were also analysed. The products were then bar-coded by nested PCR and sequenced using FLX pyrosequencing (Dupuy *et al.*, 2009) that was aimed to a target depth of ~500 reads/tumour. This resulted in a final total of 8,360 inserts with an average of ~204 inserts/tumour. These inserts were further analysed with two statistical methods to identify statistically significant common insertion sites; Monte Carlo simulation (MC) and Gaussian Kernel Convolution (GKC) (Collier *et al.*, 2005;de Ridder *et al.*, 2006) by Dr Alastair Rust and David Adams at the Wellcome Trust Sanger Institute (Hinxton, Cambs, UK).

2.17 Gene Expression Microarray and network analysis

Human expression data were taken from publicly available data comprising 119 Affymetrix HGU133p2 arrays (Kool *et al.*, 2008;Fattet *et al.*, 2009). Mouse expression data were taken from (Gibson *et al.*, 2010;Subkhankulova *et al.*, 2010). All analyses were performed on per chip and per gene normalised log2 transformed data using Genespring GX7.2 (Agilent Technologies, Wokingham, Berkshire, UK). Where multiple Affymetrix probes were present for a single gene, the probe with the highest mean expression value across all samples was analysed. Murine orthologs of human Affymetrix probes were specified using UCSC (http://genome.ucsc.edu). All t-test p-values were adjusted using a Benjamini-Hochberg False Discovery Rate of 0.05.

ARACNe network generation and Metagene analyses were performed by Dr. Dan Williamson (Northern Institute of Cancer Research, Newcastle upon Tyne UK). Briefly, publicly available Affymetrix HGU133.2 expression data generated from 119 human MB samples generated by Kool *et al.* (2008) and Fattet *et al.* (2009) were used to establish the ARACNE network. Management of CEL files was performed using the Bioconductor RMA package (Irizarry *et al.*, 2003). The Affymetrix probes were filtered down from the original ~500,000 by retaining only probes that were expressed in at

least 10% of samples with a range of expression greater than 500 and an overall fold change of 3 or more. Only probes with a significant Mutual Information (MI) association with at least one other node were retained. For retention in the full network, the genes had to be connected to each other by at least one neighbour node with a bootstrapped P-value of <10E-8. This left 8,749 probes in the network.

For Non negative Matrix Factorization (Tamayo *et al.*, 2007), metagenes defining the 4 MB subgroups generated from human tumours data were superimposed onto mouse arrays using orthologs and used as the basis of support vector machine (SVM) classification, where the classification training was performed using human MB data, and the classification was then tested on mouse primary tumours data.

Two soft wares were used to create the Ontology results one is the Bingo 2.44 Cytoscape plug (Maere *et al.*, 2005), and the other is the Generic Gene Ontology Term Finder (http://go.princeton.edu/cgi-bin/GOTermFinder).

Chapter 3. (Results 1)

Sleeping Beauty mutagenesis identifies candidate genes for involvement in Medulloblastoma.

3.1 Introduction

3.1.1 The SB transposon

The Sleeping Beauty (SB) transposon belongs to the Tcl/mariner superfamily of transposons (Largaespada, 2009) which includes a large quantity of transposons in organisms such as fish and insects. The superfamily is defined by conserved amino acids responsible for equivalent processes in both families (Ivics *et al.*, 1997). The SB transposon was first identified as a Tc1-like fossil in Salmonid species and the function element, which uses replicative transposition to insert into the target DNA, was reconstructed using phylogenetic inference (Ivics *et al.*, 1997). Structurally, it is made up of a transposase which is delimited on both sides by terminal inverted repeats. Like other transposons which insert using non-replicative transposition, SB transposes using a cut and paste mechanism. Briefly, the transposase identifies the inverted repeat ends and cleaves the transposon from surrounding DNA. The cleaved transposon is then inserted into the target DNA and the transposon and DNA are joined together (Essner *et al.*, 2005). This is depicted diagrammatically in Figure 3.1

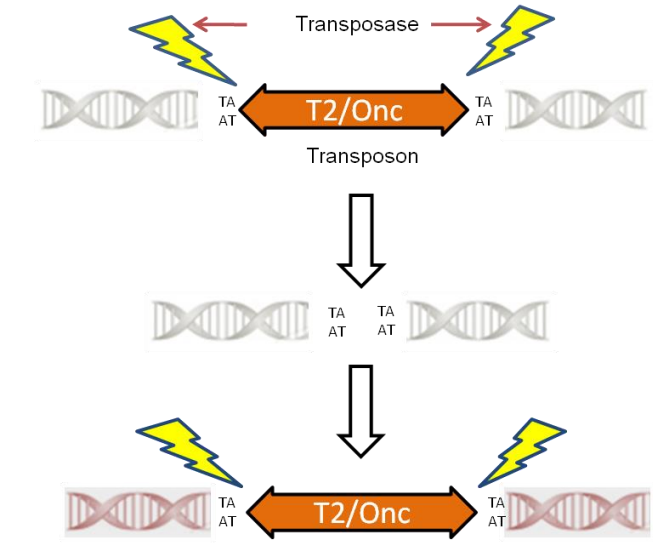

Figure 3.1. The Sleeping Beauty transposon system. This is composed of two functional parts, the transposase and transposon vector. When both are present in the cell nucleus cut and paste transposition can take place.

A major advantage of the SB system for discovery of cancer genes compared to candidate gene approaches is that it allows the investigator to perform whole genome scans that are devoid of subjectivity (Collier and Largaespada, 2007a). Whereas the use of chemicals and radiation to induce mutagenesis is effective in causing mutations, identification of the candidate genes is often wrought with many difficulties. Identification using the SB system is much easier since the mutagen acts like a molecular tag allowing the easy detection of cancer genes through subsequent molecular analyses (Collier and Largaespada, 2007b).

The ease of mutation detection is borne out by the results of landmark studies carried out by Dupuy *et al.*, (2005) and Collier *et al.*, (2005). Briefly, Collier *et al.*, (2005) introduced T2/Onc transposons with a CAGGS-SB10 transposase in a background already predisposed to cancer (*p19Arf*+/-). In total, 28 sarcomas were generated, and PCR amplification of insertion sites followed by sequence analysis identified mutations within the *Braf* gene in 22 of these tumours. Missense mutations in *Braf* have previously been reported in several human primary sarcomas and a sarcoma cell line. In contrast, Dupuy *et al,* (2005) examined the impact of a modified transposon (T2/Onc2) with a more active transposase (Rosa26-SB11) in a wild type background. Hematological malignancies, such as T-cell leukemia and lymphoma were generated at high frequency. *Notch1* was identified as an important candidate gene from this analysis. An important conclusion from these and other insertional mutagenesis studies was that the malignancies are primarily occasioned by the delivery of powerful promoters and enhancers quantitatively altering transcription, and do not occur solely as a result of insertional gene disruption.

Another factor which has a significant effect on the ability of the SB transposons to cause mutagenesis is their insertion site preference. Unlike retroviruses, lentiviruses and adenoviruses, which have a bias for promoter and transcriptional sites, SB transposons have been observed to integrate in a relatively indiscriminate manner when delivered by plasmid DNA. Other studies suggest that SB transposons do have some insertion site bias, with a tendency to integrate in regions that are rich in thymine and adenine (TA) sequences, and in places where the DNA is twisted or warped (Clark *et al.*, 2004;Liu *et al.*, 2004;Dupuy *et al.*, 2005). However, when transgenic transposon arrays are used for mutagenesis, high levels of local hopping events are observed, with 20%-50% of insertions mapping to within 40Mb of the array (Collier *et al.*, 2005;Dupuy *et al.*, 2005). Despite this well documented non-random insertion of a large subset of inserts, SB has been successfully applied to a wide variety of solid tumours.

3.1.2 The Sleeping Beauty Mutagenesis System

The SB mutagenesis system comprises two elements: The mutagen in the form of a tranpositionally competent but non coding transposon, and the transposase source which is encoded within a transpositionally incompetent element.

3.1.2.1 SB Transposons: T2Onc and its derivatives

Different types of SB transposons have been engineered for cancer gene discovery. *Collier et al.,* (2005) used the T2/Onc element to study transposon-mediated mutagenesis. The T2/Onc SB transposon contains a long terminal repeat of the murine stem cell virus (MSCV). The MCSV has powerful promoters and enhancers which play a dominant role in activities of stem cells. The MCSV is surrounded on the downstream site by a poly-A tail and this tail is delimited by a splice acceptor. Upstream of the MCSV long terminal repeat is a splice donor that causes over expression if it is located adjacent to or within a gene. Next to this splice donor site is a poly-A tail which is then flanked by a splice acceptor (Figure 3.2; Collier *et al.,* 2005;Largaespada, 2009). The function of the poly-A tail is to precipitate premature transcript termination when it is inserted into a gene. This happens regardless of the orientation in which the transposon is placed since the poly-A tails at either ends are inverted. The splice acceptor-poly-A tails are used to interrupt the function of tumor suppressor genes (Collier and Largaespada, 2005). Several transgenic lines containing arrays of the T2/Onc element have been generated to act as transposon sources, including a 25 concatemer line (76) where the transgenic array maps to mouse chromosome 1, and line 68 that has the same concatemer number mapped to chromosome 15 (Collier *et al.,* 2005).

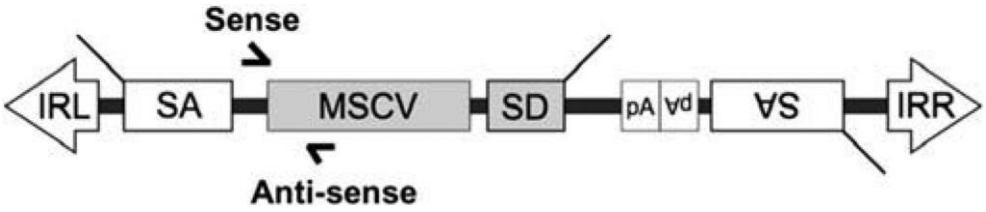

Figure 3.2 The T2/Onc transposon. The MSCV 5' LTR can enhance transcription of downstream exons picked up by the splice donor (SD). Transcript disruption can be caused by insertion in either orientation due to the presence of two splice acceptors (SA) (Largaespada, 2009).

Another type of SB transposon which has been used in mutagenesis studies is the T2/Onc2 element developed by Dupuy *et al*., (2005). The T2/Onc2 system resembles the T2/Onc in many respects. The only differences between the 2 are that one of the splice acceptors in the T2/Onc2 system is much longer, and the element is slightly shorter which may enhance transposition efficiency. Perhaps more importantly, the number of transposon copies found in the chromosome concatemer of the T2/Onc2 transgenic line used in these analyses has approximately 6-14 times more copies than the T2/Onc lines used by Collier *et al.,* (2005). It is thus referred to as a high copy line while the T2/Onc lines are termed as a low copy line. However, an increase in embryonic lethality rate has been recorded when the high copy number line is used (Dupuy *et al*., 2005;Collier and Largaespada, 2007b).

More recently, a third Sleeping Beauty transposon concatemer has been derived; T2/Onc3. In this advanced version, the murine stem cell virus (MSCV) promoter has been replaced with chicken β-actin (CAG) promoter. This harbours a hybrid cytomegalovirus enhancer / beta-actin promoter that is active in a wider range of tissues than the MSCV promoter (Dupuy *et al.,* 2009). As a result, this transposon has been able to generate more than 20 different tumor types in mice, suggesting that it can be widely applied to understand the underlying molecular basis of a wide variety of cancers.

3.1.2.2 SB Transposases: the CAGGS-SB10 and Rosa26SB-11 Transgenic Lines

To date, two transgenic lines that can be used as sources of transposase have been developed. These are CAGG-SB10 and Rosa26-SB11 (Collier and Largaespada, 2007). CAGGS-SB10 mice were developed using the conventional method of pronuclear injection (Okabe, 1997) and were conceived to produce the transposase in all tissues. However, it has been observed that high levels of this enzyme are expressed only in a limited number of tissues including connective tissue cells (CT), mesenchymal and male germ line tissues. Studies also show that it exhibits mosaicism in many tissues (Dupuy *et al.*, 2001;Largaespada, 2009).

The Rosa26-SB11 transgenic line was created by knocking in the SB transposase into the Rosa 26 loci of ES cells by use of homologous recombination (Colllier and Largaespada, 2005). Unlike the CAGGS-SB10 system, the Rosa26-SB11 line produces transposase in all tissues, but when used with the T2/Onc2 system it

causes both a high degree of embryonic lehality, as well as a variety of tumours including leukaemias, sarcomas, Gastrointestinal Tract (GIT) tumours, and Hepatocellular carcinoma in almost all animals which survive to term (Collier *et al.*, 2005; Dupuy *et al.*, 2005; Keng, *et al.*, 2009; Largaespada, 2009; Starr *et al.*, 2009). To overcome this drawback, some studies have utilised tissue specific promoters to regulate the transposase (Largaespada, 2009). Collier and Largaespada (2007) have found lower transposition rates when the Rosa-SB11 lines are used compared to the CAGGS-SB10 lines. This was attributed to the high level of expression of the transposase enzyme in the former, a factor that has previously been associated with a reduction in the efficiency of transposition (Geurts *et al.*, 2003).

3.3 SB mutagenesis of *Ptch1*+/- mice.

To identify genes which interact with the predisposition to MB in the *Ptch1*+/- transgenic mouse model, an SB mutagenesis breeding program was initiated in the Institute of Genetic Medicine, Newcastle University, in 2007 utilising both the SB10 and SB11 transposases in combination with T2/Onc to mutagenise mice heterozygous for the ortholog of the *PTCH1* gene *(Ptch1+/-)*. Initial experiments using SB10 failed to generate tumours in either a *Ptch1*+/- or wild type background, despite the fact that tests indicated that transposition was occurring in cerebellar tissue (M. Lastowska, personal communication). Previous studies showed that the combination of *T2/Onc* with *SB10* resulted in tumour production, predominantly sarcoma (Collier *et al.*, 2005), suggesting that this system has tissue specificity. Subsequent experiments in Newcastle, utilising SB11 Rosa26 transposases in combination with T2/Onc line 76 to mutagenise mice heterozygous for the ortholog of the gene (*Ptch1+/*), did appear to impact upon the frequency of medulloblastomas in *Ptch1*+/- animals and generated haematological tumours such as lymphoma and leukaemia in both *Ptch1*+/- and wild type backgrounds. In addition, rhabdomyosarcomas and liver adenomas were also observed. This combination of transposase and transposon was therefore chosen to perform a full mutagenesis screen of the *Ptch1*+/- model, as outlined below.

3.1.3 SB11 mutagenesis breeding scheme

All mice strains were initially backcrossed to the CBA wild type strain for a minimum of 3 generations to minimise background affects known to influence tumour pentrance in the *Ptch1*+/- model (Goodrich *et al.*, 1997). Mice homozygous for T2/Onc and heterozygous for the *Ptch1* null allele were then mated to mice heterozygous for the

SB11 transposase (Figure 3.3). This cross generated 4 genotypes, 3 of which were used for subsequent analysis: One genotype is both predisposed to cancer and subject to experimental mutagenesis (*Ptch1*+/-;T2/Onc+/-;SB11+/-), the second is a predisposition only control genotype (*Ptch1*+/-;T2/Onc+/-), and the third is a transposition only control genotype (T2/Onc+/-;SB11+/-).

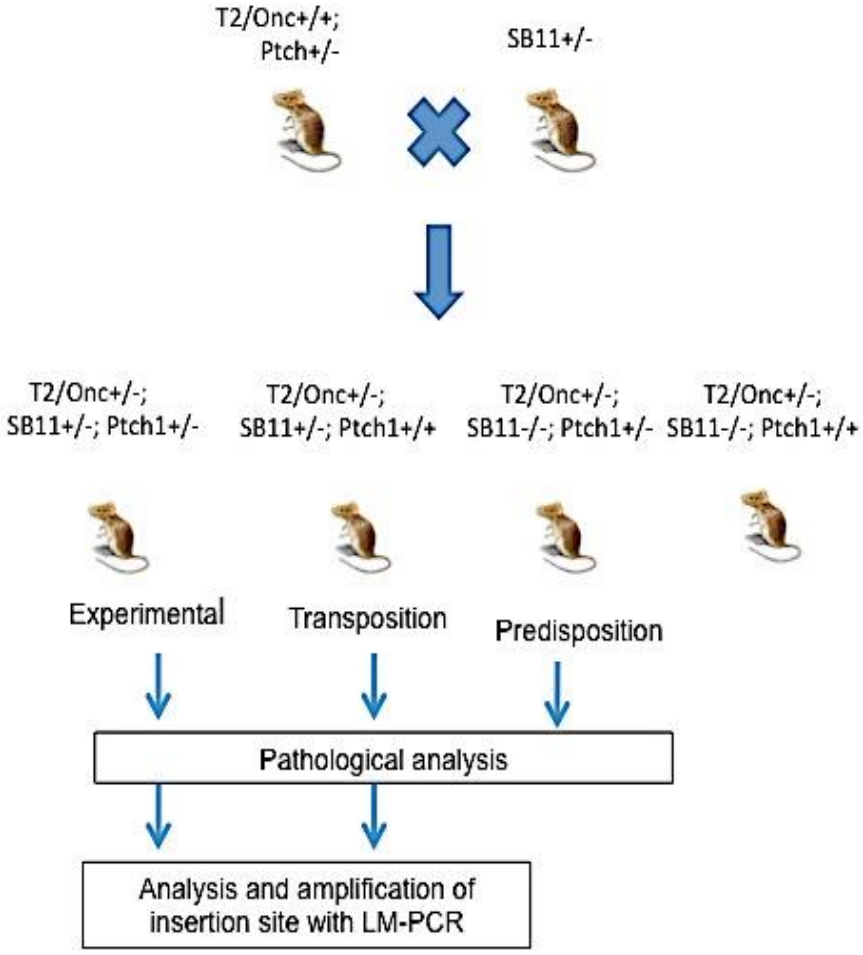

Figure 3.3: Animal breeding, genotyping and sample analysis. The experimental animals (T2/Onc+/-;SB11+/-;*Ptch1*+/-), transposition controls (T2/Onc+/-;SB11+/-;*Ptch1*-/-) and predisposition controls (T2/Onc+/-;SB11-/-;*Ptch1*+/-) are analysed.

In total, 243 experimental and 195 controls mice were aged for approximately 15 months, with tumours arising in mice of all 3 genotypes being retained for pathological and immunohistochemical analysis. In addition, tumours arising in the two genotypes where transposition has occurred, together with normal tissues from transposition only controls, were also retained for insertion site analysis.

3.1.4 Initial aims of the project

When I became involved in this project, the SB11 breeding scheme had already been underway for over 18 months, and the phenotypic analysis of tumours was being performed. The initial aims were therefore to complete the mutagenesis scheme in collaboration with those already involved in the project, and perform the intial analysis to identify lists of potential candidate genes. Specifically, the initial aims were to:

1. Complete the genotyping of all remaining mice within the cohort.
2. Use genotype and tumour incidence data collected to establish the impact of SB mutagenesis upon survival and frequency of tumour formation.
3. Use linker mediated-PCR to amplify across the T2/Onc insertion sites in all tumours and control tissues.
4. Sequence at high depth all insertion sites amplified.
5. Apply statistical analysis to the sequence data to identify common insertion sites (CISs) within the tumours generated.

3.2 Results

3.2.1 Genotyping of *Ptch1, SB11 and T2/Onc* in experimental mice

The majority of the breeding programme had been completed at the outset of this project by Dr Michael Jackson and Dr Maria Lastowska. However, a small cohort of approximately 200 animals remained to be analysed. To establish the genotype of these mice, *Ptch1* and SB11 status was established using ear punch DNA as template for established Polymerase Chain Reaction assays (see Methods). As the design of the breeding scheme generated animals heterozygous for *T2/Onc*+/-, this transgene was only assayed to confirm the genotype of breeding pairs. Examples of results obtained for *SB11* and *Ptch1* are shown in Figure 3.4. Animals 407-1 and 407-2 are positive for *Ptch1*+/- (~460bp amplicon, Figure 3.4A), while animals 407-2 and 411-1 are positive for SB11 (~370bp amplicon, Figure 3.4B). From these results, animals which carry the T2/Onc transgene array are easily categorised. For instance, animal 407-1 can be categorised as predisposition only control (*Ptch1*+/- only). However, animal 407-2 can be categorized as the experimental genotype (*Ptch1*+/-;SB11+/-). Moreover, animal 411-1 can be identified as a transposition only control (SB11+/-). Finally, animal 411-2 can be categorised as wild type and omitted from future analyses.

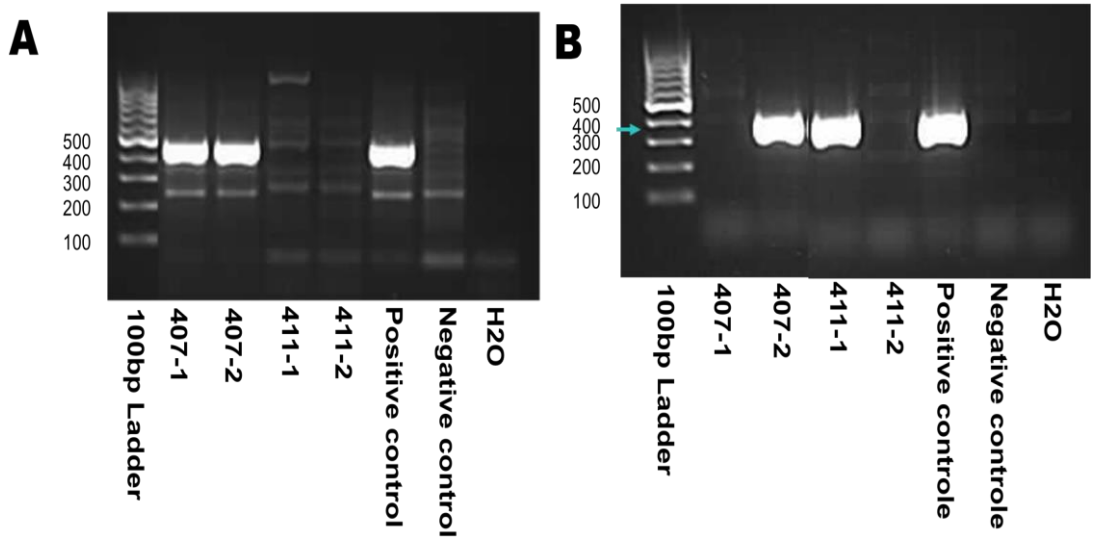

Figure 3.4: *Ptch1 and SB11* PCR genotyping assay. Ear punch DNA was used for genotyping (A) *Ptch1*; expected amplicon ~460bp, arrowed. (B) SB11; expected amplicons ~370bp, arrowed. Positive controls are samples known to carry each transgene. Negative controls are wild type. H2O = no template.

In addition to genotyping the individual genes which are segregating, an additional check for transposition was also performed in a subset of animal to confirm transposition, and to rule out genotyping errors in tumours arising in predisposition only control animals. The excision assay (Collier *et al.*, 2005) is able to identify complete, partial or even no excision. Figure 3.5 exemplifies the excision essay within rhabdomyosarcoma samples. This figure shows short (~210bp) amplicons generated from samples PM29, PM96, and PTS11-481, which indicates total excision of the transposons within the array. The larger amplicon observed in samples PM25, PTS11-165, PTS11-215 and PTS11-432 (>600bp) indicates that no transposon excision events have occurred. However, sample PTS11-324 shows both small and large bands suggesting that some but not all of the transposons have excised from the array.

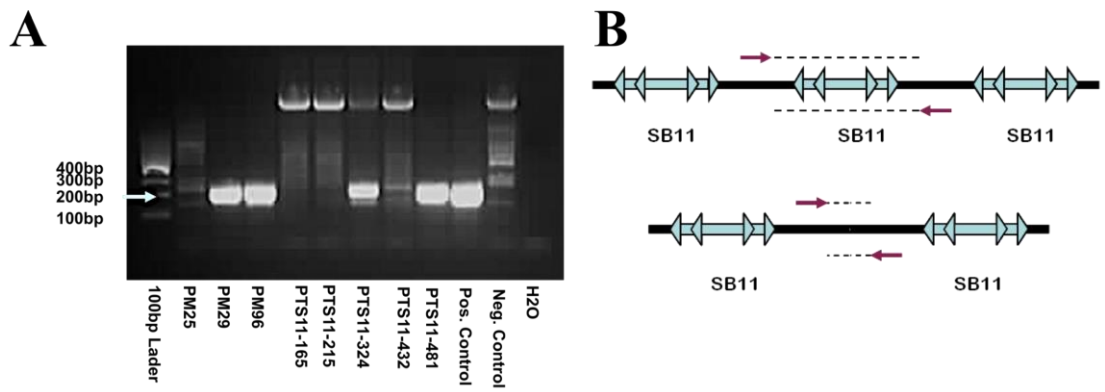

Figure 3.5: Example of Excision essay in rhabdomyosarchoma samples. (A) PCR results showing the products size and different levels of excision. The arrow represents the expected size ~210bp of the excised transposon. Positive control shows complete excision. H2O = no template. 100bp ladder is on the extreme right of the gel. (B) Schematic diagram showing transposons and the relative amplicon structures/sizes before and after excision.

In total, ~200 animals were genotyped and/or checked for transposition to finalise the tumour dataset for subsequent analysis.

3.2.2 Tumour types observed in the mutagenized *Ptch1*+/-;T2/Onc;SB11 mice

A total of 243 mutagenised experimental mice and 195 transposition and predisposition control mice were aged for up to 15 months and were monitored for tumour formation. The total number of tumours generated in animals of each genotype are summarised in Table 3.1. Examples of the most frequent tumour types observed in the mutagenised Ptch mice are also shown in Figure 3.6. The pathology analyses were performed by Dr Chris Bacon and Professor Alastair Burt (Faculty of Medical Sciences, Newcastle University). Medulloblastoma lesions developed exophytically in approximately 23% of animals aged for >6 months (Figure 3.6 panels A, B). Macroscopically, an absence of meningeal masses has been reported in these affected animals. However, microscopically, these tumours were classified as desmoplastic forms, a typical form of human SHH-subgroup of Medulloblastoma. Collectively, these features presented them as a typical of *Ptch1*+/- model. Furthermore, MB like lesions were also observed within predisposition only controls *(Ptch1*+/-*;T2/Onc*+/-) but at a low frequency (6%).

Slow growing rhabdomyosarcoma (RMS), the other form of tumour that is usually generated as a result of *Ptch1*+/- heterozygosity, arose in approximately 22% of mutagenized animals (Figure 3.6 panels C, D). Immunohistochemical staining analysis showed small round mitotical active cells and rhabdomyoblasts and large eosinophilic differentiated cells. Both the MBs and RMSs were morphologically similar to those

obtained from non-mutagenised *Ptch1*+/- mice and other murine MB models (Johnson *et al.*, 1996;Marino *et al.*, 2000;Fults, 2005;Hatton *et al.*, 2008;Mancuso *et al.*, 2008).

In addition, multiple liver adenomas were generated in 19 mutagenised *Ptch1*+/- mice compared to none within predisposition controls. Immunohistochemically, these lesions were usually characterised by leukaemic infiltration and usually arose in animals who also had haematological tumours or rhabdomyosarcomas (Figure 3.3 E-G). It is noteworthy that the morphological features of these adenomas are similar to those observed and analysed in a previous SB screening study within mutagenized mice with a conditional SB11 for hepatocellular carcinoma (Keng *et al.*, 2009).

Finally, approximately 28% of mice succumbed to haematological malignancies that presented either as lymphoblastic leukaemia (presenting as hepatosplenomegaly) or large thoracic (probably thymic) lymphomas (data not shown). These haematological phenotypes were also observed in transposition only control mice at similar frequency, indicating that they can be induced by SB mutagenesis only. This is consistent with data from Collier *et al.*, (2009) who first used the SB11 transposase in combination with T2/Onc.

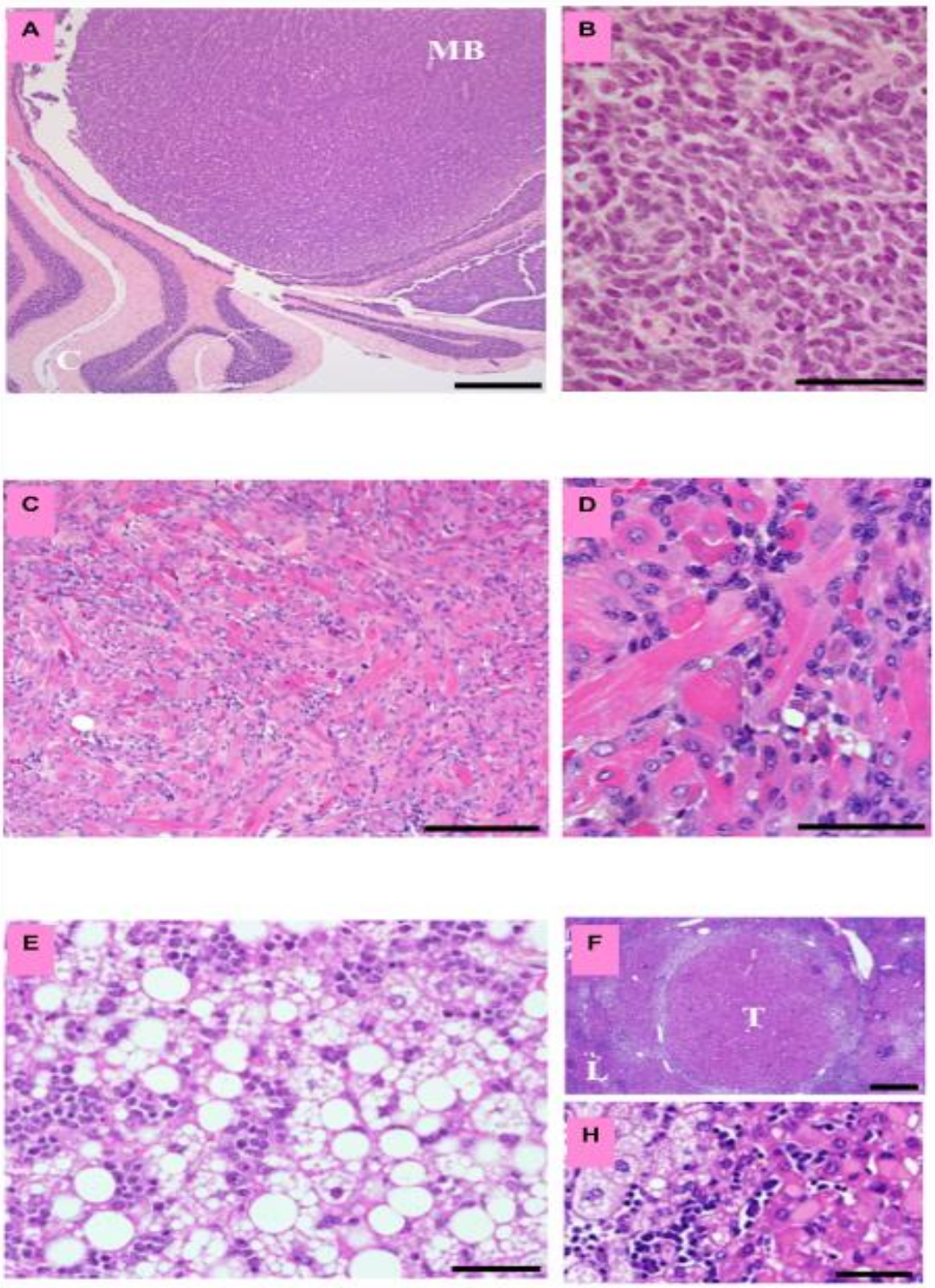

Figure 3.6: Main tumour types obtained from the mutagenised Ptch+/- mice: A. a large prominent MB originating from the cerebellum; B and C. immunohistochemical staining showing a large exophytic MB which grew from the cerebellum. D. a large rhabdomyossarcoma on the right extremity. E and F. immunohistochemical appearance showing small round cells mitotical active and rhabdomyoblast and large eosinophilic differentiation. G and H. a typical haematological malignancies that appears in form of hepatosplenomegaly and large thoracic lymphoma lesion, respectively. I. microscopic feature representing hepatocellular carcinoma with leukaemic infelteration. J and K. microscopic appearance of immunohistchemical staining from adenomatous lesion with hepatocelular carcinoma, micro and macrovacular steatosis that often associated with leukaemic infiltration (Lastowska *et al.*, 2013).

Table 3.1: Primary tumours obtained by genotype

	Genotype					
	Experimental		**Predisposition**		**Transposition**	
	No.	**Median (range)**	**No.**	**Median (range)**	**No.**	**Median (range)**
Cohort Number	243	-	95	-	120	-
Medulloblastoma	42	17.3 (7 - 27)	4	19 (15 - 28)	0	-
Other Brain lesions	5	29.2 (23 - 38)	0	-	4	19.5 (12 - 25)
Haematological	51	27.9 (8 - 52)	0	-	25	34 (13 - 54)
Rhabdomyosarcoma	33	21.3 (14 - 36)	29	25.5 (8 - 49)	0	-
Liver Adenomas	19	29 (20 - 45)	0	-	2	-
Other Tumours	2	-	1	-	5	-
All Tumours	152	-	34	-	36	-
Hydrocephalus	9	-	0	-	1	-

Median age of onset and range of onset is given for common tumour types. Genotypes: Experimental – *Ptch*+/-;SB11+/-;T2Onc+/-. Predisposition - *Ptch*+/-;T2Onc+/-. Transposition - SB11+/-;T2Onc+/-. [1]Parenchymal brain lesions included one with pseudopalisading necrosis consistent with glioblastoma multiforme. [2]Morphological and immunohistochemical analysis established that ~50% of haematological malignancies were precursor T-cell lymphoblastic lymphomas/leukaemias, ~5% were confirmed myeloid neoplasms, and the remainder were poorly differentiated haematological neoplasms, mostly of probable myeloid lineage. [3]RMSs developed most frequently on the hindquarters and lower limbs. [4]Adenomas were mildly dysplastic with frequent intralesional steatosis, although sufficient atypia and eosinophilic cytoplasmic inclusions were observed in 2 nodules to warrant classification as hepatocellular carcinoma. All but 2 were identified in animals where other malignancies were also present (4 RMS, 10 Haematological, 1 with RMS+Haematological, and 2 unclassified). All animals with liver adenomas were male, a highly significant bias ($p<0.0001$). [5]Hydrocephalus has previously been reported in this model (Goodrich *et al.*, 1997)

3.2.3 Impact of predisposition and mutagenesis upon survival

After genotyping was completed, the data was integrated with survival information to establish if mutagenesis alone is capable of inducing tumours, and if mutagenesis interacts with the tumour predispositions of the *Ptch1* allele. Kaplan Maier survival curve comparing experimental and control genotypes were analysed using the Log rank test; these analysis were performed by Dr Mauro Santibanez-Koref and Dr Maria Lastowska (Institute of Genetic Medicine, Newcastle University). These are shown in figure 3.7. The results indicate that the mortality rate in experimental animals was significantly higher (90%) than the predisposition or transposition only control groups (Figure 3.7A). Clearly, *SB11* induced mutagenesis may interact with *Ptch1* in all tumour types, although the impact is most clearly seen for medulloblastoma where approximately 23% of *Ptch1*+/-;T2/Onc;SB11 animals developed MB compared to less than 6% of predisposition controls ($p<0.0001$ log rank test), with no tumours in the transposition only control (Figure 3.7B). This confirms that the transposition enhanced the tumour predisposition and that further analysis of the insertional mutation should be informative. The frequency of rhabdomyosarcomas in the experimental and predisposition controls was similar, and the Kaplan Meier curve suggests no impact of transposition upon the frequency of these malignancies ($p=0.31$ log rank test, Figure 3.7C). Finally, the impact of haematological malignancies upon survival was not significantly different between experimental or transposition controls (data not shown) and suggests that transposition alone induced these malignancies.

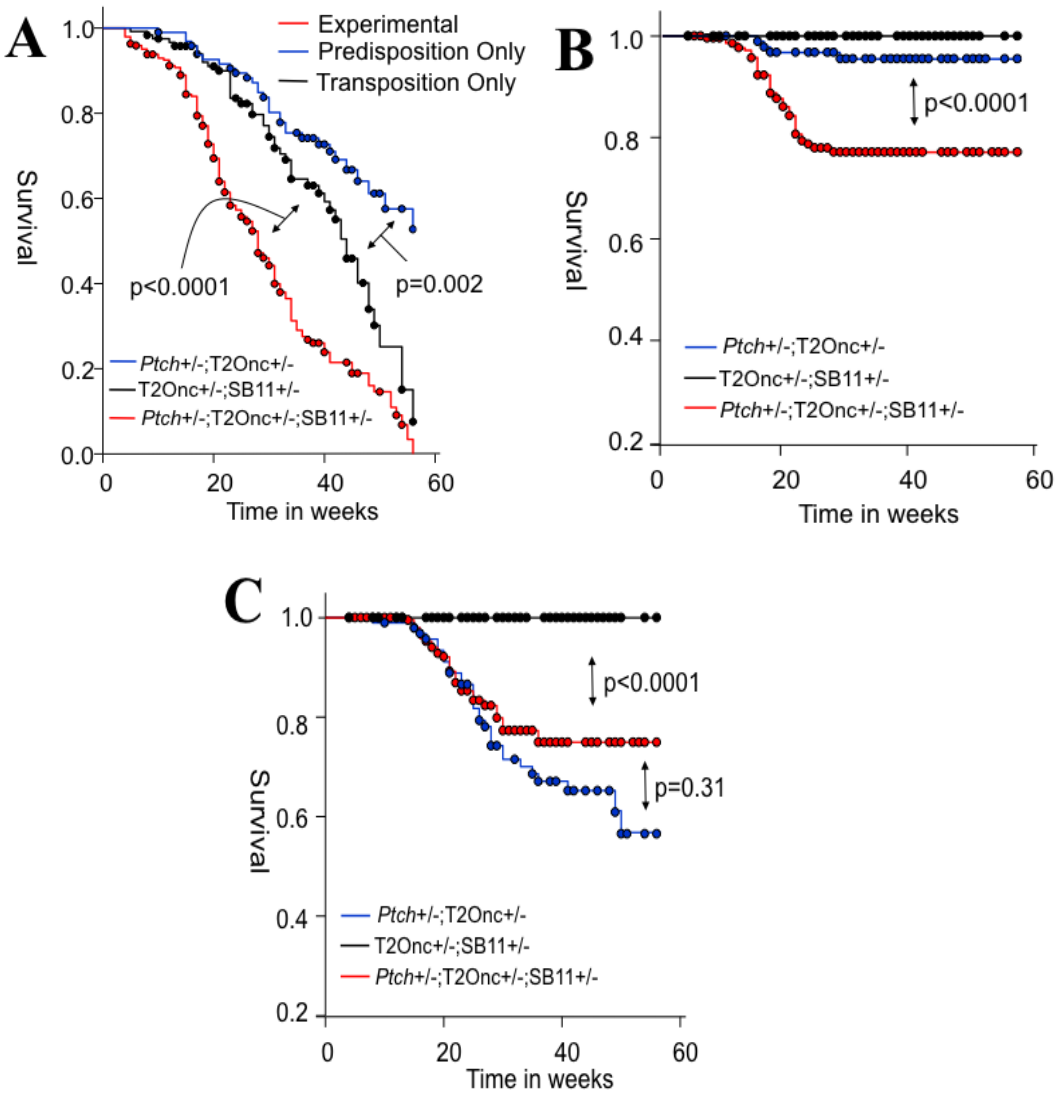

Figure 3.7: Kaplan Mier survival curve of SB mutagenised mice and controls. A. Mortality is significantly higher within transposition only than predisposion only controls (p<0.002), and was significantly higher in experimental than both controls groups (p<0.0001). B. Mortality due to MB alone is significantly higher in the experimental cohort compared to both control groups (p<0.0001), however, mortality between both controls was not significant (p=0.118). C. Mortality due to rhabdomyosarcoma malignancies was not significantly different between experimental and predisposition cohorts (p=0.31). The log rank test was used in all analysis.

3.2.4 LM-PCR amplifications of T2/Onc insertion sites

A total of 219 tumours from animals subjected to SB11 mutagenesis were harvested for linker mediated-PCR (LM-PCR) amplification (see Table 3.2). These consisted of 41 medulloblastomas (MB), 34 Rhabdomyosarcomas (RMS), and 123 haematological malignancies manifesting as either leukaemic spleens with enlarged liver (LS) or thoracic Lymphoma (Lym) plus some other tumours generated by this mouse model (see Table 3.1). The haematological malignancies included 32 generated

during an independent analysis of mice positive for the human *MYCN* oncogene (M. Jackson and M. Lastowska unpublished, (Schwab *et al.*, 1983). This was looking for interaction between SB11 and the predisposition of mice with the *MYCN* transgene to develop neuroblastoma; none was observed (M. Jackson unpublished). In addition, approximately 30-35 thymus, spleen, cerebellum and muscle samples from SB11+/-; T2Onc+/- transposition only mice were harvested at 6-7 weeks as to enable the distribution of insertion sites to be analysed in tissues relevant to the tumours under study. In total, therefore, the dataset for LM-PCR and insertion site sequencing comprised 230 tumours and >120 controls.

Table 3.2: Number of samples processed and sequenced; tumours and controls.

Samples	Total	Extracted and analysed by myself	LM-PCR – Left	LM-PCR - Right
Spleen control	35	35	35	35
Thymus control	35	10	10	10
Muscle control	35	35	35	35
Cerebellum control	35	10	10	10
Haematological	123	84	84	84
Rhabdomyosarcoma	34	34	34	34
Medulloblastoma	41	20	20	20

The full LM-PCR and sequencing pipeline is described in the Methodology section. DNA from all samples underwent automated extraction following overnight digestion with Proteinase K, and an example of the DNA produced is shown in Figure 3.8. All DNAs from leukemic samples (PTS11-348, PTS11-263 and PTS11-70a) are of relatively high molecular weight, migrating at a size of between 2 and 23kb. Lambda-*Hin*dIII and a 100bp ladder (Fermentas) were used as molecular weight markers.

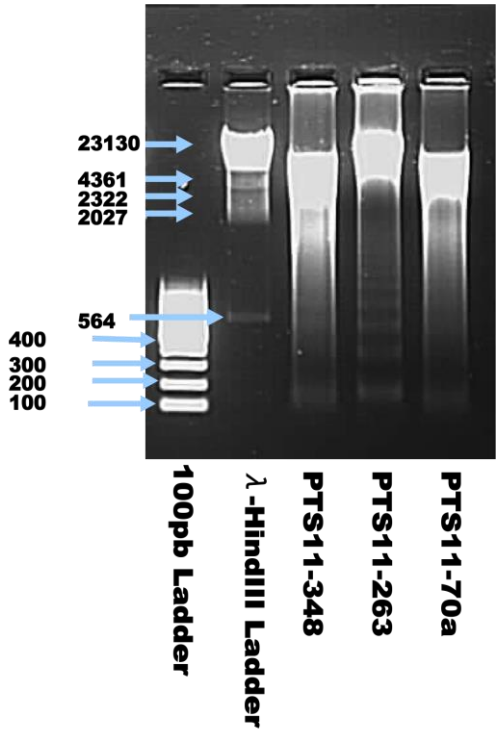

Figure 3.8: Example of DNA quality: Spleen tumour samples of undigested DNA show very high molecular weight (23130bp). Lamda-HindIII was used as a marker as well as 100bp ladder.

To define the insertion sites of T2/Onc elements, splinkerette PCR (Collier *et al.*, 2005) was used to amplify both the left and right end of T2/Onc elements in all tumours and controls. This involved DNA extraction of 359 samples (including both tumours and controls) followed by digestion with *Bfa*1 (left) or *Nla*III (right). Typical products of *Bfa*1 and *Nla*III digestion are shown in Figure 3.9. All samples gave a smear of digestion products. The *Bfa*I digests produce a larger average size than *Nla*III, as *Bfa*I cuts the DNA less frequently than *Nla*III.

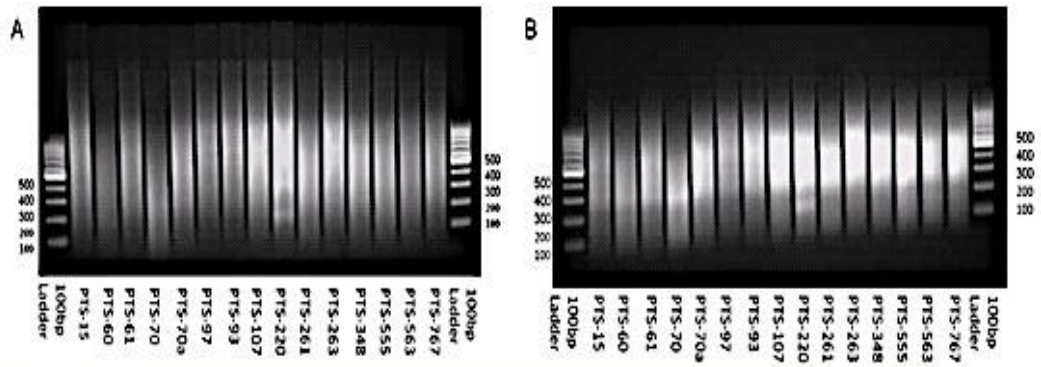

Figure 3.9: A typical DNA digestion with *Bfa1* and *NlaIII*: Results for 15 enlarged spleen are shown (A) *Bfa1* (left side of *T2/Onc*) and (B) *NlaIII* (right side of *T2/Onc*) digestion. This pattern is typical of DNA after digestion. Marker is a 100bp ladder.

To digest any untransposed *T2Onc* elements within the transgenic array and prevent these generating amplicons in subsequent PCRs, linkers were ligated and a second digestion was performed, using *Bam*H1 and *Xho*I for *Bfa*1 and *Nla*III digested samples respectively. These were then subject to primary PCR to specifically amplify from inside integrated T2/Onc elements out to the nearest *Bfa*1 or *Nla*III site within surrounding genomic DNA (see Methods). Although undigested DNA acted as a control for *Xho*I and *Bam*HI digestion efficiency, the ligation and first PCR were only checked if secondary PCR was unsuccessful (see Methods).

Examples of secondary PCR products generated by the 2nd nested PCR reaction, which eliminates non specific amplification products and incorporates the barcode and sequencing primer, are shown in Figure 3.10 and Figure 3.11 for *Bfa*I and *Nla*III respectively. *Bfa*I secondary PCR generates a sample specific ladder of amplicons, consistent with the presence of a different spectrum of integration sites in each tumour. The majority of the amplicons range in size from 200-1000bp, an appropriate size for FLX amplification and sequencing. As the T2/Onc left hand priming site lies approximately 120bp from the terminus of the transposon, the minimum possible size of an intact amplicon is approximately 180bp (120bp + 60bp of primers). Amplicons smaller than 180bp are likely to be primer dimers, or ligation artefacts. It is also clear from Figure 3.10 that two ligations, MTS11-224ST and PTS11-93ST, failed to yield products. Subsequent checks on excision status in such cases (see methods) were used to establish whether this represented experimental failure, or no excision.

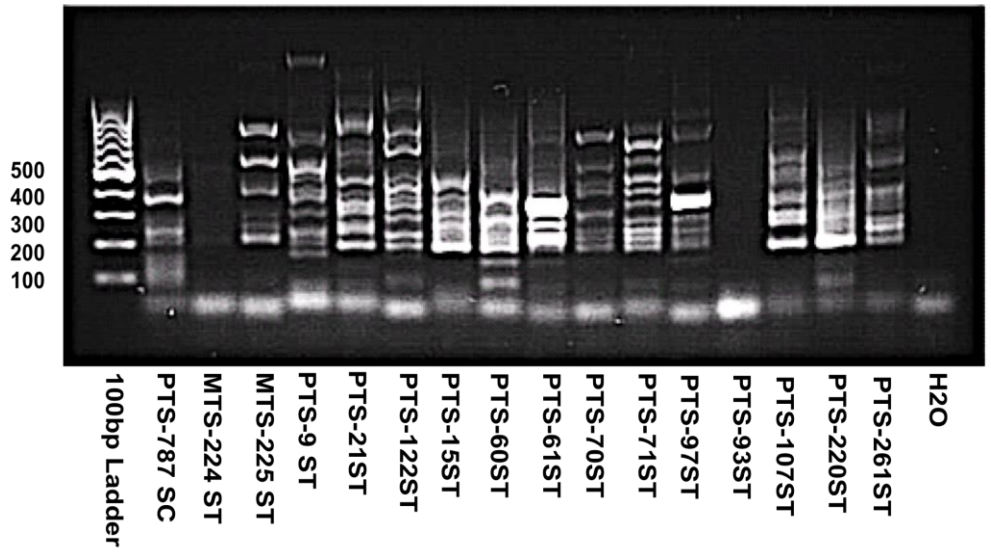

Figure 3.10: Amplification of *Bfa*I products. Mixed tumours samples *Bfa*I digested and subjected to PCR. Amplicons range from 200 – >1000bp. Two samples failed to produce products (MTS11-224ST and PTS11-93ST). A 100bp Ladder is on the left end lane. H2O = no template.

The results for secondary PCR of *Nla*III digested DNA is shown in Figure 3.11. The results are essentially identical to the *Bfa*I results, except that the average size of the PCR products is smaller, and appear to be more complex (smeared). This is consistent with results from other laboratories, and is due both to the right hand primer being closer to the terminus of the T2/Onc transposon (30bp), giving a minimum amplicon size of about 90bp, and the higher frequency of *Nla*III restriction sites within mouse genomic DNA resulting in a larger number of inserts being recoverable by this method.

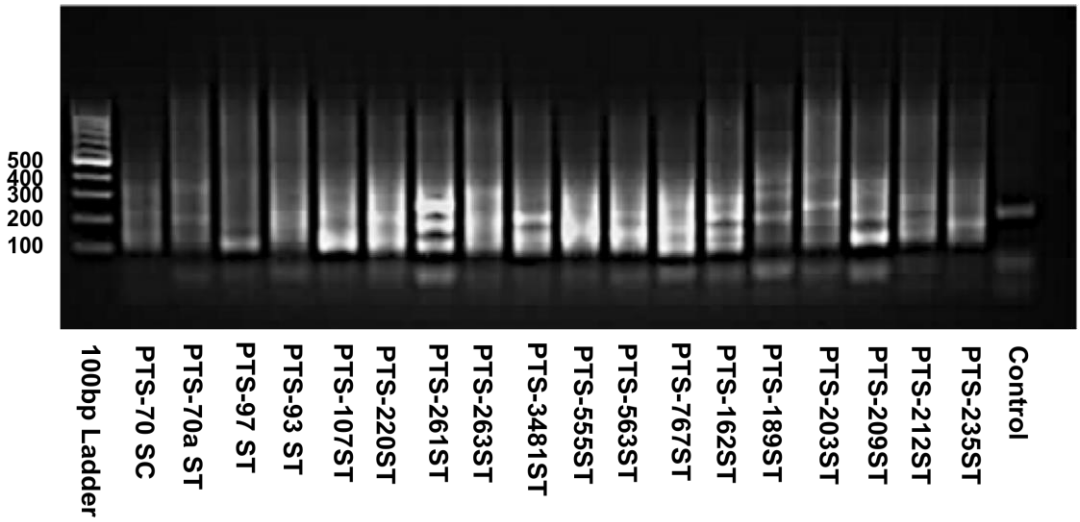

Figure 3.11: Amplification of *NlaIII* products. Mixed tumours samples *Nla*III digested and subjected to PCR. Amplicon sizes range from ~90bp->1000bp. A 100bp Ladder is on the left end lane. H2O = no template. is on the extreme right lane as a negative control. *Nla*III digested T2Onc+/- DNA is shown as a negative control on the extreme right lane; the single band from the transgenic array is clearly visible.

3.2.5 Identification of T2Onc insertion sites by Pyrosequencing

In total, 8 DNA sequencing runs were implemented; 3 FLX-GS20 runs were performed with a maximum of 48 samples per GS20 lane, and 5 FLX-Titanium runs with a maximum of 120 samples per lane. This produced a total of 265,793 raw sequencing reads from 359 tumour and control PCR reactions, with an average number of reads per sample of 1433 (left) and 778 (right).

Generally, the reads lengths obtained from processed samples (tumours and controls) appeared to be distributed evenly. An example of the output from a single FLX run is summarized in Figures 3.12 and 3.13. The distribution of read lengths obtained is shown in Figure 3.12, and it is clear that there is a wide range of read lengths from <100bp to over 400bp. The prominent peaks between ~50–80bp are likely to be primer dimers and ligation artifacts. Additionally, the very prominent peak around ~170bp is known to represent amplicons from the T2/Onc concatemer, suggesting that the second enzyme digestion was not complete. However, these observations are congruent with the gel pictures of secondary PCR of both *Bfa*I and *Nla*III (see Figure 3.10 and 3.11, respectively).

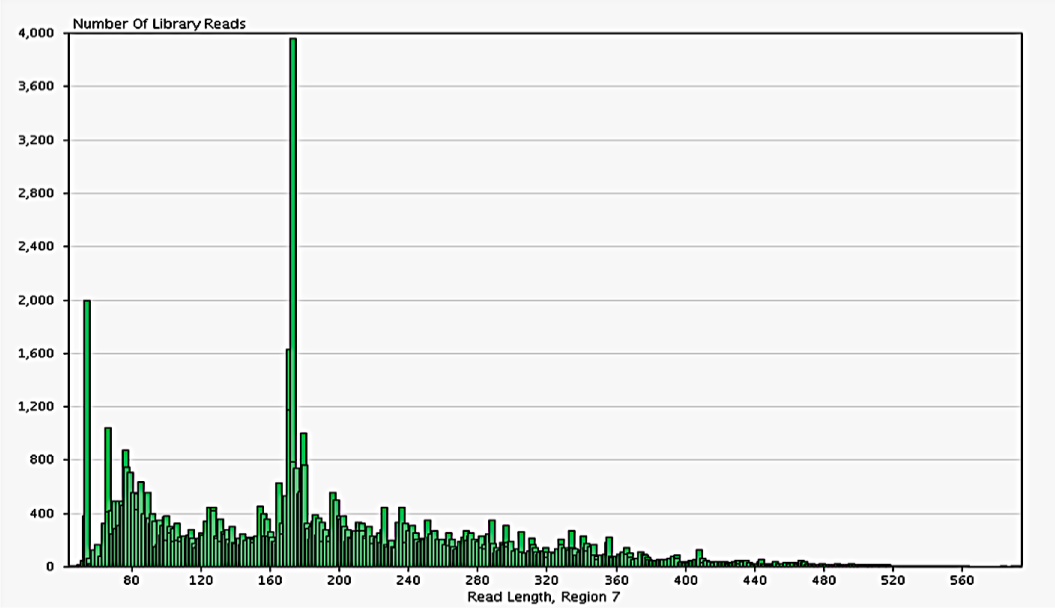

Figure 3.12: FLX pyrosequencing read length distribution. Read lengths from 1 FLX run are shown. The prominent peak at ~170bp is known to be a product from the T2Onc concatemer. The peaks at ~50 – 80bp are primer dimers.

To illustrate sample coverage in more detail, the number of reads for *BfaI* (left) side reactions of both tumour and control samples from this run are shown in Figure 3.13. Most of the samples give read numbers between ~200 and 1000, significantly more than obtained using subcloning methods (Collier *et al.*, 2005, Dupuy *et al.*, 2005). However, two control samples (surrounded by the red rectangle) produced a very low number of reads (<~180bp), and several give read numbers of over 1000 (e.g. PTS11-791). Samples with low read numbers were re-processed. As the number of individual amplicons in each sample was low relative to read number (e.g. see Figure 3.10) no normalization of read numbers between samples was performed prior to downstream processing.

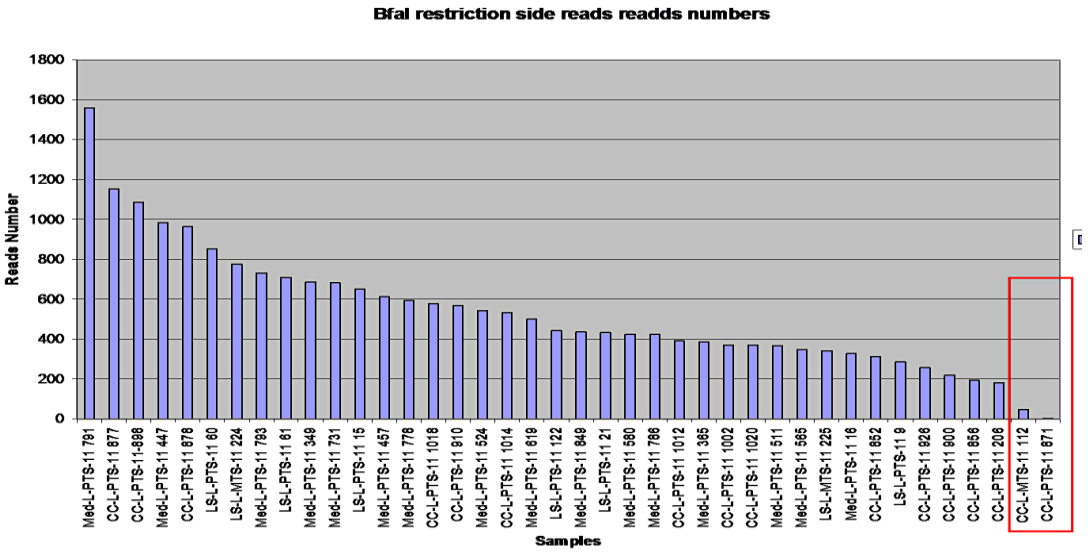

Figure 3.13: Example of read numbers recovered from *Bfa*1 digestion. Number of reads obtained by FLX pyerosequencing for subsets of SB induced tumours and Controls from *Bfa1* digestion are shown. The minimum target number of reads was 200 – 1000. Two controls samples (highlighted in red) produced a lower number of reads.

3.2.6 Manual assessment of sequence read consistency with SB system

To confirm the quality of the sequence reads prior to analysis, and to establish if the insertion site distribution is consistent with what is known about SB mobilisation, the insertion sites of transposons identified from both *BfaI* and *NlaIII* products for 1 spleen (leukaemic) tumour (PTS11-21 - run 3) was analysed in detail using BLAT. In this analysis, sequence reads were only called as inserts if the following criteria were met:

1. The start of the identity to the mouse genome was approximately at position 140bp for the left and 40bp for the right (+/-5bp): These positions are consistent with the known position of the amplification primers within T2Onc.

2. The identity of the high scoring hit was between 96 and 100%.

3. No higher scoring hit was obtained which mapped to different co-ordinates within the read (indicative of a PCR artefact).

The insertion sites identified from this BLAT analysis are presented in Table 3.3 and Table 3.4 for *Nla*III and *Bfa*I sites respectively. The initial observations obtained from the tables are as follows: First, 26 insertion sites have been identified. This is a

large number of inserts compared to the average obtained from standard subcloning and sequencing protocols (Collier *et al.,* 2005; Dupuy *et al.,* 2005) illustrating the power of FLX sequencing. In two cases, both the left and the right sides have been recovered; The *A230020J21Rik* gene has inserts scored on chromosome 1 at positions 192853626 (left) and 192853627 (right) respectively. The *Gm5069* gene, localized on the same chromosome, has an insert scored at positions 182259106 (left) and 182259109 (right), respectively. This confirms that the LM-PCR and sequencing from both Left and Right reactions has worked. Furthermore, approximately 54% of the insertions are located within genes, with 45% localised within intergenic DNA. This frequency of genic hits is consistent with previous analyses (Collier *et al.,* 2005). Interestingly, the *Prdm16* gene located on chromosome 4 has generated the most reads. The insertions into this gene accounts for ~11.6% of all reads identified from the *Nla*III digest. This may be due to variation in digestion/amplification efficiency, or may reflect variation in insert sites between cells. However, it is interesting that this gene is already known to be involved in the promotion of haematological malignancies (Shing *et al.,* 2007), suggesting that this insertion may be a key event in the development of this specific tumour.

Approximately 39.2% of the transposition positions are within chromosome 1. This is much higher than the frequency expected by chance, and is due to the well known local hopping behaviour exhibited by the SB transposon. The T2/Onc transgenic array within line 76 is known to lie at position 166.788 Mb on mouse chromosome 1, as sequence flanking the array was recovered by LM-PCR (Collier *et al.,* 2005). It is noteworthy that 12 out of the 19 insertions on chromosome 1 map to within 30Mb of this position.

Overal, therefore, the insertion sites identified in this tumour indicate that the SB system is behaving as anticipated in terms of its transposition dynamics, and that the insertion site recovery is efficient. This suggested that the insertion site data generated was of sufficient quality to be subject to a full bioinformatic analysis to automatically assign insert position, and identify common insertion sequences (CISs) (Dupuy *et al.,* 2005). All samples were therefore taken through the CIS analysis pipeline (see Methods).

Table 3.3: Insertion sites identified within *Nla*III restriction fragments. Both genic and intergenic insertions are presented. The inserts name and code, with the chromosomes (Chr) and gene orientation (Or) are shown together with the position of inserts mapped (Position start – end). Gene names and insert numbers are shown.

Insert code	Chr	Or	Position Start – End	Genes	No. of inserts
GIUO1OD06G6TA7	1	-	192853553 – 192853627	A230020J21Rik	4
GIUO1OD06HILZS	1	-	182259050 – 182259109	Gm5069	3
GIUO1OD06HNVQO	1	-	128137770 – 128137856	Nckap5	4
GIUO1OD06HJ4R3	1	-	107694701 – 107694832	Tnfrsf11a	1
GIUO1OD06HI7UG	1	+	182604447 – 182604533	Lin9	1
GIUO1OD06G9UOR	1	+	153867681 – 153867712	1700025G04Rik	1
GIUO1OD06HA0WC	1	+	36210622 – 36210678	Ugcgl1	1
GIUO1OD06HABGF	1	-	162517215 – 162517276	Rabgap1l	3
GIUO1OD06HLBA3	1	+	110405740 – 110405800	intergenic	1
GIUO1OD06HAJI1	1	+	192348485 – 192348523	intergenic	1
GIUO1OD06HE9H5	1	-	163485580 – 163485654	intergenic	1
GIUO1OD06HK4ZK	1	-	94066947 – 94067068	intergenic	1
GIUO1OD06G7RE0	2	-	120253125 – 120253208	Ganc	1
GIUO1OD06HPQ7I	4	-	153936641 – 153936686	Prdm16	13
GIUO1OD06G5VZL	4	+	62734919 – 62734941	Zfp618	1
GIUO1OD06G626K	4	+	32407598 – 32407625	intergenic c	6
GIUO1OD06G5B02	6	-	31087927 – 31087962		5
GIUO1OD06HMNX4	7	+	114459761 – 114459787	intergenic c	2
GIUO1OD06G5RUT	8	-	91238964 – 91239019	Cy1d	1
GIUO1OD06HOECH	10	-	20849509 – 20849662	Myb	3
GIUO1OD06G3KTP	10	-	103129070 – 103129114	intergenic	1
GIUO1OD06G4Q45	11	-	90134905 – 90134931	Mmd	4
GIUO1OD06G3KTP	11	-	9222831 – 9222865	Abca13	1
GIUO1OD06G3KTP	12	-	4626012 – 4626051	Itsn2	1
GIUO1OD06HDJF8	12	+	119741135 – 119741259	intergenic	1
GIUO1OD06HKHEL	14	-	61748860 – 61748887	intergenic	3
GIUO1OD06HBPMD	15	-	42342061 – 42342216	Angpt1	1
GIUO1OD06HNOXV	15	-	25487216 – 25487242	intergenic	1
GIUO1OD06HOECH	17	+	55893137 – 55893156	Emr4	1
GIUO1OD06G9QJW	17	+	3493680 – 3493708	Tiam2	1
GIUO1OD06HLBA3	X	-	157677067 – 157677086	Scm12	1
GIUO1OD06HJO8O	X_rand	-	1508536 – 1508602	intergenic	1
GIUO1OD06HF9K4	X	-	18832143 – 18832166	intergenic	1
GIUO1OD06G3KTP GIUO1OD06G3KTP GIUO1OD06G3KTP	3 16 19	- - +	137977321 – 137977364 61030171 – 61030214 19783450 – 19783493	Unmappable	2
GIUO1OD06G3KTP GIUO1OD06G3KTP GIUO1OD06G3KTP	3 4 4	- + +	110954681 – 110954707 96211275 – 96211301 27073100 – 27073126	Unmappable	1

Table 3.4: Insertion sites identified within *Bfa1* restriction fragments. Both genic and intergenic insertions are presented. The inserts name and code, with the chromosomes (Chr) and gene orientation (Or) are shown together with the position of inserts mapped (Position start – end). Gene names and insert numbers are shown.

Insert Code	Chr	Or	Position Start - End	Gene	No. of inserts
GIUO1OD06HB41H	1	+	192853626 – 192853811	**A230020J21RiK**	1
GIUO1OD06G5LCV	1	+	182259106 – 182259128	**Gm5069**	1
GIUO1OD06HN4RD	1	-	172298826 – 172298953	**Nos1ap**	1
GIUO1OD06HG0J5	1	+	192853626 – 192853811	**Intergenic**	2
GIUO1OD06HCYLB	1	+	192334903 – 192335003	**Intergenic**	2
GIUO1OD06G8RTF	1	+	94068028 – 94068161	**Intergenic**	2
GIUO1OD06HBA7N	1	+	166047280 – 166047301	**Intergenic**	1
GIUO1OD06G4D04	1	-	415644 – 415756	**Intergenic**	3
GIUO1OD06HAAA0	2	+	54158306 – 54158359	**Intergenic**	4
GIUO1OD06HISWH	4	-	37938914 – 37938972	**Intergenic**	3
GIUO1OD06HIRI3	6	+	31087959 – 31088181	**Intergenic**	1
GIUO1OD06G9TMJ	9	+	44584022 – 44584092	**Ift46**	1
GIUO1OD06HKVGV	9	-	69469549 – 69469829	**Intergenic**	1
GIUO1OD06G5QPE	10	+	127902897 – 127902987	**Smarcc2**	1
GIUO1OD06HK8D9	11	-	68244102 – 68244251	**Intergenic**	2
GIUO1OD06HPLSF	17	+	80840228 – 80840298	**Sos1**	4
GIUO1OD06HHVWI	X	+	108807692 – 108807829	**Hdx**	4
GIUO1OD06G4D04	X	+	99741305 – 99741325	**Phka1**	1
GIUO1OD06HDIG9 GIUO1OD06HDIG9 GIUO1OD06HDIG9	7 18 10	- - -	66411379 – 66411411 39593716 – 39593748 101757704 – 101757736	**Unmappable**	2

3.2.7 Identification of insertion sites within MB tumours and controls

The main aim of this project is to identify cancer genes and determine the initiation events in tumours generated within the *Ptch1*+/- mouse model, such as MB and rhabdomyosarcoma. To identify the insertion sites, 41 MBs, 28 rhabdomyosarcoma, 123 haematological tumours and 27 adenomas from the experimental genotype, together with tissue specific control samples for MB and rhabdomyosarcoma from the transposition-only genotype, were analysed by linker-mediated PCR to amplify across insert/genomic DNA junction. The products were then bar-coded by nested PCR and sequenced using FLX pyrosequencing (Dupuy *et al.*, 2009) aimed to a target depth of 200-1000 reads/tumour. This gave a final total of 8,360 inserts with an average of ~204 inserts/tumour in MB. These reads were mapped to the mouse genome (NCBI37/mm9). For more details about the inserts in the whole cohort of tumours, see Table 3.5. These inserts were further analysed with two statistical methods; Monte Carlo (MC) simulation and Gaussian Kernel Convolution (GKC) analysis by Dr David Adams and Alistair Rust (Welcome Trust Sanger Institute, UK) (Collier *et al.*, 2005;de Ridder *et al.*, 2006).

Table 3.5: Tumours by type, with the number of reads. For details of stacking (inserts at the same dinucleotide) and firestorm (multiple inserts in one locus from the same sample) filtering, see following sections.

Sample sets \ Counts	Number of samples	Insertions after filtering by samples	Stacking insertions removed (one max per location)	Firestorm insertions removed (max of 2 per CIS)	Final insertions count
Adenomas	27	1220	14	0	1206
Brain controls	31	4147	131	35	3981
Haematological	123	14389	373	182	13834
MB	41	8573	134	79	8360
Muscle controls	40	2208	60	6	2142
RMS	28	3308	20	14	3274

3.2.7.1 Monte Carlo (MC) and Gaussian Kernel Convolution (GKC) analysis

In the MC method, the original data, consisting of insert positions within the genome, is used to generate pseudo-replicates (like bootstrap analysis) in which the insert positions are randomly distributed. Insert position in tumours is then analysed using different sized windows e.g. 10Kb, 30Kb, 60Kb, and 100Kb, and the original data is compared against the pseudo-replicates to localise sites within the genome where the number of inserts is greater than expected [termed common insertion sites (CISs)]. This method has the benefit of being relatively straightforward. However, it becomes less sensitive with larger window sizes, as it does not take into account genome structure (Copeland and Jenkins, 2010;Dupuy, 2010).

The GKC method uses a kernel function, which is localised at every single insert within the genome; i.e. it estimates an insert probability function. This involves measuring the probability of an insert occurring within a particular region within a window of defined size. This analysis is performed several times using a variety of window sizes. The original data is analysed against a probability distribution curve to identify CISs. This allows for a higher number of CISs to be identified at a wider variety of scales, and to generate probabilities which are comparable across different Kernel widths (window sizes). This method is relatively fast and more sensitive than MC, but is mathematically more complex (Copeland and Jenkins, 2010;Dupuy, 2010).

3.2.7.2. ***In Silico*** **analysis to Identify MB CISs**

From a final cohort of 40 MB samples (1 of the 41 tumours failing to yield inserts), a set of 46 genic and 9 intergenic regions were defined as common insertion sites (CIS) using Monte Carlo simulation and Kernel Convolution methods. However, Some CISs were defined by inserts from different tumours at exactly the same dinucleotide position. Other CISs were defined by multiple inserts from the same tumours (e.g. Figure 3.14). The former could be due to PCR contamination or sequencing artifacts, while the latter could be due to secondary local hopping events. Both of these could confound the analysis (Starr *et al.,* 2009) and could lead to false positives, as they violate the assumption of independence of insertions.

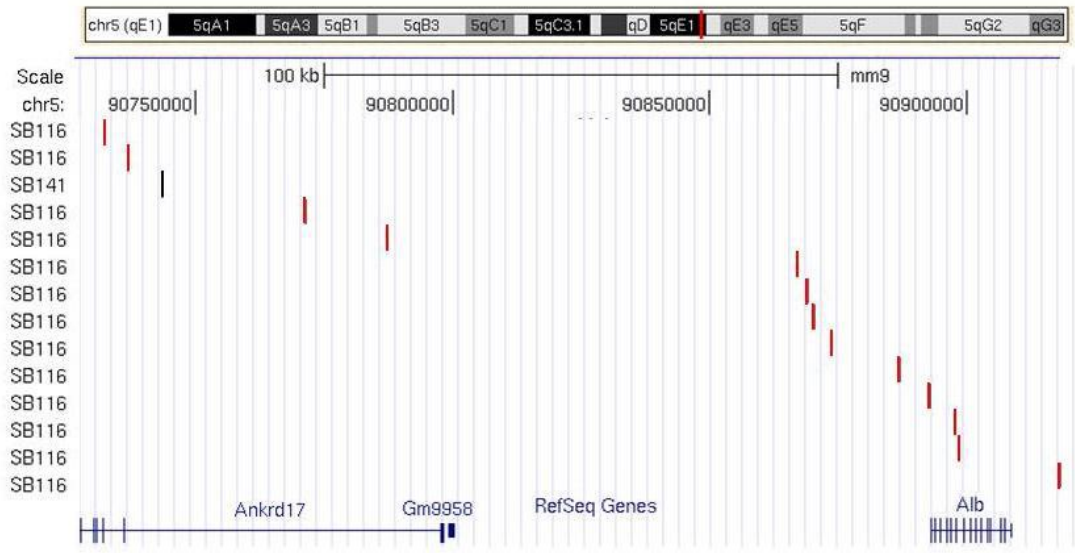

Figure 3.14: Example of CIS defined by inserts within one tumour. Image adapted from UCSC browser showing T2/Onc insert positions within a CIS from Cerebellum Controls. Thirteen out of fourteen inserts (red) are derived from a single tumour (SB116).

To take account of these possible confounding issues, the sequence data were reanalysed as follows: First, where multiple inserts were identified at the same position in more than one sample, the sample with the largest number of reads at that position was retained and others discarded. In the event of a tie, one sample was arbitrarily chosen. This conservative approach has been used before (Dupuy *et al.*, 2009;Starr *et al.*, 2009). To remove the "firestorm" CIS pattern, the GKC was re-run once with a kernel window of 15Kb and a maximum of 2 insertions from any single sample within a single CIS was retained. Additionally, in the case of tied numbers of mapped reads in a CIS, only the closest insertion to the kernel peak of the CIS was retained. Finally, any mapped reads with insertion sites shared with control samples were omitted from the analysis. The final reads number retained from all tumours types after this analysis are compiled within Table 3.5.

3.2.7.3 CIS candidate genes in MB samples

After reanalysis, 18 candidate genes and 3 intergenic regions were identified within 21 CISs recovered with median P-value = 0.008, and these are listed in Table 3.5. Those genes were analysed bioinformatically to establish the number of inserts, insert orientation bias, inserts clusters with respect to gene structures, and gene transcriptional orientation, to infer the likely impact on the gene transcription (Starr *et al.*, 2009). Several databases were investigated for previous identification for these CISs as tumourgenic genes.

It is to be noted that 11 candidate genes (61.9%) were identified by both statistical methods. Interestingly, of the remaining 7 candidate genes, 6 were identified by GKC only, and 1 (*Gli3*) by MC only, suggesting that the combination of such statistical methods is highly recommended (Table 3.6). In the GKC analysis, there were two statistical cut-offs applied based on either genomic or chromosomal levels. For each chromosome and GKC kernel size, potential CISs are assigned p-values. For a CIS to be classified as significant, its p-value must be below 0.05 which is adjusted using a multiple test correction. In the case of chromosomes this is 0.05 divided by the total number of CIS peaks on the current chromosome. CIS selected in this way are thus significant for the specific chromosome they lie in. Then the multiple test corrected p-value for the CIS was implemented. In this case it is the (raw p-value X the number of CIS peaks on its chromosome). Values given must be less than 0.05 to satisfy the chromosome-by-chromosome cut-off.

At the genomic level however, a CIS that is significant on a certain chromosome may not be significant on the genome-scale. Therefore, a CIS to be called significant on the genome-scale, its p-value should be less than 0.05 but in this case will be corrected for the total number of CIS peaks over the whole genome. Since this total number of peaks is greater than the individual numbers of peaks on chromosomes, the cut-off for a peak to be called significant becomes far more stringent (i.e. 0.05 divided by the total number of genome CIS peaks). This was carried out as the (raw p-value X the number of CIS on its chromosome X 19). This is an approximate calculation adjusting by the "number of chromosomes". It is only 19 as we don't look at chr1 and there are very few CIS peaks on chromosome Y. Accordingly, 6 genes (*Nfia, Atxn2, Tead1, Tgif2, Crebbp, Dscr3*) showed a significant p-value ($p = <0.05$) at the genomic level. While the rest are significant only at the chromosomal level ($p = <0.05$) (Table 3.6). However, the genome level cut off is considered to be very stringent, as it cannot account for chromosomal variation in size, gene number etc. As a result, the chromosome cut off is generally reported in exploratory datasets.

3.2.7.4 Investigating the inferred Mode of Action of inserts upon candidate genes

The orientation bias of the inserted transposon with genes, as well as the location of the inserts within the genes, can be used to infer of mode of action (MOA), and this information is specified in Table 3.6. Approximately 52.6% of the inserts were

in the positive orientation with respect to gene transcription, compared to about 47.5% that were orientated negatively, indicating no significant bias in orientation across the whole dataset. However, 6 CIS (*Tgif2, Fgif13, Itgbl1, Gli3, Myt1l,* and *Adcy5*) were determined to have statistically significant bias in insertion orientation, suggesting that the mode of action (MOA) could be clearly inferred. If the T2/Onc is in the same orientation as the gene transcript, the MSCV promoter may lead to overexpression of the gene. However, if the T2/Onc is in the opposite direction to the gene, this may lead to gene truncation and dysregulation due to splice acceptor signals (Dupuy, 2010). The insert orientation, the number of the inserts, and the location of the inserts that clustered within the 5' end of the genes suggested upregulation effects for 4 genes (*Tgif2, Tead1, Adcy5,* and *Fgf13*). In contrast, the rest of the genes have high numbers of inserts distributed on a wider scale, suggesting loss of function or more complex effects of insertion upon the affected genes.

Two of the CISs, namely *Pten* and *Crebbp*, were identified as showing somatic mutation at low frequency within human MB during unsupervised sequencing assays; nonsense mutations within the human ortholog of *Crebbp* have been identified previously within WNT-subgroup tumours (Parsons *et al.*, 2011;Robinson *et al.*, 2012). Strikingly, 6 identified CISs (*Nfia*, *Nfib*, *Tead1*, *Crebbp*, *Sfi1* and *Pten)* have been identified as MB CISs within an independent tissue-conditioned mutagenesis screen study of MB (Wu *et al.*, 2012). Within the cerebellar controls, 3 out of the 5 CISs are genic, all with p-values between 0.01 and 0.05 (Table 3.6). This does suggest that a small number of the MB CISs with low p-values may be false positives.

Table 3.6: Identified Common insertion sites (CISs) in Medulloblastoma and normal Cerebellum controls.

Gene	P-value	Gene orientation	N Tumours	N Insertion Sites	Insert Bias	Inferred MOA	Chr – Mouse	Chr- Human	MB-Mut	MB-CIS	CGC	Other genes
MC+ GKC Genomic p<0.05												
Nfia	<1E-14	+	20	27	.4/22	ORF Disruption	4	1p31.3-p31.2	-	Ptch	-	-
Atxn2	<1E-14	+	6	6	.1/5	Unclear	5	12q24.1	-	-	-	-
Tead1	7.34E-13	+	6	7	.2/3	Gain	7	11p15.2	-	Ptch	-	-
Tgif2	0.0115	+	3	4	.4/0	Gain	2	20q11.2-q12	-	-	-	-
Crebbp	0.0170	-	5	5	.2/3	ORF Disruption	16	16p13.3	Y	Ptch	S	-
Dscr3	0.0178	-	5	6	.3/2	ORF Disruption	16	21q22.13	-	-	-	-
CIS13:72336914	0.0042		5	6			13					
MC+ GKC Chromosomal p<0.05												
Pten	0.0046	+	4	5	.1/4	ORF disruption	19	10q23.3	Y	Both	S/G	
Itgbl1	0.0062	+	5	5	.5/0	Unclear	14	13q33	-	-	-	*Fgf14*
Nfib	0.0072	-	7	9	.5/4	Gain	4	9p24.1	-	Ptch	S	-
Tmem45b	0.0195	-	4	5	.4/1	Unclear	9	11q24.3	-	-	-	-
Sfi1	0.0214	-	5	6	.3/3	ORF Disruption	11	22q12.2	-	Both	-	-
MC only												
Gli3	n.a.	+	3	4	.4/0	ORF Disruption	13	7p13	-	-	-	-
GKC-only Chromosomal p<0.05												
Myt1l	0.0089	+	8	10	.7/3	ORF Disruption	12	2p25.3	-	-	-	*Pxdn,Tpo, Sntg2*

Ankrd5	0.0092	-	8	15	.3/2	Unclear	2	20p12	-	-	-	*Plcb4,Pack7,Snap25*
Slit3	0.0122	+	8	12	.6/6	ORF Disruption	11	5q35	-	-	-	-
Fgf13	0.0254	-	3	4	.1/3	Gain	X	Xq26.3	-	-	-	-
CIS15:70979306	0.0279		4	4			15					
CIS3:147532546	0.0430		4	7			3					
L3mbtl4	0.0432		7	7	.2/5	ORF Disruption	17	Chr18p11.31	-	-	-	*Tmem200c*
Adcy5	0.0443	+	4	6	.4/2	ORF disruption	16	3q13.2-q21	-	-	-	*Ptblb*
Cerebellar Controls												
Faf1	0.01		4	4		Unclear	4					
I7Rn6	0.018		3	3		Unclear	7					*Ccdc81*
CIS36:31346217	0.045		6	8			6					
Ric3	0.035		4	4		Unclear	7					*Tub*
CIS9:68415409	0.016		9	11			9					

Genes identified the CISs. Chromosome in which the CISs mapped; P-value is from GKC analysis unless otherwise stated;Gene orientation and direction of gene transcription is according to UCSC - (+) positive , (-) negative orientation; (N tumour) is the number of tumours that defined the CIS; (N insertion site) is the number of T2/Onc insertions; (Insertion Bias) is the number of inserts in the same orientation right/ number of inserts opposite to gene orientation; (Inferred MOA) is the inferred mode of action based on the position and orientation of T2/Onc insertions - (open reading frame disruption, or upregulation); (Chr-Mouse) mouse chromosome in which the CISs mapped; (Chr-Human) orthologous human chromosome; (MB-Mut) Somatic mutation previously identified (Parson et al., 2011; Jones et al., 2012; Pugh et al., 2012; Robinson et al., 2012), (Y) yes; (MB-CIS) gene previously identified as MB CISs (Wu *et al.*, 2012) in (*Ptch1*) *Ptch* model or *Ptch and p53* model (Both); (other gene) lists additional genes identified within the same GKC width.

3.2.8. Whole body SB mutagenesis identifies common Insertion sites in primary tumours other that MB

Although the main focus of this study and thesis is CISs in MB and their impact upon MB development, there were a variety of other tumours generated within the model used in this experiment (Table 3.1). The vast majority were haematological tumours, and these are not dealt with here as they have been documented previously in this model (Collier *et al.,* 2009). However, a number of rhabdomyosarcomas and liver adenomas were also identified, and the next sections will explore the CISs identified in these two tumours.

3.2.8.1 CISs identified in Adenomas

Transposon distribution was also analysed in liver adenomas as these have not previously been described in whole body SB mutagenesis. Because adenomas were often recovered in animals with haematological tumours and shown evidence of lymphocytic infiltration, CISs identified in this dataset which were identified at a higher p-value in the haematological tumours were excluded. Seven genic CISs were identified as being specific to adenomas (Table 3.7A) including 2 within genes recently shown to be subject to rare somatic mutation (*TLK1*) or focal amplification (*ERBB2IP*) in hepatocellular carcinoma (Guichard *et al.,* 2012). However, the most significant CIS was centred on intron 24 of the *Egfr* gene ($p<1E-14$), the same location as the top scoring CIS in a conditional SB mutagenesis screen for genes associated with hepatocellular carcinoma (Keng *et al.,* 2009), confirming both the specificity of the mutational event in this gene (truncation) and its importance in neoplasia of the liver.

Table 3.7: CISs identified within liver Adenomas by SB mutagenesis:

Gene	Chr	CIS Width	N	I	p-value	Other Genes	MOA	CGC
A. CISs in Adenomas								
Egfr	11	21.9	7	8	<0.0001		Orf Disruption	S
Ubr2	17	260	4	5	0.0049	*Trerf1*	Unclear	NA
Erbb2ip	13	780	6	6	0.0102		Orf Disruption	S
Rapgef6	11	8.8	3	3	0.0143		Unclear	NA
Ovol2	2	27.7	3	3	0.0155	*Csrp2bp*	Unclear	NA
Tlk1	2	134	3	4	0.0203		Orf Disruption	S
Rad23b	4	31.3	3	3	0.0264		Orf Disruption	S
B. CISs in Adenomas also identified as CISs in Haematological tumours								
Pten	19	36.1	5	6	<0.0001		Orf Disruption	S
Notch1	2	1.2	3	3	<0.0001		Gain	G
Ikzf1	11	62.7	5	7	0.0001		Orf Disruption	G
Mecom	3	24.6	3	4	0.0009		Unclear	S
Myb	10	3.4	3	3	0.0003		Gain	G
Foxp1	6	15.5	3	3	0.0089	*Eif4e3*	Orf Disruption	S

For details of columns see Legend to Table 3.6

3.2.8.2 CISs identified in rhabdomyosarcoma

The RMS and muscle control datasets were also analysed using both GKC and MC analyses. The only GKC CIS in the RMS data which was significant at the genome scale is the Chr7 insertion hotspot recovered in all CIS analyses, including controls. Two CISs were significant after correction at the chromosomal level (CIS2:99636326 and *Rin3*), but two CISs were also recovered at this level of significance in the muscle control dataset (Table 3.8), suggesting that these CISs are likely to be false positives. The lack of a clear impact of transposition upon RMS incidence (Figure 3.7) and the small number of CISs recovered suggest that there is no interaction between the SB mutagenesis system used here and the *Ptch* induced predisposition to RMS.

Table 3.8: CISs identified within rhabdomyosarcoma by SB mutagenesis:

Genes	Chr	N	I	P=value	MOA
A. CISs in RMS					
Ptgis	2	6	9	0.011	Orf Disruption
CIS2:99636326	2	3	3	0.017	Orf Disruption
Rin3	12	3	3	0.018	Gain
Zfx	X	3	3	0.022	Orf Disruption
CIS13:48649631	13	3	3	0.024	Unclear
Golim4	3	6	8	0.030	Orf Disruption
Zmynd11	13	8	8	0.035	Orf Disruption
Diap2	X	4	4	0.037	Unclear
Ap3m1	14	8	12	0.040	Orf Disruption
Foxp1	6	3	3	0.006	Unclear
Erg	16	4	4	0.023	Orf Disruption
Fgf14	14	3	4	0.037	Orf Disruption
B. Muscle Controls					
Iqcg	16	10	10	0.016	
Ankrd52	10	4	4	0.020	
Nfia	4	4	4	0.034	
Mecom	3	3	4	0.004	
Myb	10	3	3	0.028	
Pten	19	3	3	0.032	
Ikzf1	11	3	3	0.036	

For details of columns, see Legend to Table 3.6.

3.3 Discussion

The Sleeping Beauty system has been successfully used to generate tumours in a variety of mouse cancer models and wild type mice. These tumours resemble human tumours, enabling novel genes that contribute to tumour formation to be identified and analysed. For instance, such identification has been carried out for sarcomas, T-cell lymphomas, hepatocellular carcinomas, and colorectal cancer (Collier *et al.*, 2005; Dupuy *et al.*, 2005; Collier *et al.*, 2009, Keng, *et al.*, 2009; Starr *et al.*, 2009).

In this screen, a cohort of *Ptch1*+/- mice and their wild type littermates were subjected to SB mutagenesis using the novel combination of the SB11 transposon and the *T2/Onc* mutagen. Several tumour types were generated, the main cancers being medulloblastomas, rhabdomyosarcomas, haematological malignancies (lymphoma and leukaemia) and liver adenomas. Critically, from the point of view of MB candidate gene identification, SB mutagenesis significantly increased the frequency of MB formation within *Ptch1*+/- animals relative to predisposition only controls, providing clear evidence that genes mutated by SB enhance the tumour predisposition, meaning that CISs identified within these tumours are likely to be involved in tumorigenesis. Although rhabdomyosarcomas were also generated by the model, there was no evidence of an impact of SB. Although the reason for this is unclear, it is possible that the MSCV enhancer does not function within the cell of origin of these tumours. Despite this, tumours of all types, together with control samples, were taken through the insert identification process.

In this project two different FLX next generation sequencing chemistries (GS20 and Titanium) were used in order to improve deep sequencing of inserts from tumours generated in the SB mutagenised mouse model. run: Both refer to chips used by FLX next generation sequencing machine, but GS20 has lower capacity in terms of read number and read length. Titanium-chemistry generates approximately 1.2 million reads with an average length of 350 – 450bp. In contrast, . GS20 chemistry can generate approximately 400,000 reads with an average read length of 350 – 450bp. Here, Titanum runs were used to analyse approximately 90 samples per run, while 40 -50 samples were loaded on GS20 runs. In our study the length of the reads recovered was important as we have used a random cloning approach where the enzyme can cut within a position distributed within the genome. Titanum enables recovery of more TE inserts due to the longer reads been analysed (Bard-Chapeau *et al.*, 2014).

The preliminary analysis of 100 reads recovered by FLX sequencing within a single haematological tumour identified several genes, including the *PR* domain containing 16 (*Prdm16*) gene on chromosome 4. This gene has been shown to be over-expressed in acute myeloid leukemia (AML), in both 5 human and 13 mice samples, suggesting that *Prdm16* may contribute to AML formation (Shing *et al.,* 2007). This result suggested that the Splinkerette PCR had worked effectively, and that insert identification would be successful. In addition, in this tumour the incidence of local hopping was about 39.2% on chromosome 1. This percentage is only an estimate as it was calculated from only 100 insertions. However, in previous analyses local hopping has been seen at frequencies ranging from 23% within 28 sarcomas (line 76, *Collier et al.,* 2005), 6 – 11% from 16 haematological tumours (Dupuy *et al.,* 2005;Collier *et al.,* 2009), and about 11.2% local hopping events within 25 Mb, 15.6% within 40 Mb and 31.5% within the entire chromosome within 59 haematological malignancies (line 76, Collier *et al.,* 2007).

Analysis of the MB insertion site data identified a total of 18 CIS genes using both GKC and MC methods. Until recently, only two of these (*Pten* and *Crebbp*) have been observed in previous analyses of other tumour types using the *SB* screening system (Collier *et al.,* 2009;Berquam-Vrieze *et al.,* 2011). These 18 CISs are likely to have a synergistic effect with the predisposition of *Ptch* during tumour development, as *SB* does not induce MBs alone. It is noteworthy that six CISs identified in this analysis have also recently been identified as MB CISs using a tissue-specific MB mutagenesis screen of the same model (Wu *et al.,* 2012), The overlap between these two studies suggests that the genes identified are relevant to MB specifically, and are likely to be relevant to MB formation/development.

Many of these CISs identified in the current study are known to play a role in neural development. This is consistent with the suggestion that a limited number of somatic mutations may be required to develop MB, compared to adult solid tumours, given the short time period of cerebellar development (Parsons *et al.*, 2011).

Of those CISs which are neuronal development genes, *Nfia* is the most highly inserted CIS, and a paralog of this gene, *Nfib*, is also a CIS gene. Nfi genes act as transcriptional factors, and are involved in gliogenesis of the spinal cord and axonal outgrowth (Steele-Perkins *et al.*, 2005;Deneen *et al.*, 2006). In the cerebellum specifically, *Nfia* and *Nfib* were found to be highly expressed in the cell nuclei of immature GNPCs as they evolved to the postmitotic stage in the external germinal layer (Wang *et al.,* 2007). And this expression was found to remain high until these cells

reach their final maturation in the internal granule cell layer of the cerebellum. Additionally, *Nfia* and *Nfib* have a significant role in axonal outgrowth, so when *Nfi* family expression was repressed by an *Nfi* dominant repressor, the outgrowth of axons was significantly affected. Moreover, *Nfia* knockout mice showed a dramatic defect in dendrite development, and large numbers of GNPCs were unable to migrate to the internal granule layer of the cerebellum, remaining in the molecular layer. Furthermore, *Nfib* knockout mice present with perinatal mortality and cerebellar foliation impairment was recorded (Wang *et al.*, 2007).

The identification of 3 CISs with low GKC p-values within the cerebellum control samples suggests that some of the CISs identified here may be false positives. Despite this, some of the CISs that have been identified by only a single statistical method are also plausible candidates. For instance, *Gli3*, pulled out by the MC anaylsis, is a known downstream target of the *SHH* signalling pathway and it is also a known tumour suppressor gene (Erez *et al.*, 2002). Recently, the role of *Gli3* was identified in mouse brain development (Magnani *et al.*, 2012). The *Gli3* hypomorphic mouse mutant Polydactyly Nagoya (*Pdn*) shows features of agenesis of the corpus callosum (CC, a flat array of neural fibres that connect the two cerebral hemispheres) and misplaced glial and neuronal guidepost cells. Additionally, the expression levels of *Nfi* family members (*Nfia*, *Nfib* and *Nfix*) in the cortex (where they have to be at a higher expression level) and in the dorsal-most septum (where they have to be expressed at lower levels) were investigated. The results revealed the disappearance of all aforementioned gene expressions in the cortex, while the only *Nfia* expression was at higher levels in the septum. Finally, disruption of the *FGF* and *WNT/β-catenin* signalling pathways was also observed (Magnani *et al.*, 2012).

Of the genes only identified by GKC, *Myt1L* and *Slit3* are of particular interest. *Myt1L* is a pan neural transcription factor which can help to convert fibroblasts to neurons (Vierbuchen *et al.*, 2010), *Slit3*, with other Slit genes (*Slit1* and *Slit2*) are known to play a role in stimulation of axonal branch formation and their expression is upregulated postnatally (Marillat *et al.*, 2002). Furthermore, *Slit3* controls axonal guidance and neuronal cell migration in the CNS midline (Yiin *et al.*, 2009). Additionally, *SLIT3* was found to be methylated in about 29% human glioma cell lines and in approximately 35% of human primary glioma tumours (Dickinson *et al.*, 2004). In addition, in zebrafish neuronal development *Slit3* has a role in directing glial cells localisation and position (Barresi *et al.*, 2005). *Slit3* is considered a tumour suppressor gene as its over-expression in the *Sdf1*/*Cxcr4* mouse model of breast cancer reduced

tumour formation (Marlow *et al.*, 2008). The inferred mode of action of the insertions within *Slit3* is ORF disruption, consistent with this suggestion. The identification of these two genes by GKC alone is also to be expected, as both are very large genes (spanning ~385 and 590kb respectively. Both were only identified in the widest Kernel width analysis (120kb GKC scale), as the inserts were fairly evenly distributed within both genes (consistent with loss of function). This illustrates the effectiveness of combining the two statistical methods.

Although, Liver adinomas generated in this study were small in number (19 tumours) they are, to my knowledge, the only adenomas generated by whole-body SB mutagenesis. About 1206 transposon reads were recovered from these and 7 CISs were identified. However, Keng *et al.*, (2009) generated 68 nodules and recovered 8060 insertions, and 19 CISs were identified by utilising conditioned mobilised SB transposase by Cre recombinase. Furthermore, as *Egfr* gene was identified by both study revels that this gene is specific and important for hepatocellular carcinoma generation.

In conclusion, a whole-body *SB* mutagenesis system used with the *Ptch1*+/- murine model has increased the frequency of MB and decreased the latency. Efficient genotyping, high depth sequencing analysis for 40 MB samples and MC GKC statistical methods have identified 18 genic CISs, many of which are plausible candidates for involvement in MB. The validation of these gene CISs within experimental MB samples would be a high priority before downstream analyses, and this will be presented and discussed in the next chapter.

Chapter 4. (Results 2)

Validation of CISs and their inferred mode of action

4.1 Introduction

Insertional mutagenesis is often used for somatic cell forward-genetic-screening in model organisms to identify tumour suppressor genes and oncogenes relevant to human cancer (Ivics *et al.*, 1997) (Callahan and Smith, 2000;Uren *et al.*, 2005;Collier and Largaespada, 2006). The mutagen has the ability to work as a molecular tag, allowing for the prompt identification of mutagenized genomic loci through the identification of common insertion sites [CISs (Starr *et al.*, 2009)]. The Sleeping Beauty (SB) mutagenesis system is one of the current methods used for insertional mutagenesis screens (Collier *et al.*, 2005;Kool and Berns, 2009;Starr *et al.*, 2011) and has been coupled to next generation sequencing (NGS) technology to facilitate deep sampling and recovery of insertion sites. For instance, in a recent study conducted in colorectal cancer, where the SB system was used to identify candidate CISs, the insertion read recovery was about 347,993 sequence reads from 446 tumours (Starr *et al.*, 2011).

In Chapter 3 I presented the results of a whole body SB mutagenesis screen of the *Ptch1*+/- model of MB. Mutagenesis increased the frequency of MB in predisposed animals, but did not generate MBs in wild type littermates. It also generated additional tumour types including rhabdomyosarcomas and adenomas. However, as the primary aim of this screen was to identify genes involved in MB, the focus of my work was the CISs identified within the data from MB tumours. FLX sequencing recovered approximately 204 inserts per tumour, comparable to other analyses utilising this technique, and in total 18 CIS genes were identified using two different statistical methods, with 11 being identified using both methods (Table 3.6).

Prioritisation of these candidate genes for downstream analysis could be done based solely on CIS p-value and known function. However, the SB mutagenesis system, and subsequent CIS identification steps, can yield false positives, and can confound such a direct approach. Tumour cells may comprise both clonal and subclonal insertions that enable the position of cancer-causing genes to be identified by mutagenesis. Most inserts in tumour cells consist of passenger mutations, which are also known as non-lesion-causing insertions, and these could lead to false positives. Second, some tumours

may contain contaminating normal cells that will present with a variety of SB insertion patterns (Copeland and Jenkins, 2010;Brett *et al.*, 2011). The SB system does also show some insertion site bias, which may also contribute to false positives. The use of statistical analyses and normal cerebellar samples act as controls for such false positives, and the recovery of a small number of CISs from the control samples suggests that some of the MB CISs with low p-values may indeed be false positives.

In addition to the mutagenesis step, CIS identification uses both PCR and high throughput sequencing, both of which could be affected by experimental error such as PCR contamination, or errors in mapping due to low quality sequence reads. Removal of inserts in different tumours sharing the same insertion site is used here, and elsewhere, to control for the former. In addition, our HTG sequencing has identified a significant number of "firestorm" insertions, likely to be caused by secondary transposition events being picked up by the high depth of sequencing. Because of these different sources of potential false positives, it is also necessary to validate some of the mutation data experimentally.

The insertion sites which define CISs need to be validated at the genomic level, to confirm the integrity of individual T2/Onc insertions. This validation process is presented here in this chapter, with the insertion's location and impact being addressed at both the genomic and transcriptomic level for a subset of MB CIS genes, and for one CIS identified in liver adenomas; *Egfr*. Although adenomas were not the main focus of this thesis, *Egfr* inserts were chosen for analysis due to the large number and unusual distribution of the inserts within this gene, and because the position of *Egfr* inserts has been analysed in a previous SB screening study (Keng *et al.*, 2009) allowing the consistency of the mutagenesis to be assessed.

Validation of inserts can also be performed at the transcriptional level, to confirm the impact of individual inserts on transcript structure and by extension, gene function. The SB transposon carries the murine stem cell virus' (MSCV) long terminal repeat (LTR), and a splice donor (SD) from exon 1 of the murine *FOXP2* gene. These enable overexpression of oncogenes upon insertion upstream of, or within, a gene if insertion is in the same orientation as transcription. On the other hand, the transposon also carries a splice acceptor (SA) and polyadenylation signal (PA): Thus they have the ability to terminate transcription of a tumour suppressor gene upon insertion into the gene in either orientation (Figures 4.1 A and 4.1 B) (McIntyre *et al.*, 2012). For example, Dupuy *et al.*, 2005, established that about 60% of the T-cell leukaemias that were analysed in their first SB study contained insertions within intron 27 of *Notch-1*.

All the inserted transposons were located in the same transcriptional orientation as the gene, causing a 5' truncated protein to be expressed due to the presence of the MSCV promoter. This allows the mode of action inferred from the position and orientation of inserts within genes to be verified (Collier and Largaespada, 2006).

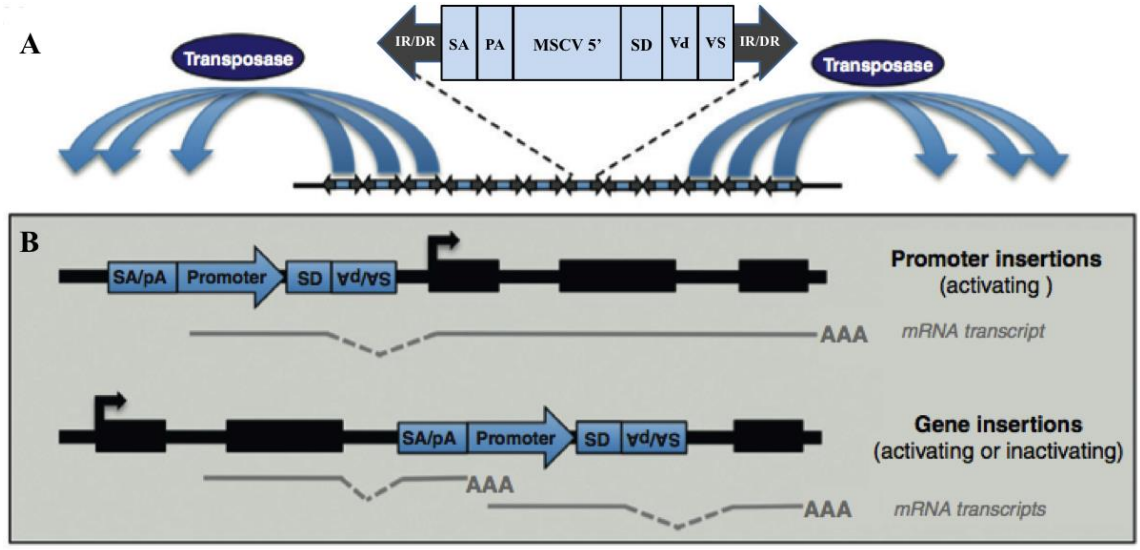

Figure 4.1. Sleeping Beauty system and the impact of insertions. (A) Sleeping Beauty mutagenesis system. The concatemerised copies of the transposon (T2/Onc) (highlighted) excise and mobilise (hop) from the donor chromosome/allele, with the aid of the transposase enzyme, and are pasted into a recipient gene (common insertion site: CIS). **(B)** Impact of the transposon on the affected gene. The transposon can trap the gene bi-directionally and, together with the promoter, it can terminate transcription prematurely (SA and PA) or overexpress genes (via the MSCV and SD). IR/DR, indirect repeat/direct repeat; SA, splice acceptor, PA, polyadenylation signalling, SD, splice donor; MSCV 5', murine stem cell virus promoter. (Figure adapted from Mclntyre et al., 2012).

Before analysing specific genes, several MB screens have used both co-occurence or mutual exclusion of CIS inserts within tumours, KEGG analyses, or Ontology analyses to infer functional redundancy (Wu *et al.*, 2012). Due to the small number of inserts identified here, I analysed Ontology only. In addition, a number of gene expression analyses were analysed which have identified genes with expression differences in certain MB subgroups. Querying of this data may also provide insight into the candidature, or subgroup specificity, of the CIS identified.

4.1.1 Aims

The aims of the work presented here are therefore:

1. To validate the SB insertion datasets at the genomic level, and assess the impact of insertion on gene expression in a small number of cases. To achieve this, I examined the inserted transposons within two genes, *Tgif2* and *Egfr*, in detail. I also examined the impact of the insertions on the expression of several genes in terms of transcript structure and/or expression levels.

2. To use Gene Ontology and analysis of gene expression in human tumour series to gain insight into the biology of the interacting mutations, and whether the genes identified are likely to be relevant to SHH subgroup MB, or the wider tumour type.

4.2 Results

4.2.1 Validation of CIS Data within Tumour Samples

To validate the insertion site data, several genes were chosen to confirm the presence of specific insertion sites and to also confirm the impact of these insertions upon transcription. The following MB CIS genes were chosen based on GKC p-value, mode of action (both loss of function and upregulation), and ease of analysis in terms of clustering of insert location: *Tgif2, NfiA, Tead1* and *Myt1l*. In addition, *Egfr*, the most significant CIS in the adenoma dataset (Table 3.7) was analysed, to allow direct comparison of insert distribution with a previous SB study of a hepatocellular carcinoma mouse model which identified *Egfr* as the top CIS hit (Keng *et al.*, 2009).

4.2.1.1 Validation of the genomic inserts

First, to validate the FLX sequence data within a subset of CISs, genomic PCR assays were designed to confirm specific transposon insertional events, by amplifying the junction between the T2/Onc insertions and the flanking genomic DNA sequences. This assay consists of two genomic primers flanking the region where the inserts are present (as a positive control for wild-type DNA) and one transposon-specific primer that amplifies out from the IR/DR(R) region of the T2/Onc insert: when an insert is absent from the genome, an amplified control product of the expected size will be obtained using the genomic primers. In contrast, when the insert is present within the genome, the transposon-specific primer can give amplification with one of the gene-specific primers (forward or reverse), depending on the insert orientation relative to the gene.

For most CISs the inserts are spread throughout the targeted gene, making validation of multiple inserts labour intensive, e.g. *Nfia*. Two genes, *Tgif2* and *Egfr*, were chosen for detailed analysis because their inserts were tightly clustered, allowing analysis of multiple independent insertions using a single gene-specific primer pair. In both cases, all the inserts were located within a single intron. In the case of *Tgif2*, all the inserts are located in the same orientation as the gene on the positive strand, implying upregulation. However, in the case of *Egfr*, the inserts are inserted in the negative orientation with respect to gene transcription, implying transcript truncation.

4.2.1.1.a Validation of Tgif2 inserts in MB samples

Five MB samples were originally identified with inserts in *Tgif*2 using FLX sequencing and they were analysed by genomic PCR. However, two samples, 791 and 793, had an insert within the same dinucleotide which suggests that one of them may be an artefact. All the inserts were located in the first intron and all were in the sense orientation (Figure 4.2 A). Therefore, the second (E2) exon-specific reverse primer and the transposon-specific IR/DR(R) primer were used to check for the presence of inserts. As shown in Figure 4.2, amplification products were generated from all of the samples with SB inserts except for sample 793, with amplicon size varying according to the distance of the transposon from E2. The expected sizes of the products were ~483 bp for sample 8, and ~1.3 kb for samples 580, 524, and 791, respectively. Additionally, four MB samples without inserts in this gene and three normal cerebellum samples were analysed, and no insertions were detected (Figure 4.2B). These results validate the FLX sequencing data. However, the failure to confirm the presence of an insert in sample 793 highlights one potentially confounding factor when using LM-PCR and FLX sequencing for transposon insert detection, and confirms both that some inserts reported to be in the same dinucleotide are false-positives, and that these do need to be controlled for during the CIS analysis. These false positives could be due either to sample/PCR contamination or misreading of the sample specific barcode during FLX sequencing.

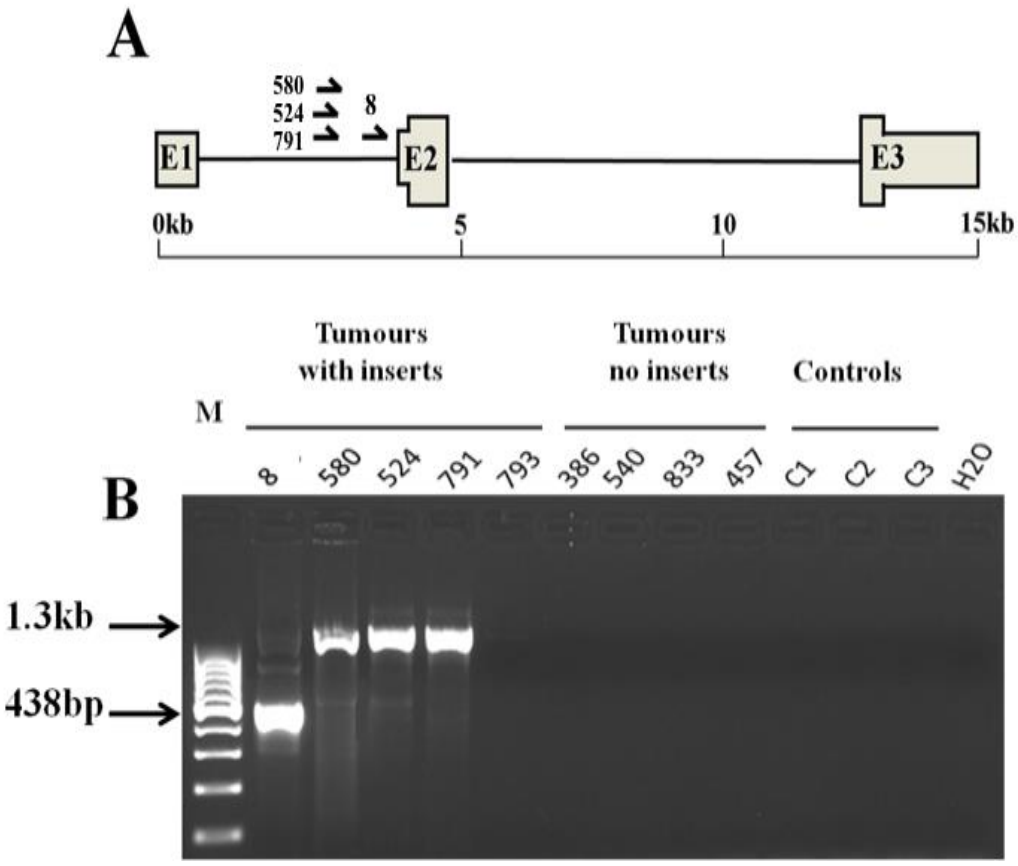

Figure 4.2. Validation of genomic inserts within *Tgif2*. (A) Location and orientation of T2/Onc within intron 1 of *Tgif2*. E - Represents the exon; half arrows represent the directions and positions of the inserts. Annotated tumours according to their distance are included. (B) Products amplified using primers from the right side of T2/Onc and Exon 2. Tumour id. and approximate product sizes are indicated. H2O = no template.

To confirm the integrity of the exonic primer, cDNA templates from all tumours were also analysed using the exon 2 primer in combination with a primer from exon 1 (Figure 4.3). The expected amplicons were generated in all the samples.

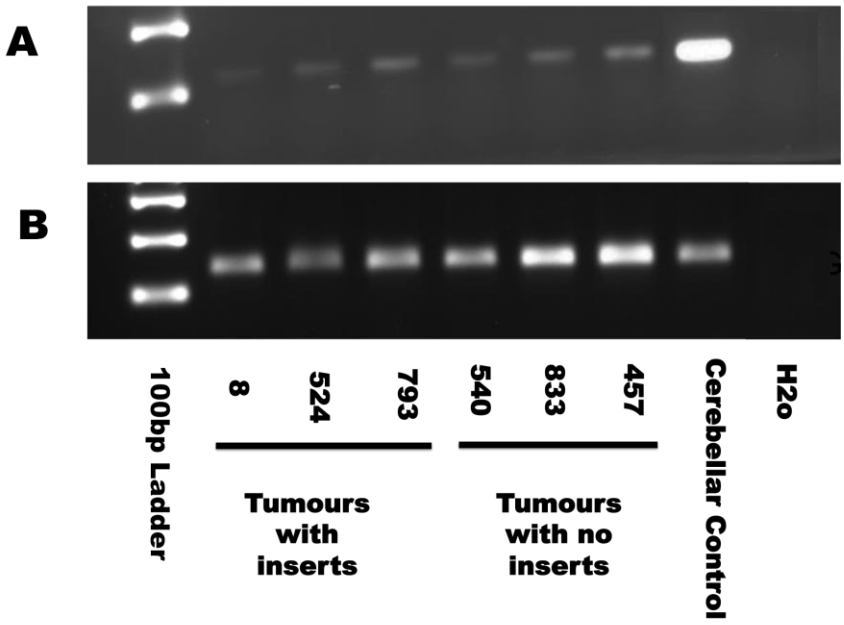

Figure 4.3. Expression of *Tgif2* in MB. (A) Expression of *Tgif2* within MBs with inserts and with no inserts, assayed using primers from exons 1 and 2. Cerebellum control sample, and no template control (H2O) are also shown. The amplicon size of *Tgif2* is ~117 bp. (B) control *Gapdh* (~131 bp) amplification products.

4.2.1.1.b Validation of the inserts within Nfia

Nfia was the gene most frequently targeted by SB insertions in the MB dataset with 27 independent transposition events, and the majority of the inserts were in the antisense orientation (relative to *Nfia* transcription) (Figure 4.4 A). To confirm insertional events within this gene, genomic PCR was performed on one MB tumour (MB-P-540) and one wild-type cerebellar control, as shown in Figure 4.4 B; a product of the expected size was amplified from the primer pair flanking the junction between an intronic primer and the right IR/DR primer (3+1) in tumour with insert. Moreover, the expected ~200 bp amplicon was obtained from the wild-type allele using intronic primers (1+2). This finding provides additional evidence for to the insertional validation series within the MB samples.

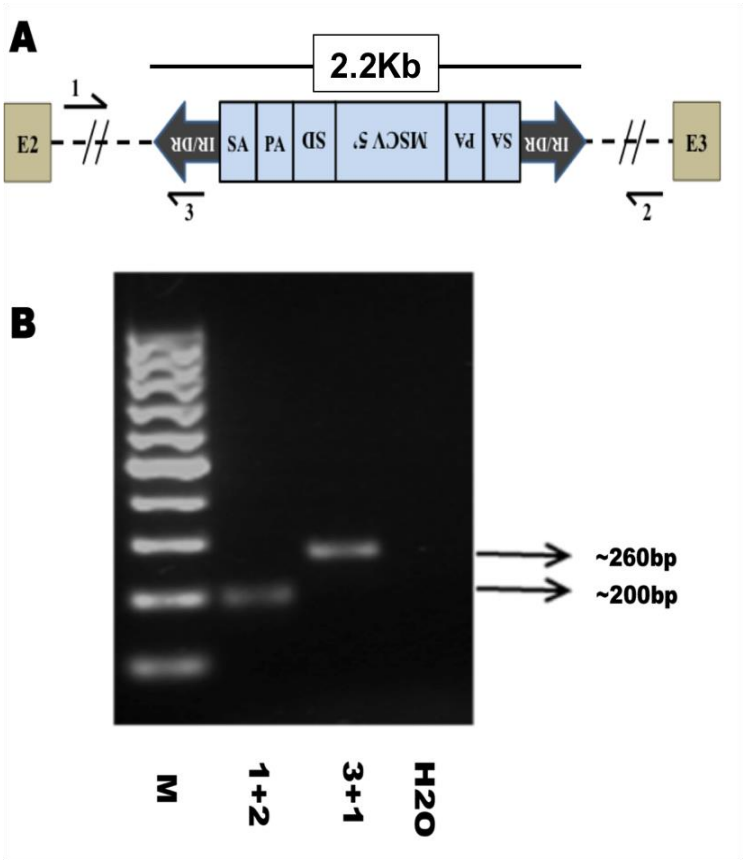

Figure 4.4. Validation of genomic insert within *Nfia*. **(A)** Schematic diagram representing the location, size (2.2Kb)and orientation of T2/Onc within intron 2 of *Nfia*. The half arrows indicate the primer positions within the intron and T2/Onc. E2 and E3 represent the exons. **(B)** Amplification products obtained using the intronic primers. Primers 1+2 generate a product from the wild type Nfia allele (expected size ~200 bp). Use of intronic primer 1 with the primer from the right side (IR/DR) of T2/Onc generates a product when SB is inserted in the negative strand giving a product which is larger than the normal allele (primers 3+1, expected size of ~260 bp). A water sample was used as a negative control (H2O)

4.2.1.1.c Validation of Egfr genomic inserts within liver adenomas

As a further check of the insert site data, *Egfr* was also analysed within twelve liver adenomas by genomic PCR. All the inserts were located and distributed within intron 24 of *Egfr* and all were negatively (antisense) orientated with respect to the gene (Figure 4.5A). A specific endogenous forward and reverse primer pair, spanning all the inserted transposons within intron 24, was used to amplify the wild-type intronic region. The expected 713 bp amplicon was successfully amplified from 12 liver nodules with inserts, and from six liver nodules without inserts (Figure 4.5B). The endogenous

forward primer and the transposon-specific IR/DR(R) primers were then used to amplify the transposon junctions. Amplicons of the expected sizes were amplified successfully from all the adenomas with inserts, with the expected sizes varying according to the position of the inserted T2/Onc element (ranging from ~195 - ~813bp) (Figure 4.5C). Furthermore, two samples, 277 and 751, generated two different product sizes corresponding to two different insertions at different positions within these adenomas. The other six liver nodules without insertions (630, 631, 262, 571, 312 and 207) were included to confirm the integrity of the assay, and they showed no amplification. It is noteworthy that seven of the samples included insertions which were removed by the second statistical analysis as potential false positives as they were inserted into the same dinucleotide, leaving only a total of five totally independent insertion sites. These results establish that in this case the T2/Onc insertions in the same dinucleotide within different samples are genuine independent insertions at identical genomic locations and not artefacts. Moreover, the distribution of inserts within *Egfr* was focused within intron 24 in agreement with a previous SB screen analysis, and all were not in the same genic orientation (Keng *et al.*, 2009), providing further evidence that *Egfr* is a key gene in hepatocellular carcinoma.

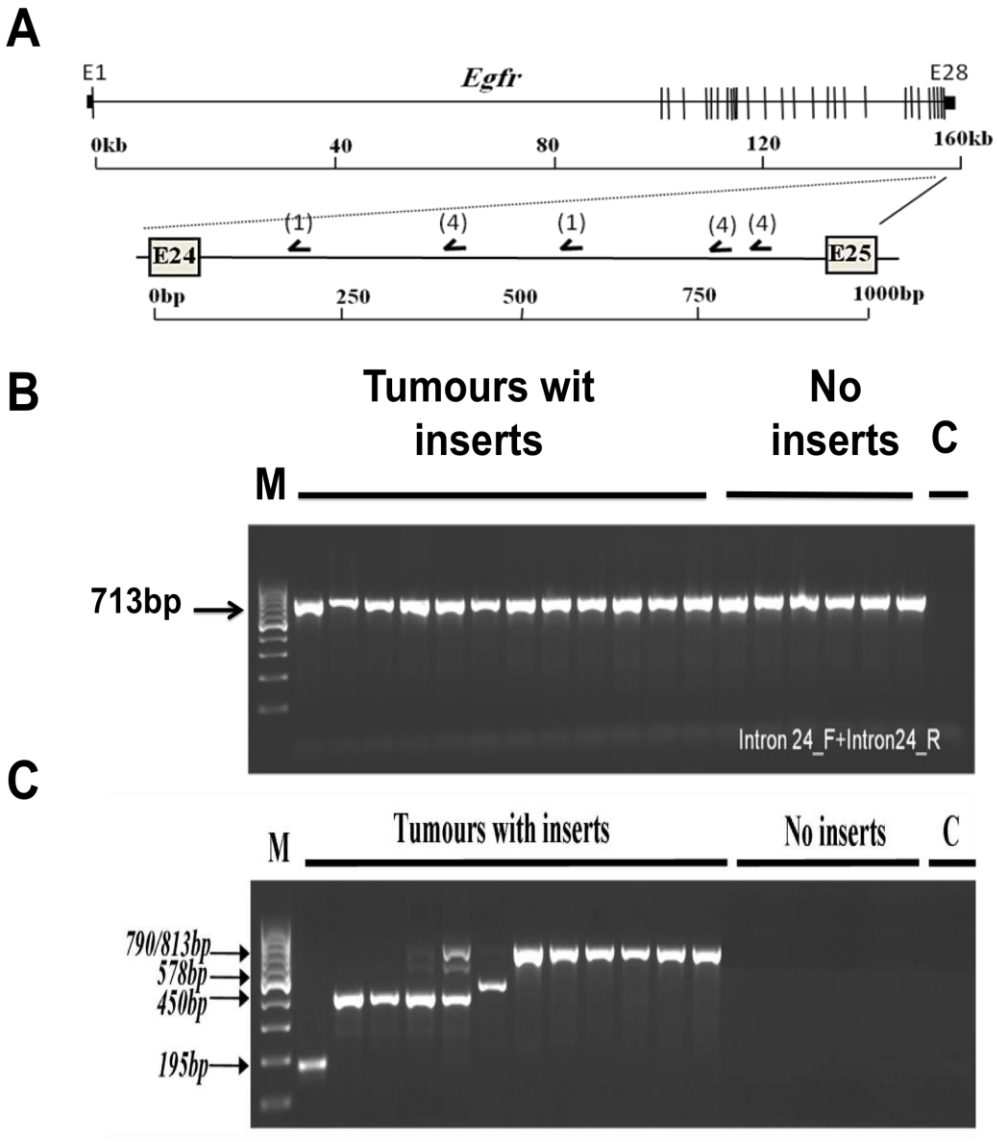

Figure 4.5. Identification of transposons within the *Egfr* gene in liver adenomas. (A) Transposons number and orientation within intron 24 of *Egfr*. (B) Products amplified by the endogenous primers of intron 24 (713 bp) from both liver adenomas with inserts and liver adenomas without inserts. (C) Product amplification between the endogenous forward primer from Egfr intron 24 and the T2/Onc forward primer; the product sizes vary according to the position of T2/Onc within the gene.

4.2.1.2 Validation of structural impact upon transcripts

The genomic PCR above confirmed the integrity of the SB insertion site data generated by FLX sequencing. From the *in silico* analysis and the orientation bias of the inserts, it is clear that chimeric or truncated transcripts are expected to underpin the biological impact of the candidate CISs. To confirm the existence of chimeric transcripts in these genes, sets of primers were designed to amplify the exonic and

transposon junction at the transcriptomic level using RT-PCR. The results for *Tgif2* and *Egfr* are shown in Section 4.3.1.2.a and Section 4.3.1.2.b below.

4.2.1.2.a Identification of chimeric transcription from Tgif2

As shown in Figure 4.6 A and B, expression of the wild type *Tgif2* allele was confirmed (320 bp) using E1 and E2 primers in both inserted and non-inserted tumours, as well as in control cerebellum samples (with very low expression). In contrast, amplification between an E1-specific forward primer and a T2/Onc Splice Acceptor reverse-specific primer was only observed in one of the samples, indicating low levels or no transcriptional integration between exon 1 and the downstream T2/Onc element in tumours with inserts (Figure 4.6C). However, amplification between the T2OncSD and *Tgif2* E2 primers did occur (~230 bp) in the four tumours where genomic inserts were validated. This confirms that expression of the allele containing the inserts is being driven by the promoter within T2/Onc. No amplification was observed in tumours without inserts, or in the control cerebellum samples using these primers (Figure 4.6D). It is noteworthy that tumours with no inserts also show *Tgif*2 expression, suggesting that this gene may be upregulated in these tumours via an SB-independent mechanism.

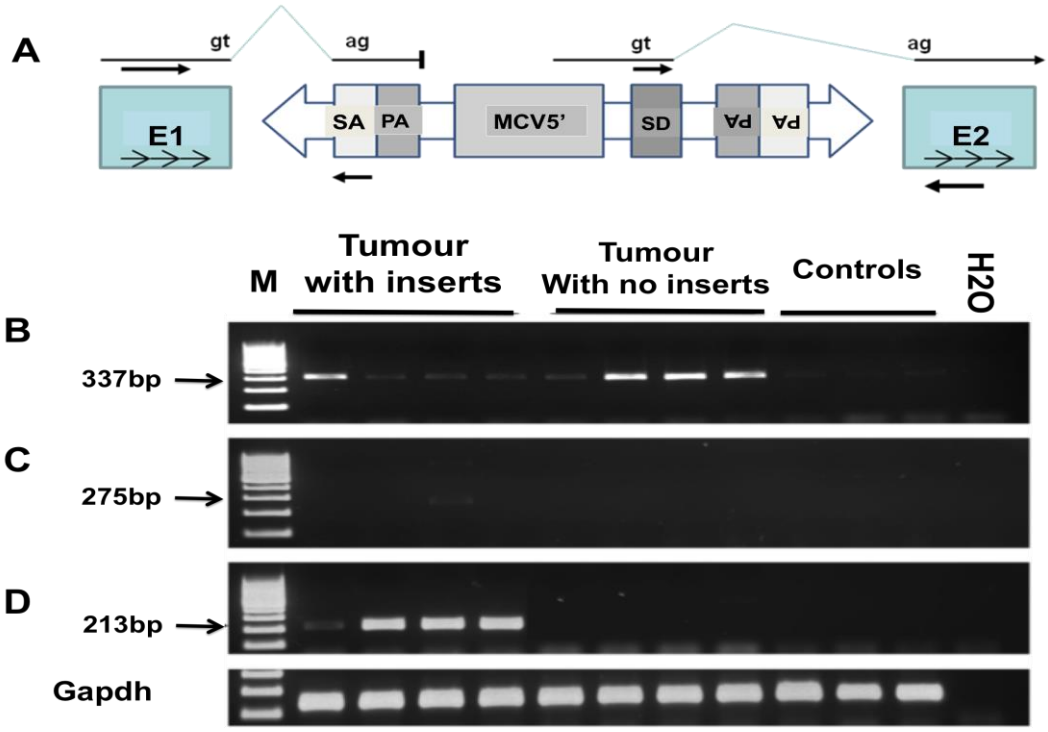

Figure 4.6. Validation of chimeric transcriptional events for *Tgif2* using reverse transcriptase PCR. (A) Primers between T2/Onc and *Tgif2* exons, insert orientation, and the hypothetical transcriptional process of *Tgif2*. (B) Amplification of the exonic primers spanning exon 1 and exon 2 of *Tgif2*: the products sizes are 337 bp. (C) Amplification between SA of T2/Onc and exon 1 of *Tgif2*. (D) Amplification between SD of T2/Onc and exon 2 of *Tgif2*, confirming the integration of T2/Onc within the transcriptional process of *Tgif2*. The last panel shows *Gapdh* (131 bp) amplification for all the samples used.

4.2.1.2.b Identification of chimeric transcripts from Egfr

Analysis of *Egfr*, which was identified as a common insertion site within liver adenomas, is shown in Figure 4.7. The expression of the wild-type allele was confirmed (183 bp) using primers from the 24th (E24) and 25th (E25) exons of *Egfr* in both inserted and non-inserted adenomas (Figure 4.7 B). In tumours with inserts in intron 24, transcripts were captured between E24 and the polyadenylation (PA) signal of the transposon (393 bp, Figure 4.7 C), suggesting that a truncated protein is the outcome of insertion. Comparable results were also reported in a previous Sleeping Beauty study that sought to identify the cancer genes involved in hepatocellular carcinoma (Keng *et al.*, 2009).

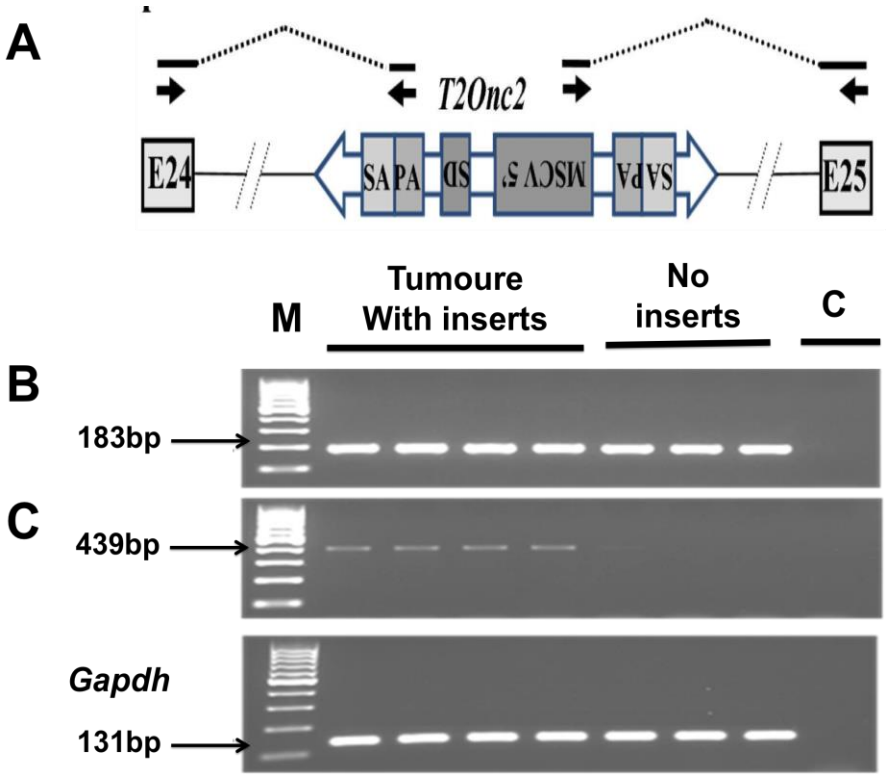

Figure 4.7. Validation of chimeric transcriptional events for *Egfr* using reverse transcriptase PCR. (A) Schematic showing the primers between T2/Onc and the *Egfr* exons 24, insert orientation, and the hypothetical transcriptional process of *Egfr*. (B) Amplification using the exonic primers spanning exon 24 and exon 25 of *Egfr*; the expected product size is 183 bp. (C) Amplification between PA of T2/Onc and exon 24 of *Egfr*. The expected product was amplified from tumours with inserts (~393 bp) confirming the integration of T2/Onc within the *Egfr* transcripts. The last panel shows *Gapdh* amplification.

Overall, these results validate the genomic insertion sites, and confirm the existence of inferred transcriptional products which are consistent with a specific mode of action upon each gene (upregulation for *Tgif*2 and truncation for *Egfr*).

4.2.2 Quantification of the Impact of Transposition on Gene Expression

The above analyses confirmed the presence of chimeric T2Onc/target gene transcripts: in several cases upregulation of the entire open reading frame is the inferred outcome of insertion when inserts are clustered upstream of the exon containing the translation start site, while downregulation of exons 3' of T2Onc insertions is inferred for other genes. However, to fully validate the inferred mode of action it is necessary to quantify gene expression for specific genes. Three MB CISs, namely *Tgif2*, *Myt1l* and *Tead1*, were chosen for analysis using TaqMan® qPCR.

4.2.2.1 Quantification of Tgif2 expression

Initially, a real-time PCR assay was calibrated using tumour samples with *Tgif*2 inserts to measure the efficiency of the assays and to select appropriate concentrations of sample, primers and probe (see Methods). For each sample, triplicate reactions were carried out to measure the average cycle threshold (Ct) value per sample. The average relative expression for *Tgif2* was determined using the ΔΔCt method and normalised to a control gene *(β-Actin*) and to the median ΔCt value obtained from six wild-type cerebella controls. As shown in Figure 4.8, *Tgif2* expression in the tumours with inserts (pink) is appreciably higher than it is in the tumours without inserts. Moreover, the differential expression of *Tgif2* between the inserted tumours and the tumours without inserts using qPCR was statistically significant (t-test p = 0.025). The differential expression was statistically highly significant between the tumours with inserts and the wild-type tumours (p = 0.00051). The significantly higher expression of *Tgif2* within the inserted tumours, as compared to the tumours without inserts, is consistent with the transposition having driven the expression upwards. Additionally, these results suggest that *Tgif2* may exhibit oncogene-like behaviour and that it may play a role in MB formation.

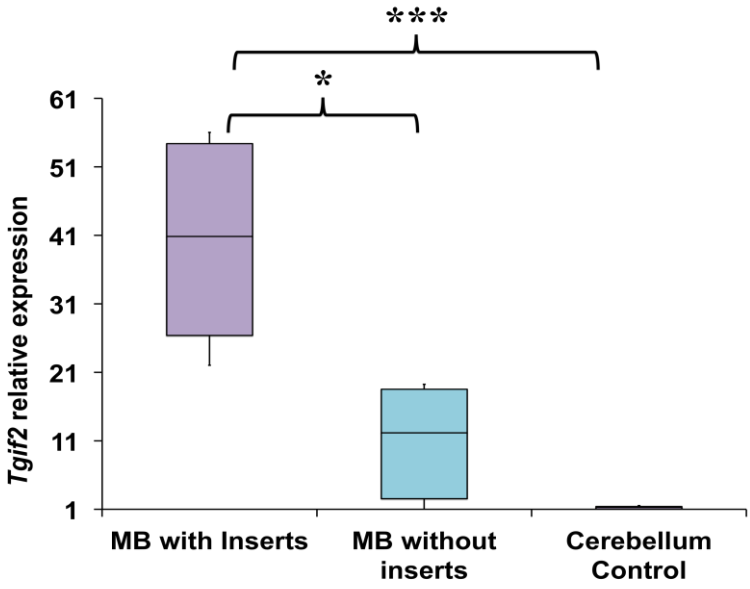

Figure 4.8. Expression of *Tgif2* in MBs. Relative expression of *Tgif2* analysed by qPCR in SB mutagenized MB mice. Relative expression was calculated according to the median expression of the six cerebellum controls. * (p=0.025), *** (p=0.00051).

4.2.2.2 *Quantification of Myt1l expression*

As *Myt1l* plays a major role in neuronal development (Kim *et al.*, 1997), the impact of insertion upon expression levels may have an impact on tumour development. The inferred mode of action of *Myt1l* was downregulation, as most of the inserts were distributed evenly within the gene and in reverse orientations to the gene. *Myt1l* expression relative to wild-type cerebella controls was determined (Figure 4.9) using ΔΔCt methods, normalised to a reference gene (β-Actin), but no statistically significant difference was observed (p = 0.59). However, as expected in all seven of the MB tumours, *Myt1l* was expressed at very a low level compared to the controls.

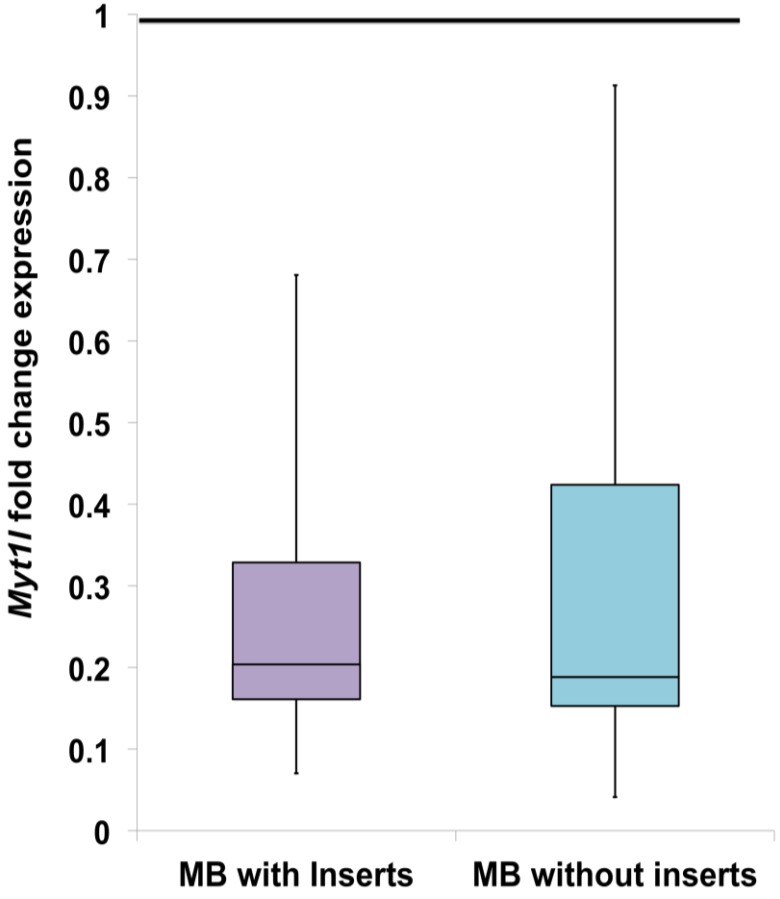

Figure 4.9. *Myt1l* relative expression by qPCR. Differential expression of *Myt1l* within tumours with inserts (purple bars), and within tumours without inserts (blue bars). The expression was not significantly different (p = 0.59) between the tumours with inserts and the tumours without inserts. The reference control is set to 1 and indicated by a thick line.

4.2.2.3 Quantification of Tead1 expression

The inferred mode of action for *Tead1* expression was downregulation, based on the position and orientations of the inserts (see Table 3.6). Figure 4.10 presents the expression of the *Tead1* gene relative to the reference cerebellar controls in MB tumours with inserts using the ΔΔCt method, normalised to an endogenous gene (B-Actin). The higher expression of *Tead1* within both the five MB tumour samples with inserts (purple box) and the five MB tumour samples without inserts (blue bars) was statistically highly significant (p = 1.2E-05) compared to the expression in the cerebellum controls (indicated as the base line of 1). Furthermore, the expression of *Tead1* in the tumours with inserts is 2.1X higher that the expression in the tumours without inserts, and this difference was also statistically significant (p = 0.04).

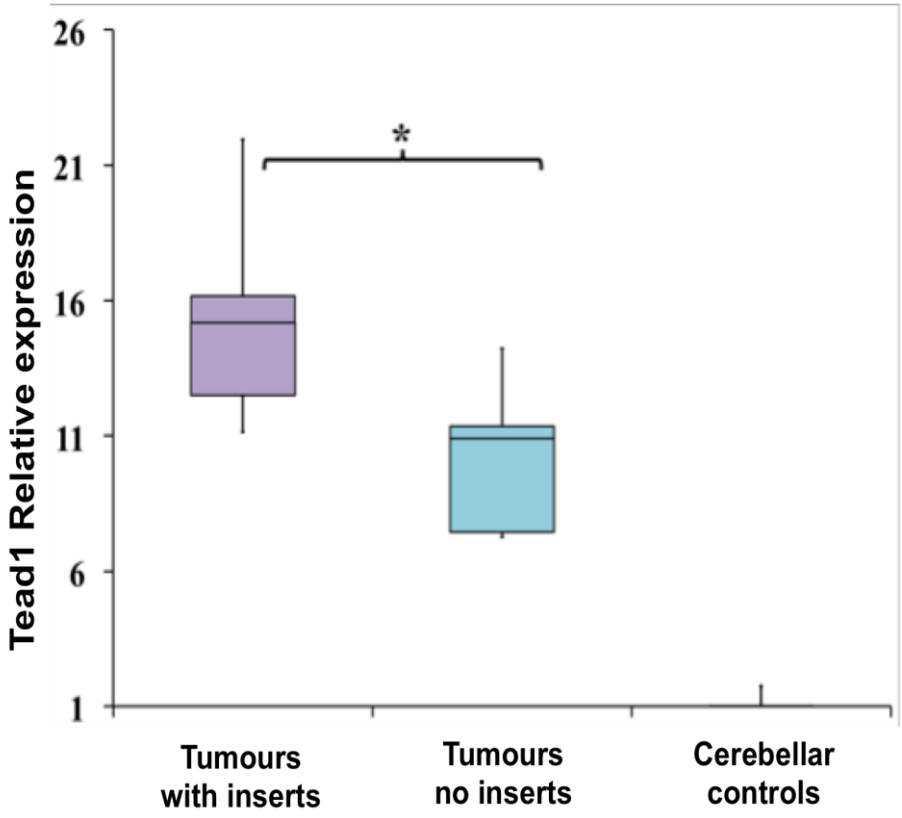

Figure 4.10. Relative expression analysis of the *Tead1* gene in the MB samples and cerebellum controls. The expression of the *Tead1* gene was significantly upregulated in MB tumours (inserted or without insertional mutation) compared to the cerebellum controls (p = 1.2E-05). Tead1 expression is also significant between tumours with inserts and tumours without, as shown in the figure (* p= 0.04). The average ΔCt of the cerebellum controls was set to 1 for presentation purposes.

The aforementioned results show that the use of qPCR assays to measure expression levels in tumours with inserts and tumours without inserts was highly informative in the case of the *Tgif2* gene. The high expression in tumours with inserts is consistent with the inferred MOA and existence of chimeric transcripts. Moreover, consistency with the inferred MOA was also checked in two other CISs, namely *Tead*1 and *Myt1*l. The former also showed evidence of upregulation, consistent with the inferred mode of action. Although no evidence for downregulation was observed for *Myt1*l, the inferred mode of action is loss of function due to truncation. Overall, this study's expression results are consistent with the identified CISs directly affecting transcription of the genes into which they insert, consistent with a role in tumour formation.

4.2.3 CISs are neuronal developmental regulators

After the inferred mode of action was confirmed for three genes using qPCR, the question remained as to whether or not these CISs have any biological function that is consistent with the development of cancer. To examine and potentially answer this question, Gene Ontology (GO) analysis was performed using the Generic Gene Ontology Term Finder (Boyle *et al.,* 2004). This revealed that the identified MB CISs (Table 4.1) are enriched for genes involved in neuronal development (8/18 CIS genes; $p = 0.0003$), and transcriptional factors (7/18 CIS genes; $p = 0.00002$) or regulators of gene expression (9/18 CIS genes; $p=0.00003$). All 8 CIS genes implicated in neuronal development are transcriptional regulators (*Atxn2, Tead1, Tgif2, Crebbp, Nfib, Pten, Slit3* and *Gli3*). It is noteworthy that several genes were identified as being implicated in cellular growth and/or fate determination (data not shown). This suggests that the CISs are involved in fundamental cellular developmental processes. As a result, it appears plausible that any imbalances in expression could lead to tumourgenesis.

4.2.4 CISs are Differentially Expressed between MB subgroups.

As the inferred mode of action was generally consistent with qPCR for the tested genes, and the results of the ontology analysis revealed important roles for CISs genes in neuronal development, I next examined the expression of CIS genes within human medullobastoma (MB) to establish whether MB CIS genes are differentially expressed in human clinicogenetic subgroups. To achieve this, I investigated the differential expression of human orthologs of CIS genes within the data generated by Kool *et al.,* (2008) and subsequently re-grouped by Northcott *et al.,* (2011). This data

consists of 63 human MB samples that exemplify the currently defined four different MB clinicogenetic subgroups—WNT, SHH, Group 3 and Group 4. Their expression patterns are presented in Table 4.1. Interestingly, 15 of the 19 CISs genes showed differential expression in one or more clinicogenetic subgroups, and strikingly, 4 genes show differential expression in the SHH subgroup only (*ATXN2*, *CREBBP*, *TEAD1* and *ITGBL1*). However, some of the CISs show significant differential expression in non-SHH subgroups, such as *NFIA* differentially expressed in Group 3, *FGF13*, *SLIT3* and *PTEN* differentially expressed in group 4 only tumours, and *NFIB* which was differentially expressed in WNT subgroup tumours only. Finally, some CIS genes were expressed differentially in more than one subgroup beside SHH subgroup, such as (*TGIF2*, *GLI3*, *MYT1L*, *ADCY5* and *L3MBTL4*).

To view and compare the expression of CISs between the human MB clinicogenetic subgroups, a heatmap of differential expression of CISs using the method recommended by Kool *et al.*, (2008) was generated (Figure 4.11). As expected, the majority of the CISs show clear differential expression between one or more of the human MB subgroups. Within the SHH subgroup tumours, it is striking to note that the genes show expression which is generally consistent with the inferred MOA. For instance, *MYT1L* is downregulated in the SHH subgroup, while *TEAD1*, *TGIF2* and *CD2AP*, *ANKRD5*, *ITGBL1* and *NFIB* are upregulated within the same tumours. As expected, the expression of *GLI3*, an effector of SHH signalling, was upregulated within the SHH subgroup as the SHH pathway was stimulated.

Because several genes showed differential expression within the SHH human MB subgroup, this finding suggests that these CISs may interact with the *PTCH* gene, and, consequently, may contribute to MB formation within the current mouse model. However, the fact that some of the other genes were differentially expressed within other subgroups suggests that those genes may have a wider importance in MB development.

Table 4.1. CIS genes, enriched ontologies, and their differential expression within human MB subgroups. (ORF) open reading frame; (Up-R) upregulation; Arrows show the gene expression depending on the direction and relevance to the subgroup colour, (↑) up-regulation, (↓) down-regulation (expression data taken from Kool et al., 2008) .

Gene	Inferred MOA	Gene Ontology			Subgroups with Differential Expression	Expression
		Neuronal Development	Transcription Factor	Regulation of Expression		
Nfia	ORF disruption		✓	✓	**C**	↓
Dscr3	ORF disruption				**-**	
Atxn2	Unclear	✓			**SHH**	↓
Tgif2	Up-R	✓	✓	✓	**SHH/D**	↑/↓
Crebbp	ORF disruption	✓	✓	✓	**SHH**	↓
Tead1	Up-R	✓	✓	✓	**SHH**	↑
Fgf13	Up-R				**D**	↑
Nfib	Up-R	✓	✓	✓	**Wnt**	↓
Itgbl1	Unclear				**SHH**	↑
Pten	ORF disruption	✓		✓	**D**	↑
Sfi1	ORF disruption				**-**	
Slit3	ORF disruption	✓		✓	**D**	↑
Gli3	ORF disruption	✓	✓	✓	**SHH/D**	↑/↓
Myt1l	ORF disruption		✓	✓	**Wnt/SHH/ D**	↓/↓/↑
Tmem45b	Unclear				**-**	
Adcy5	ORF disruption				**C/D**	↓/↑
L3mbtl4	ORF disruption				**SHH/D**	↑/↓
Ankrd5	unclear				**SHH/C**	↑/↓
		42.90%	38.90%	47.60%		
		p=0.0003	p=0.00002	p=0.00003		

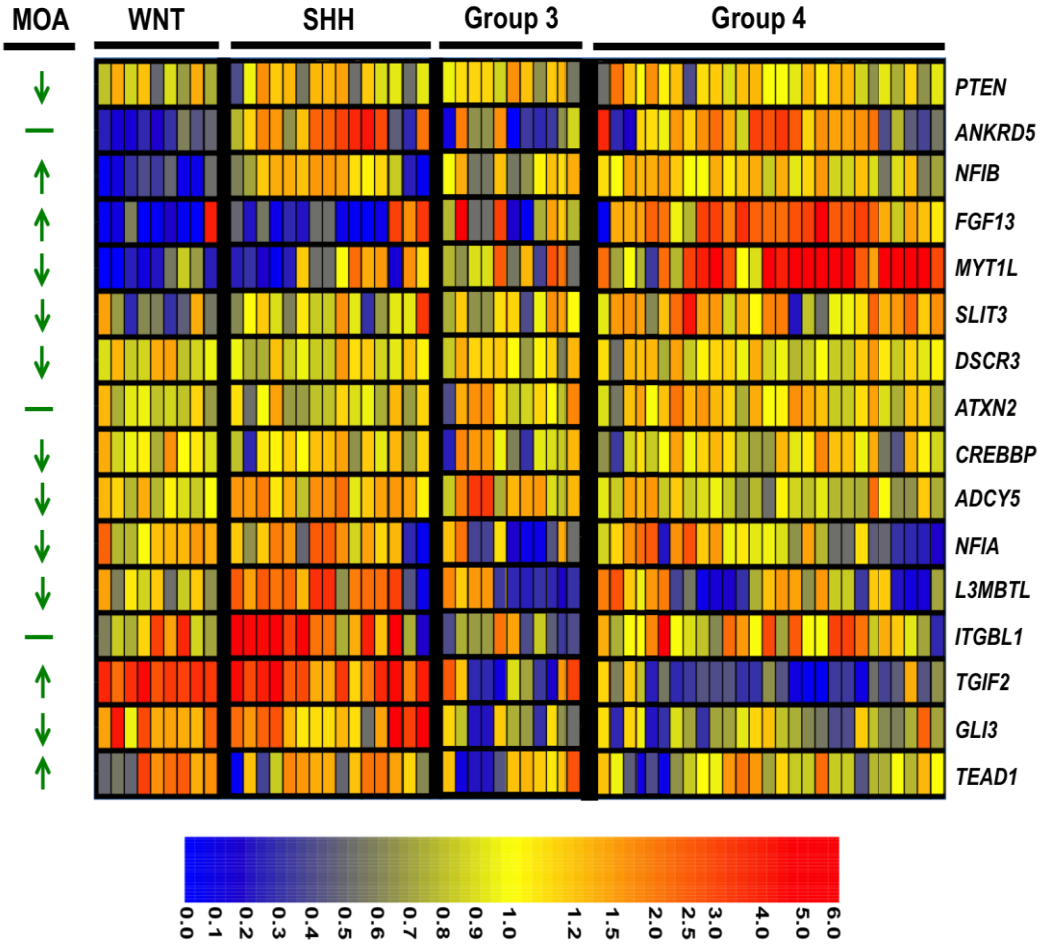

Figure 4.11. Heatmap of CIS gene expression in human MB clinicogenetic subgroup tumours. The probe with the highest mean expression levels across all samples is presented. MOA represents the inferred mode of action with the SB insertion; the arrows represent the upregulation or downregulation pattern and dashes represent unclear interpretation. Two genes,*Tmem45b* and *Sfi1*, are not represented within the Affymetrix chip used. A colour bar representing specific expression levels is shown. Expression data taken from Kool et al., (2008).

4.3 Discussion

In this chapter I confirmed the presence and impact of transposon insertions within a subset of CISs identified in the SB screen. This was achieved at both the levels of genomic sequence and chimeric transcripts. Q-PCR analysis was then employed to confirm the inferred mode of action of inserts upon *Tgif2, Tead1* and *Myt1l*. Ontology analysis was also performed, and this identified a significant enrichment of neuronal and transcription factor CISs.

To validate the bioinformatic identification of CISs, genomic PCR was first applied to three genes from different tumour types, *Tgif2* and *Nfia* from the MB dataset and *Egfr* which was identified as the top CIS in liver adenomas. Despite the depth of next generation sequencing, CIS identification is error-prone, and this was exemplified by the *Tgif*2 analysis where an insertion in one of the MB samples proved to be a false positive. Additionally, chimeric transcripts were successfully identified for *Tgif2* and *Egfr* using reverse transcriptase PCR, confirming the presence of T2/Onc fusion transcripts from both genes. These results support the inference that the effect of insertion on gene function was upregulation in the case of *Tgif2* and truncation of the *Egfr* gene product.

A previous study proved qPCR's superior sensitivity in comparison to that of the Illumina microarray (Guo *et al.*, 2010). To confirm the impact of SB insertion upon gene expression, the *Tgif2, Tead1* and *Myt1L* genes were all analysed by qPCR. *Tgif2* expression was found to by significantly higher within the tumours with inserts as compared to the tumours without inserts (2.1X), consistent with *Tgif2* exhibiting oncogene-like behaviour. This is the first time this gene has been implicated in MB. *Tgif2* is a member of TG-interacting factor gene family (TGIFs), which belongs to the TALE superfamily of homeodomain proteins (Imoto *et al.*, 2000). *Tgif2* has the ability to repress Smad-mediated transcription through interaction with TGF-β-activated Smad proteins (Melhuish *et al.*, 2001). *Tgif2* expression has been detected in the developing brain tissues of mouse embryos, specifically within the proliferating cerebellar neuroepithelium (Bertolino *et al.*, 1996). In a recent study conducted by Taniguchi et al. (2012), both *Tgif*1 and *Tgif*2 have been reported to be involved in Holoprosencephaly (HPE) in mouse. This occurs when mutations of *Tgif*1 and 2 inhibit *Shh* signalling and cause malformation of the forebrain of mouse models. The authors concluded that the mutation of *Tgif* family members in mouse is involved in HPE, and suggested that this occurs due to Shh signalling pathway interruption.

I also showed that the *Tead1* gene was significantly upregulated within the MB samples, as compared to the controls, with a significantly higher expression within tumours with inserts compared to tumours without inserts. This finding is also consistent with the MOA inferred from the CISs *in silico,* as all inserts are clustered at the 5' end of the gene consistent with upregulation. *Tead1* is a family member of the *TEAD* (TEA domain) group of transcription factors that possess a DNA-binding domain (Anbanandam *et al.*, 2006). The activity of this gene depends on a transcriptional coactivator called *YAP* (Cao *et al.*, 2008;Sawada *et al.*, 2008). *YAP* is known to be the downstream effector of the Hippo pathway, and recent studies have demonstrated its role in controlling cellular proliferation in *Drosophila* (Fernandez *et al.*, 2009). Cao *et al.*, 2008, demonstrated that upregulated *YAP* and *TEAD1* promote the growth of neuronal progenitor cells via the activation of the cell cycle through the induction of *cyclin D1*. Moreover, they provided evidence that both these genes have the ability to disrupt progenitor cell differentiation via the suppression of *NeuroM.* Importantly, a recent study investigated the role of *YAP1* and *TEAD1* in MB subgroups. Fernandez et al. noted that the overexpression of both *YAP1* and *TEAD1* was associated with the SHH MB subgroup. This is congruent with our identification of *Tead1* as a candidate cancer gene in MB.

Myt1l, myelin transcription 1-like, is a zinc finger gene (*MYT1* or *NZF3*) and belongs to the *Myt1l* family. In rodents, it has been shown to be predominantly expressed in the developing brain. The highest expression of this gene was recorded in the brain and spinal cord in adult rodents, whereas no expression was observed in the glial cells, suggesting that *Myt1l* is strictly expressed in neurons (Kim *et al.*, 1997). Despite the significant difference in *Tgif2* and *Tead1* expression between tumours with inserts and tumours without inserts, no difference could be found for *Myt1l.* However, *Myt1l* expression was significantly lower in all tumours compared to the controls and this finding is consistent with the inferred MOA. The most plausible explanations for this finding are either that the expression of the disrupted gene is irrelevant, and that knock out of one allele through truncation of the protein is the MOA in this case (perhaps interfering with function of the wild type protein), or that reduction in *Myt1l* expression has been achieved by an SB-independent mechanism in tumours with no inserts in this gene. It is also possible that the expression of *Myt1l* within the MB precursor cell, rather than the mature tumour, is critical. Although these are all plausible explanations, distinguishing between them would be technically challenging and time consuming, as MB insertions within the gene would have to be engineered (or be

germline) to allow straightforward analysis. As the *Myt1l* CIS is highly significant, and *Myt1l i*s a known pan neural differentiation gene, it remains a strong candidate gene despite the lack of a clear impact of insertion upon expression in mature tumours.

Primers were also generated for *Nfia* and *Tead1*, and these genes were also analysed at the transcriptional level (data not shown). These genes have an insertion bias opposite to their orientation, as the genes are positively orientated and the T2/Onc inserts are negatively orientated. No evidence of gene transcripts could be found linking the nearest exons and transposon from either site (data not shown). This might be due to the fact that the inserted transposons have an effect on the gene function without contributing to the transcriptional process. Alternatively, the impact on the transcripts might have occurred in the developing tumours prior to sampling.

It is important to note that nine of the 19 CISs identified in this current study (*Atxn2, Tgif2, Crebbp, Tead1, Itgbl1, Gli3, Myt1l, L3mbtl4* and *Ankrd5),* were reported to have differential expression within the human SHH MB subgroup in at least one of two recent tumour series analysed (Kool *et al.,* 2008;Northcott *et al.,* 2011). This provides further evidence that these genes could have a significant impact on the generation of medulloblastoma.

In conclusion, the insertional events, and the significant impact they have had on the expression of genes currently identified, suggest that many of the genes identified are plausible candidates for involvement in MB. Additionally, Gene Ontology investigation revealed that CISs are enriched for transcription factor genes which contribute to neuronal development and control cellular outgrowth. Moreover, It is clear from the discussion of individual CIS gene functions that several are very suitable candidate genes for MB development. However the ontology analysis has established that many genes have similar or overlapping functions in developmental processes within the neuronal system. This raises the possibility that the individual genes mutated affect the same or convergent developmental pathways which are key to SHH subgroup MB development. If so, identification and understanding of these pathways may shed light on tumour development and allow rational targets for therapeutic intervention to be identified. A more detailed investigation of these genes within MB tumours, and the possibility that they work in co-ordinated developmental pathways, is presented in the next chapter.

Chapter 5. (Results 3)

MB CISs define a network involved in neuronal development and human disease

5.1 Introduction

Brain and Central nervous system tumours comprises the second most prevalent cancer in childhood after leukaemias (Cancer Research UK, 2013b), and MB accounts for about 20-30% of the CNS in children (Li *et al.,* 2013). Cancer genome sequencing and microarray expression analyses have identified a minimum of four molecular clinicogenetic subgroups of human MB; known as SHH, WNT (where SHH and WNT signalling pathways respectively have been critically disrupted), Groups 3 and Group 4 (Tylor *et al.,* 2012; Robinson *et al.,* 2012).

Studies conducted on mice revealed that the SHH MB subtype arises from granule neuron precursor cells (GNPCs) of *Ptch1*$^{+/-}$ mice, that Wnt subtype tumours can originate from lower rhombic lip progenitors (LRLP) cells of *Blbp-Cre;Ctnnb1*$^{+/lox(Ex3)}$*;Tp53*$^{flx/flx}$ mice, and that Group 3 MB tumours are likely to arise from as yet undefined cerebellar progenitor cells. A Group 4 tumour cell of origin has yet to be identified (Schuller *et al.*, 2008;Gibson *et al.*, 2010;Kawauchi *et al.*, 2012). These data are consistent with the widely held view that Medulloblastomas, like many paediatric cancers, arise due to the breakdown of normal developmental processes (Ellison, 2010).

In the previous chapters I presented the results from a whole body SB mutagenesis screen of the *Ptch1*+/- murine model of MB which identified candidate genes for involvement in disease (Table 3.6), and showed that the majority of these are differentially expressed in one or more MB subgroup. I also established that the candidate genes are significantly enriched for transcription factors and genes involved in neuronal development. This observation is consistent with the hypothesis that the neuronal transcription factor CISs affect key signalling pathways within the developing cerebellum, and that their disruption by SB enhances MB formation. As transcription factors act within carefully regulated cascades often with complex feedback mechanisms and interactions (Vaquerizas *et al.*, 2009),it is possible that the transcription factors within the CIS list are influencing one, or a small number of

interacting, developmental pathways in the *Ptch1*+/- model. Before analysing any candidates experimentally, it was decided to investigate this possibility by analysing microarray gene expression data generated from primary human MBs.

A dataset generated from 72 paediatric MB analysed by comparative genomic hybridization (CGH) and gene expression assay (Fattet *et al.,* 2009) is publically available. The main focus of this study was the investigation of the role of WNT/β-catenin pathway in MB and to identify a general and effective biomarker for this lesion. β-catenin protein was expressed in more than 50% of the investigated cases. However, mutated *CTNNB1* was extensively expressed in the nucleus of all β-catenin positive samples. Furthermore, this study analysed the clinical features of the MB patients, and longer survival rates were recorded in patients with *CTNNB1* positive tumours compared to the negative tumours (mean 5-year of overall survivals of 53.7%). The data from this study was combined with those generated by Kool *et al.,* (2008) by Dr Dan Williamson (Newcastle University) and used in the analyses presented in this chapter.

A variety of techniques have been developed to identify putative gene interactions within complex expression datasets (Elkon *et al.*, 2003)(Elkon et al., 2003). One of the most widely used is ARACNe (algorithm for the reconstruction of accurate cellular networks), which was specifically developed to identify direct gene-gene or gene-protein interactions of the type expected to exist between transcription factors and their targets. This method can create gene interaction networks by analysing expression data from large datasets (Carro *et al.*, 2010), and is characterised by its ability to overcome the limitations of some other algorithmic methods such as correlation methods. For instance, it possesses a low computational complexity, uses the full range of expression data, and exhibits a very low false positive rate (Margolin *et al.*, 2006). ARACNe has been successfully used to identify two transcription factors (*C/EBPβ* and *STAT3*) acting as synergic initiators and the main regulators of mesenchymal transformation in human malignant glioma (Carro *et al.*, 2010).

For these reasons, ARACNe was chosen for the analysis of MB CISs in human expression data. The availability of clinical data from one of the human datasets (Fattet *et al.,* 2009) also meant that the relationship between individual genes or CIS networks and survival/clinical features could be investigated.

5.1.1 Aims

5.1.1.1 To generate a human MB gene expression network using publically available microarray gene expression data from primary tumours.

5.1.1.2 To investigate the relationship between CIS genes in human data sets through the analysis of the network and presentation of heat maps.

5.1.1.3 To use bioinformatic analysis to establish the relationship between CIS gene expression and survival in human datasets.

5.2 Results

5.2.1 MB CISs define a neuronal transcription factor gene network

To investigate the possibility that the CISs presented in chapter 3 are functionally linked within MBs, an ARACNe analysis was performed by Dr D Williamson (NICR) using the merged data of Kool *et al.*, (2008) and Fattet *et al.*, (2009) which consisted of data from 119 MB human tumours generated using the Affymetrix HGU133 Plus 2.0 genechip. Briefly, the method involves the assembly of a genome-wide collection of gene expression relationships, and systematic removal of indirect relationships by the algorithm, to reverse engineer a single network of putative TF-Target relationships. Affymetrix probes were filtered down from the original ~500,000 by inspecting for probes that were expressed in at least 10% of the samples with a range greater than 500 and an overall fold change of 3. Then, only probes with a significant mutual information (MI) association with at least one other node, connected to each other with a bootstrapped *P*-value of <10E-8, were retained. This left 8,749 probes in the network. To eliminate possible bias introduced by merging data from two different CIS calling algorithms, only the 17 MB CISs identified using the Genome Kernel Convolution method were analysed. The Cytoscape representation of the full Affymetrix network generated by ARACNe analysis is shown in Figure 5.1.

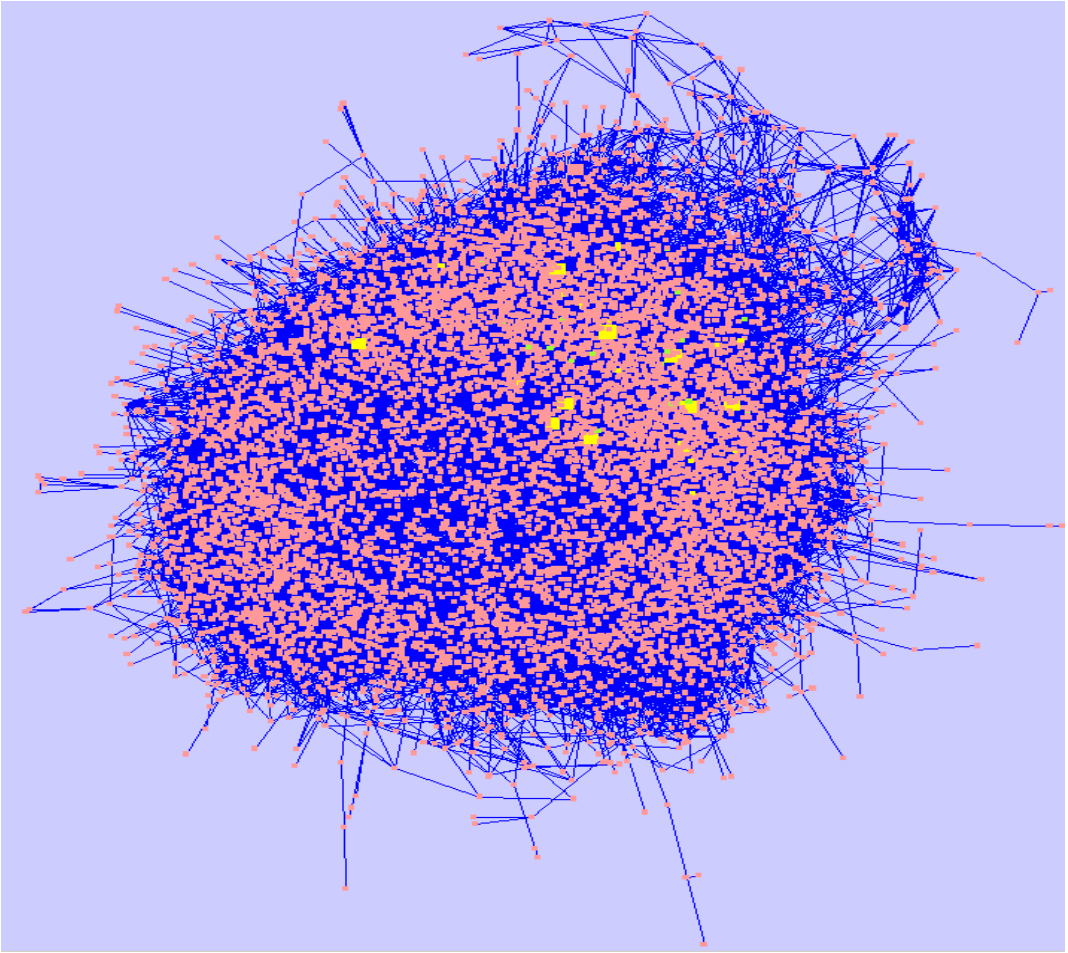

Figure 5.1: Affymetrix ARACNe network of human MB with identified CIS. The complete Affymetrix gene network generated using expression from 119 Human MB samples. It consists of ~8749 Affymetrix probes. Nodes represent probes, and edges connect nodes which show a significant mutual information association. The MB CISs within the network are shown as yellow nodes.

When the position of the CIS genes within this network was examined, seven genes, including four of those with the highest GKC *P*-values (Table 3.6), were found to be connected to each other via nearest neighbours (Figure 5.2). This clustering of 7 CIS genes was found to be statistically significant ($p<0.006$). This significance was tested by examining the topological rearrangement of the 7 genes defining the network, and performing simulations tests by calculating the mean of the shortest path distance to the nearest joining CISs genes and then permuting this 10000 times with randomly identified genes in order to generate a null distribution. The distance observed was found to be shorter for the 7 CISs than in 99.7% of simulations.

Strikingly, the CISs genes linked within the network were all defined by gene ontology analysis as transcriptional factors, neuronal development genes or both; *CREBBP*, *NFIA*, *NFIB*, *TGIF2*, *TEAD1* and *MYT1L* are neuronal transcription factors

or co-factors, while *FGF13* is a neuronal growth factor known to be required for Xenopus central nervous system development (Nishimoto and Nishida, 2007). Furthermore, the application of gene ontology analysis to the networked genes which are linked to the CISs identified significant enrichment for genes involved in neuronal architecture and functions (p=0.019–0.0042 Table 5.1, and highlighted in green in Figure 5.2). This suggests that the network as a whole consists of neuronal transcription factors mutated by Sleeping Beauty, and their targets, and that collectively these genes influence neuronal differentiation programmes.

Table 5.1: Ontology Analysis of genes linked within CIS network. Cellular Component and Molecular Function Ontologies over-represented in the Network are shown. P-value correction was achieved using a Benjamini and Hochberg False Discovery rate of 0.05. Gene frequencies in the network list and genome are shown. Ontologies marked with an asterisk are not significantly enriched when CIS genes are excluded. The neuronal and TF gene lists are highlighted.

Cellular Component					
GO-ID	***p-value***	***corr p-value***	***Network Freq***	***Total Freq***	***Ontology Description***
43226	3.82E-05	3.29E-03	65/83	9336/16374	organelle
45202	**9.87E-05**	**4.24E-03**	**9/83**	**367/16374**	**synapse**
5622	1.51E-04	5.20E-03	72/83	11306/16374	intracellular
44456	**4.47E-04**	**1.28E-02**	**7/83**	**271/16374**	**synapse part**
43228	7.43E-04	1.60E-02	24/83	2432/16374	non-membrane-bounded organelle
43232	7.43E-04	1.60E-02	24/83	2432/16374	intracellular non-membrane-bounded organelle
45211	**9.87E-04**	**1.89E-02**	**5/83**	**150/16374**	**postsynaptic membrane**
5634	1.19E-03	2.05E-02	40/83	5183/16374	nucleus
5856	1.62E-03	2.29E-02	16/83	1402/16374	cytoskeleton
43231	1.69E-03	2.29E-02	56/83	8347/16374	intracellular membrane-bounded organelle
43005	**2.83E-03**	**3.48E-02**	**7/83**	**373/16374**	**neuron projection**
43198	**3.68E-03**	**4.23E-02**	**2/83**	**18/16374**	**dendritic shaft**
Molecular Function					
GO-ID	***p-value***	***corr p-value***	***Network Freq***	***Total Freq***	***Ontology Description***
3700*	**1.71E-05**	**2.62E-03**	**16/78**	**949/15440**	**transcription factor activity***
32452	1.99E-05	2.62E-03	3/78	11/15440	histone demethylase activity
30528*	**4.18E-05**	**3.50E-03**	**20/78**	**1506/15440**	**transcription regulator activity***
8134*	**5.94E-05**	**3.50E-03**	**11/78**	**522/15440**	**transcription factor binding***
32451	6.62E-05	3.50E-03	3/78	16/15440	demethylase activity
5515	1.65E-04	7.28E-03	57/78	8119/15440	protein binding
16563*	**3.03E-04**	**1.14E-02**	**9/78**	**429/15440**	**transcription activator activity***
5522	3.73E-04	1.23E-02	2/78	6/15440	profilin binding
16706	6.07E-04	1.40E-02	3/78	33/15440	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
51213	6.17E-04	1.40E-02	4/78	77/15440	dioxygenase activity
43425*	**6.92E-04**	**1.40E-02**	**2/78**	**8/15440**	**bHLH transcription factor binding***
5488	7.85E-04	1.48E-02	73/78	12364/15440	binding
3677*	8.64E-04	1.52E-02	23/78	2327/15440	DNA binding*

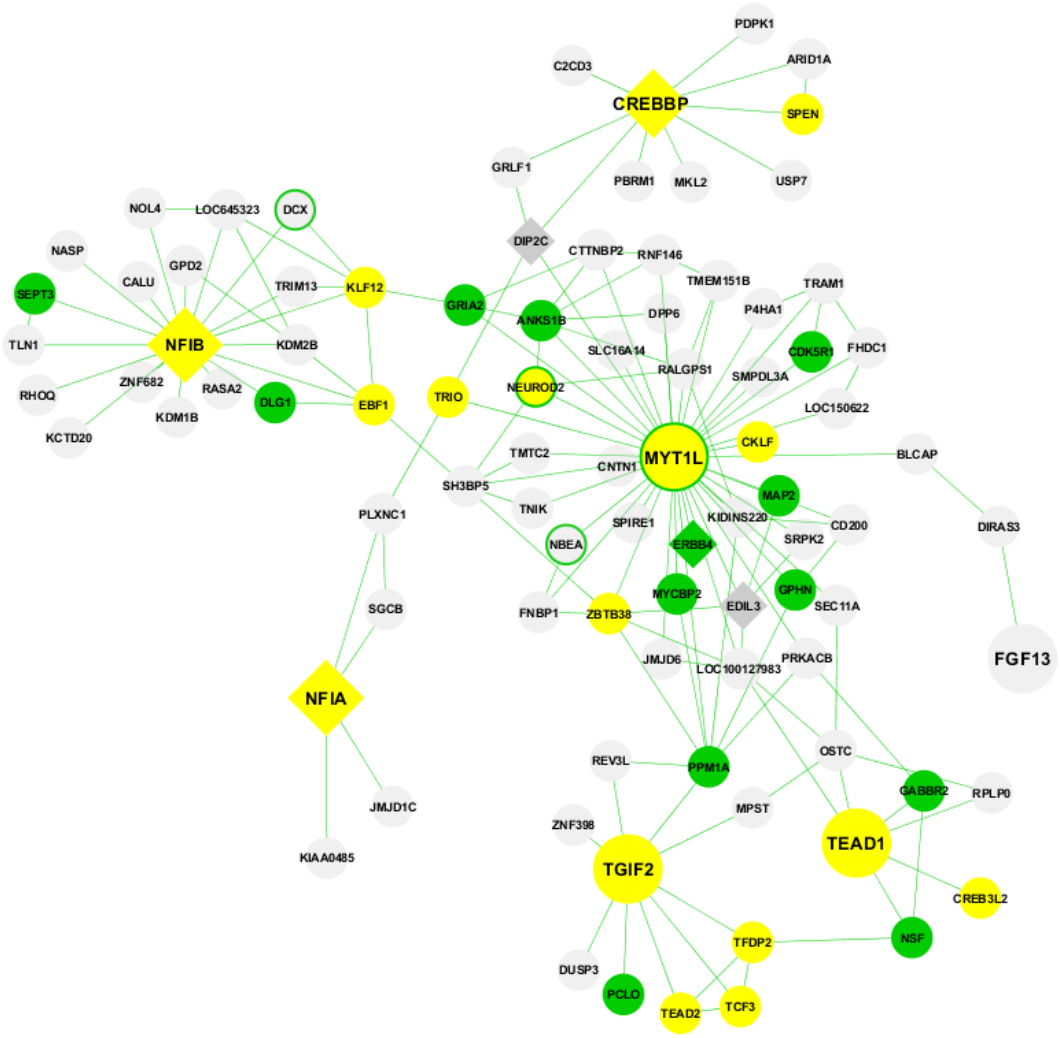

Figure 5.2: ARACNe expression subnetwork defined by MB CIS genes. Only MB CISs genes linked via nearest neighbours are shown. MB CISs are shown enlarged; genes also identified as biallelically mutated MB CISs by Wu *et al.,* (2012) are shown as diamonds. Genes with transcription factor activity are shown in yellow; Genes with enriched ontologies associated with differentiated neurons are shown in green. Genes with a known role in neuronal development are shown with a green border

5.2.2 Analysis of CIS gene expression with respect to patient survival

Only 17 CIS genes were identified using the GKC analysis (Table 3.6) out of a total of ~20,000 murine genes. This represents approximately 0.1% of all genes. To investigate whether the expression of any of these genes associate with patient survival, log rank (Mantel-Cox) tests were performed and Kaplan-Meier (KM) plots were created for all 17 CISs. Overall patient survival was analysed by Daniel Williamson (NICR) using median split data taken from (Cho *et al.*, 2011). Strikingly, the expression of two CISs genes, namely *PTEN* (a defined suppressor gene previously involved in MB development, (Hartmann *et al.*, 2006) and *MYT1L* (a pan neural transcription factor, (Vierbuchen *et al.*, 2010;Pang *et al.*, 2011) were significantly associated with poorer overall survival ((p=0.019 and p=0.016 by log rank test respectively, Figure 5.3A and B). Consistent with the poorer outcome of patients with low gene expression, the inferred mode of action (MOA) of SB insertions targeting these genes was down-regulation (Table 3.5). This is also consistent with the inference from the highly networked position of *Myt1l* that it is a key gene.

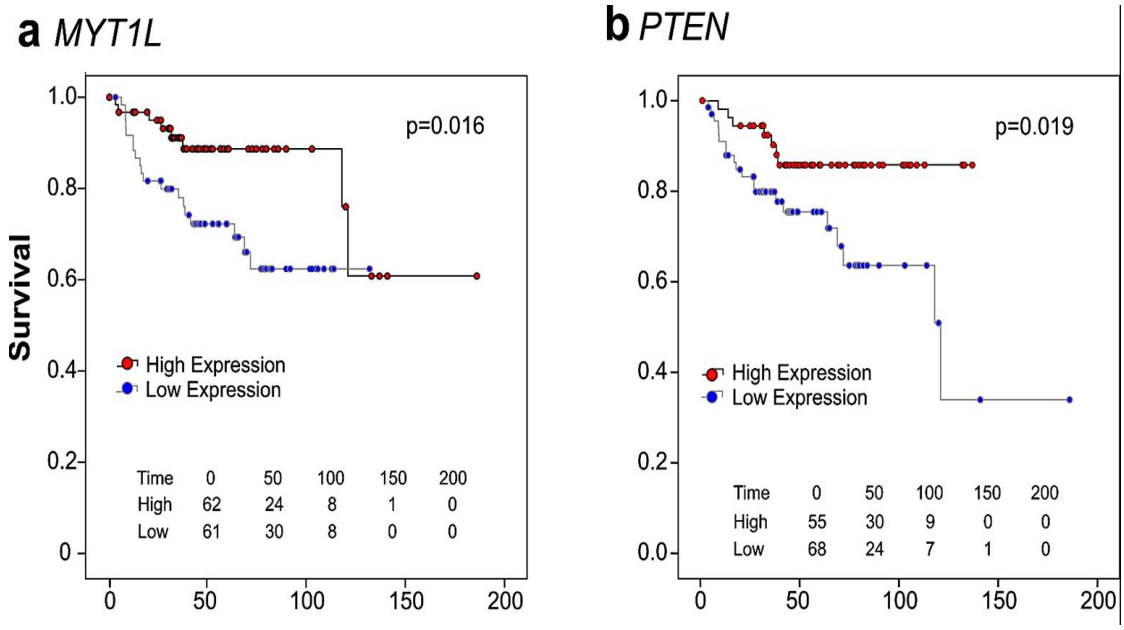

Figure 5.3: Expression correlation of *MYT1L* and *PTEN* with patients survival: Kaplan Meier curves presenting the correlation between the expression of *MYT1L* (a) and *PTEN* (b) (defined CISs) at both high and low levels, and patient's survival data (Cho, Tsherniak et al. 2011) split by median with regardless MB subtypes (follow up in months). Sample numbers and Log-ratio test p values are shown.

5.2.3 MB CISs and the network they define are differentially expressed within human MB clinicogenetic subgroups:

To investigate the relative activities of genes within the network, an expression heatmap of all probes (259) exemplifying the genes was generated using publicly available gene expression data (Kool *et al.,* 2008). To view only the most informative probe for each gene, the average expression across all samples for all available probes from all genes was calculated, and the most highly expressed probe for each gene was chosen for display. This gave a final number of 81 high quality probes, exemplifying 81 genes and gives an overview of how the network genes are behaving within human MBs (Figure 5.4).

It is clear that there are distinctions between the different MB subgroups. Interestingly, the majority of the genes in the network show a low level of expression within the WNT and SHH subgroups and higher expression in the other two subgroups. These patterns of expression may be due, at least in part, to the inferred mode of action (MOA) of the insertions as most of them were loss of function. In contrast, a small subset of genes show high expression within WNT and SHH tumours, while they showed low expression within the other sub groups, namely *PLXNC1, CALU, LOC100127983, NECAB1, TEAD1, TCF3, DC2, TRAM1*, *P4HA1, CREB3L2*, *CKLF, RPLP0, SPC18*, *MPST*, *TFDP2*, *TEAD2*, *TGIF2* and *TDLG1*. Two of these, *TGIF2* and *TEAD1*, are MB CISs identified here, and have been shown by q-PCR analysis to be up-regulated by SB insertions (Figure 4.8 and 4.10 respectively). It is noteworthy that in the original subgroup specific analysis of this expression data (Kool *et al.*, 2008), four of these genes were shown to be significantly up-regulated in the WNT subgroup (*TCF3, TRAM1, P4HA1* and *CKLF*), and others were identified as being expressed at a significantly higher level within the SHH tumour subgroup (*DC2, CREB3L2, RPLPO* and *MPST*). Only *TFDP2* was identified within Group 3 and at a significantly lower expression, with *PLXNCI* and *CALU* being found to be expressed at a low level within Group 4 (Kool, *et al.*, 2008). This diversity of expression of network genes within human MB subgroups suggests that some of these genes may contribute to the development of more than one MB subgroup of tumour.

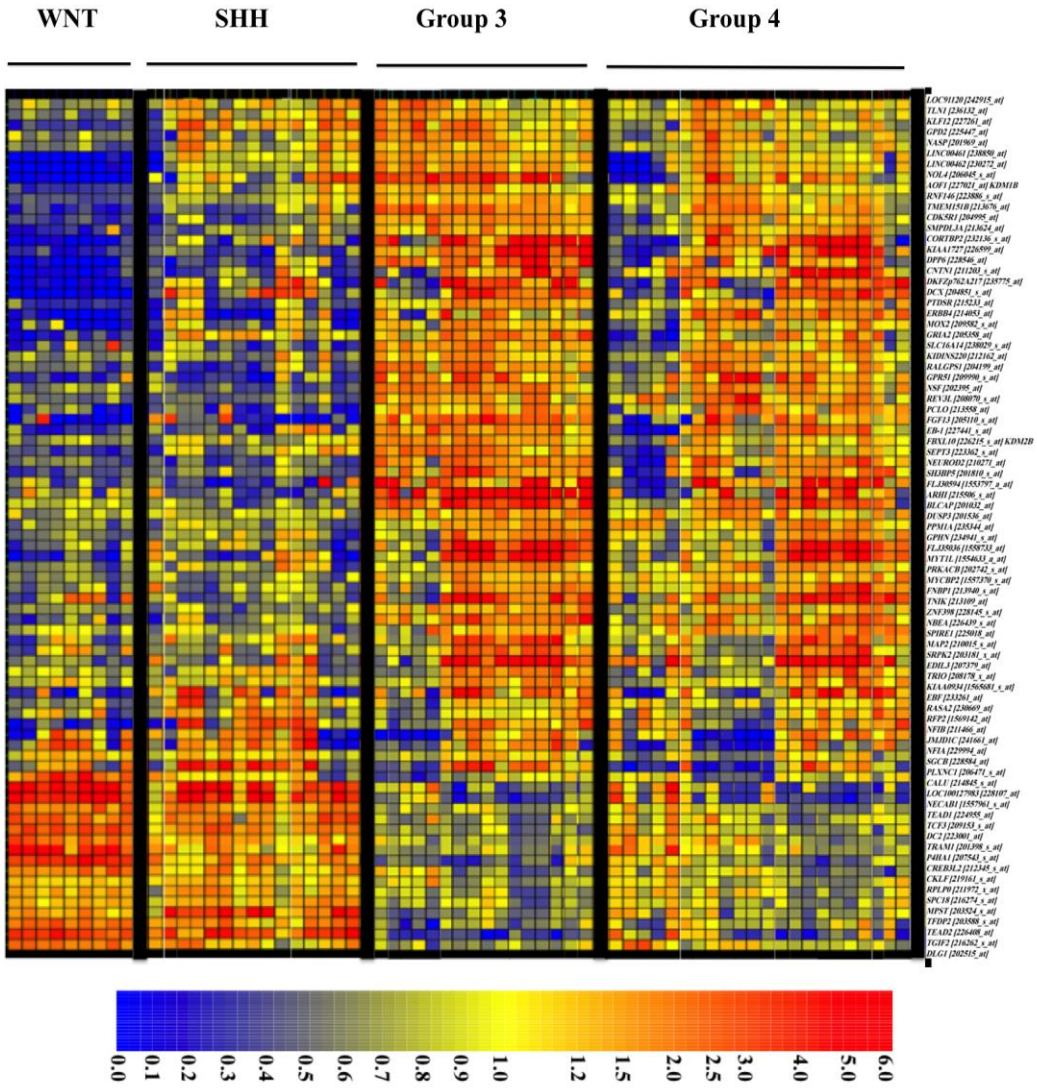

Figure 5.4: Differential expression of all CISs networked genes. An expression heatmap of 81 genes representing the CISs network within human MB subgroups. The probe with the highest mean expression level across all sample is shown for each gene. The majority of genes are expressed at relatively low levels within both WNT and SHH sugroubs, compared to a small subset being expressed at low levels within Group 3 and Group 4. The colour bar shows relative expression levels.

5.2.4 Network genes investigated separately according to GO analysis

Although the expression heatmap shows clear differences between subgroups, it does not allow any insight into the function of the genes due to the large number in the network. To specifically investigate the activity of the different enriched ontological genesets within the network, the genes were separated for the purpose of presentation into CIS genes, transcription factors other than CISs within the network, and neuronal development genes. Expression heatmaps were established separately for each of these.

Fig 5.5A shows the heatmap of the seven CISs which define the network. *MYT1L* and *FGF13* are highly expressed in Group 3 and Group 4, whereas in the WNT and SHH subgroups, *TGIF2* was highly expressed. In contrast, *NFIB* shows very low expression in the WNT subgroup, whereas *TEAD1* and *NFI*A show variable expression among all subgroups.

The expression heatmap of other transcription-factors within the network is depicted in Figure 5.5B. These genes show two basic patterns of expression within human MB subgroups. *TEAD2, TCF3, CKLF, CREB3L2* and *TFDP2* show high expression behaviour in the WNT/SHH subgroups, while *EBF1, CDK5R1, NEUROD2* and *ZBTB38* show high expression in Group 4.

Fig 5.5C presents the expression heatmap of neuronal genes in the network. The majority of these genes are expressed at low levels within WNT/SHH subgroups, including *SEPT3, CDK5R1, EB-1, GPHN, NSF, GPR51, MYCBP2, PCLO, ERBB4* and *MAP2*. Only *DLG1* is expressed at a high level (this gene is known to have an important role in cell proliferation and synaptogenesis), suggesting that the CIS network has a significant impact on the neuronal differentiation process. It is clear that most of the neuronal genes, together with most of the other genes in the network, show consistently lower expression within human SHH and WNT MB subgroups tumours, and this is consistent with the MOA of the insertion affecting the CISs (that defined the network) being loss of function. It is also consistent with previous observations that Group 4 tumours show high levels of neuronal gene expression (Kool *et al.*, 2008). Also, from the function of the genes it could be inferred that mutation of the network TF CISs work synergistically with the disrupted SHH signalling pathway in the mouse model to maintain undifferentiated, highly proliferative cells, one of the known hallmarks of tumourgenesis.

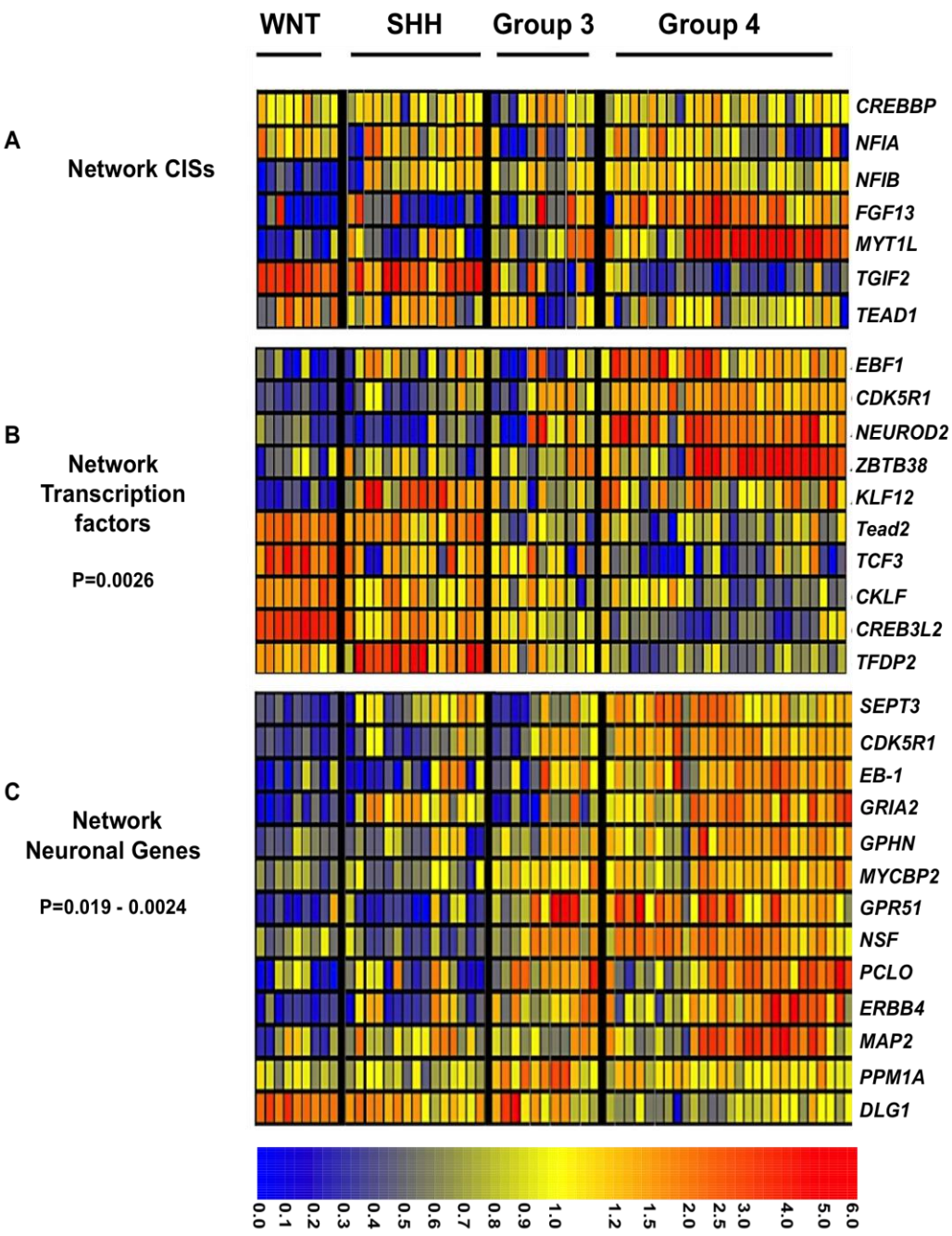

Figure 5.5: Network activity correlates with human MB clinicogenetic subgroups: gene expression heatmaps of the network genes based on enriched GO specifications in all MB subgroups are shown. Where more than one probe per gene was present, the probe with the highest mean expression level across all samples is shown. **A.** Expression of CIS genes, which define the network. **B.** Expression of other 10 transcription factors in the network. **C.** Expression of 13 neuronal genes in the network. The colour bar indicates relative expression**.**

5.2.5 Metagene activity in humans correlates with advanced disease in SHH subgroup tumours

The expression heat maps of individual genes in the network, above, showed significant differential expression between tumour subgroups, and consistently high expression of neuronal genes in group 4 tumours. This raises the possibility that network activity as a whole could both discriminate between subgroups, and be of biological/clinical relevance. To investigate the possibility, the overall activity of network genes was assessed within human datasets by deriving a single network "metagene" score. To do this, the expression of each individual gene was first marked, (signed) depending on the direction of correlation with other CISs. The direction of this sign was arbitrarily chosen, and here low expression of genes with an inferred mode of action (MOA) of loss of function (e.g. *MYT1L*) correlates with a low metagene score, while low expression of CISs with an inferred MOA of upregulation (e.g. *TGIF2*) correlates with a high metagene score. A single metagene metric for the network was established, and the clinical features of all patients within the Kool *et al*., (2008) and Fattet *et al.*, (2009) studies were analysed.

The results of this analysis, which was performed by Dr. D. Williamson (NICR), are presented in Figure 5.6. As expected from the heat maps previously shown in Figures 5.3 and 5.4, the activity of the metagene is significantly different between MB subgroups ($F=26.8$; $p < 0.0001$, as shown in Figure 5.6A). The highest network activity was recorded within Group 4 tumours. This activity is presented at moderately higher levels within subgroup SHH and Group 3. However, the metagene activity is shown to be expressed at the lowest levels within the WNT subgroup.

Moreover, in the same dataset, the metagene was observed to be differentially active in tumours with metastasis compared to those without ($t = 2.107$, $p<0.038$) (Figure 5.6B). This is consistent with the human MB Group 3 and Group 4 being associated with a high frequency of metastatic disease. Interestingly, the investigation of publicly available survival data for human MB patients (Cho *et al.*, 2011) revealed a significant correlation between network activity and survival outcome, but only within the SHH subgroup tumours (Log-rank 8.03, $p<0.005$, as shown in Figure 5.6C).

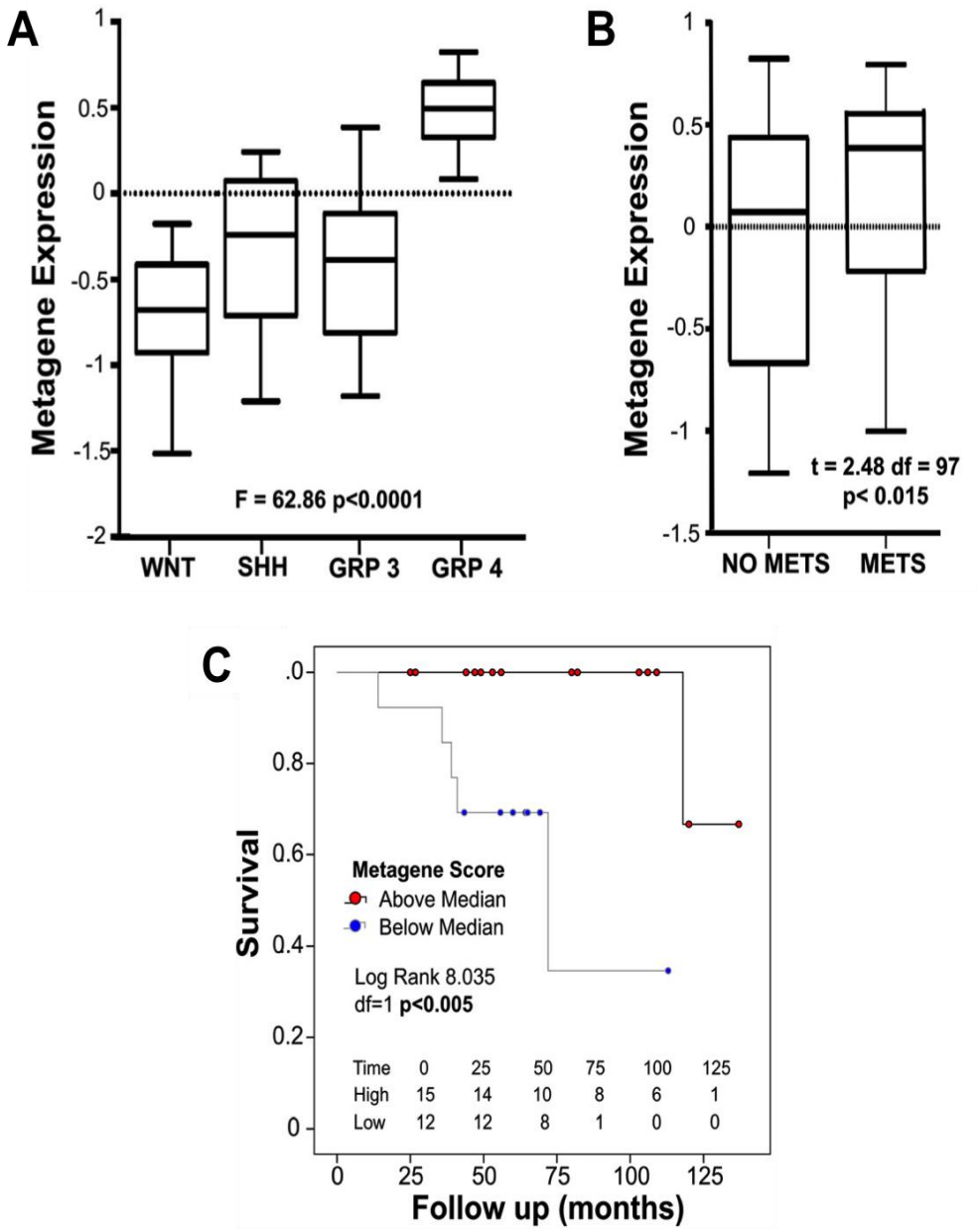

Figure 5.6: Network activity: relationship with Metastasis and survival. A. Box and whisker plot showing log of metagene expression in MB clinicogenetic subgroups. ANOVA p-value is shown. Boxes represent the 25th to 75th percentile, with median values shown as solid lines. Whiskers represent 95th percentile, and outliers are shown individually. **B.** Log of metagene expression in tumours with metastases and those without. T-test p-value is shown. **C.** Kaplan Meier curve of SHH subgroup tumours relative to Metagene score. Data generated by D. Williamson (NICR).

5.2.6 SB mutagenised murine model is representative of human MB SHH tumours

The results above support the hypothesis that the neuronal transcription factor network identified within human MB expression data using genes mutated in the *Ptch1*+/- murine model is central to the development of MB in both species. To corroborate the human data, and to attempt to identify specific outputs of the network, I generated expression data from SB mutagenised MBs using the Illumina bead array platform. This technology was selected based on several advantages: increasing popularity, low cost, and the highly replicated probe features that increase the accuracy of beads type measurements. This technology is principally based on extremely small sized Bead-Arrays, with a 3-micron silica bead size, with the beads being covered with a high copy number of oligonucleotides (Cordero *et al.*, 2007).

I isolated total RNA from thirty tumours from the SB mutagenized experimental group (24 tumours with network mutations), together with six tumours from the *Ptch1*+/- predisposition control group, and six cerebella from 6-8 week old wild type mice. The quality of all RNA samples used in the microarray analysis was then checked using an Agilent Bioanalyser. All the samples had an RNA integrity number (RIN) 8.0 or above, with the exception of two samples, MB524 (7.7) and MB244 (6.7) (Table 5.2). An example of the quality of two SB induced murine MB (experimental group) samples are shown in Figure 5.7A. The investigated RNA shows two distinct peaks corresponding to the 18S and 28S ribosomal RNAs. About 200ng of each sample of high quality RNA was then taken through to amplification and cRNA was generated. Each cRNA size distribution was tested using an Agilent Bioanalyser and an example of two SB induced murine MB samples of cRNA are shown in Figure 5.7B. These cRNAs were labelled and used to probe the Illumina Mouse 8 Reference Array (see Methods).

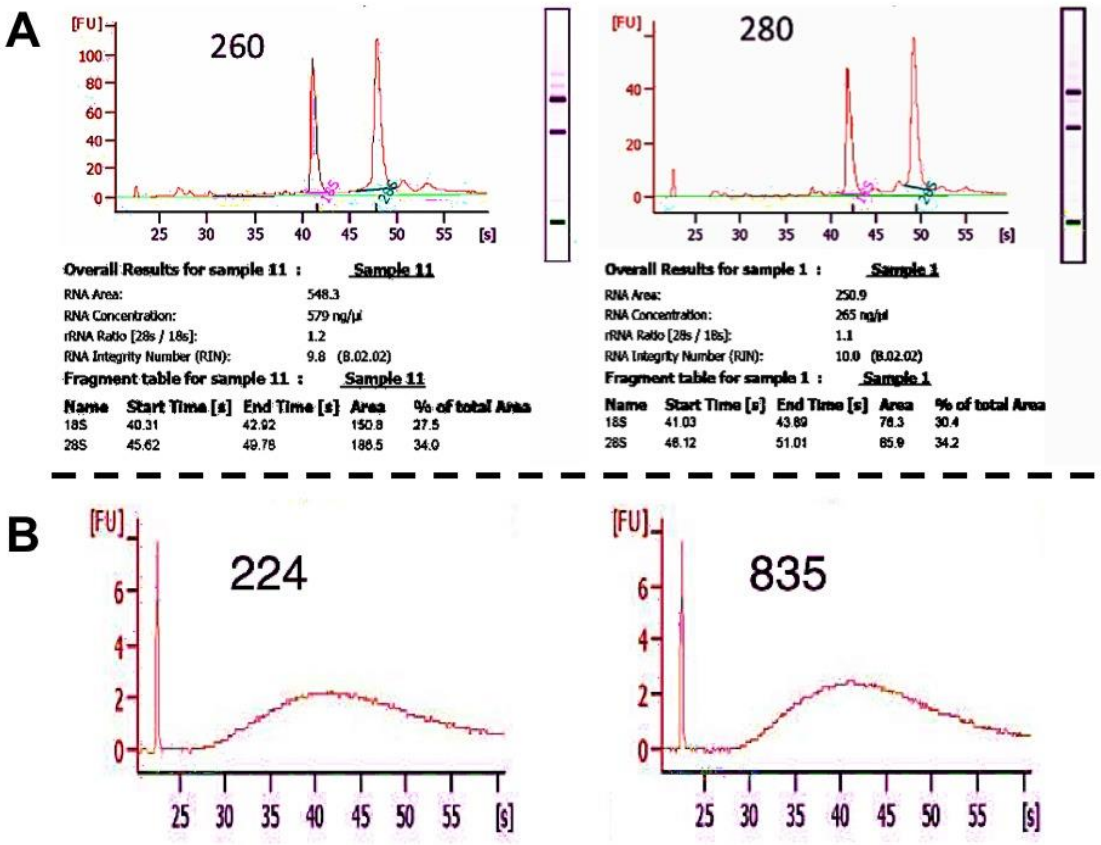

Figure 5.7: RNA and labelled cRNA quality and distribution. A. Agilent Bioanalyser electropherogram results show the high quality and size distribution of 2 experimental group samples (SB induced murine MB) used for Illumina bead microarray. Gel electrophoresis distributed bands representing rRNA isoforms are included. Detailed samples information about RNA concentration, ratio and RIN number are shown. **B.** Agilent Bioanalyser electropherogram results show cRNA size distribution of 2 experimental group samples used in Illumina bead microarray are shown. the X axis shows the time in seconds (S) and the Y axis shows florescence (FU).

Table 5.2: SB induced murine samples, *Ptch1+/-* control tumours and normal cerebella used in Illumina beads microarray analysis. RIN - RNA integrity number.

Sample ID	RIN
MB/SB11	
MB-P-008-T-Brain_PTS11-008	**9.3**
MB-P-048-T-Brain_PTS11-048	**9.8**
MB-P-068-T-Brain_PTS11-068	**8.8**
MB-P-116-T-Brain_PTS11-116	**8.9**
MB-P-146-T-Brain_PTS11-146	**9.8**
MB-P-224-T-Brain_PTS11-224	**6.7**
MB-P-260-T-Brain_PTS11-260	**9.8**
MB-P-280-T-Brain_PTS11-280	**10**
MB-P-286-T-Brain_PTS11-286	**10**
MB-P-349-T-Brain_PTS11-349	**8.3**
MB-P-366-T-Brain_PTS11-366	**9.6**
MB-P-386-T-Brain_PTS11-386	**8.7**
MB-P-421-T-Brain_PTS11-421	**9.7**
MB-P-457-T-Brain_PTS11-457	**8.1**
MB-P-524-T-Brain_PTS11-524	**7.7**
MB-P-540-T-Brain_PTS11-540	**8.2**
MB-P-565-T-Brain_PTS11-565	**9.8**
MB-P-580-T-Brain_PTS11-580	**8.7**
MB-P-619-T-Brain_PTS11-619	**10**
MB-P-730-T-Brain_PTS11-730	**10**
MB-P-751-T-Brain_PTS11-751	**9.2**
MB-P-754-T-Brain_PTS11-754	**9.2**
MB-P-756-T-Brain_PTS11-756	**9.5**
MB-P-786-T-Brain_PTS11-786	**9.4**
MB-P-791-T-Brain_PTS11-791	**8.6**
MB-P-793-T-Brain_PTS11-793	**9.4**
MB-P-833-T-Brain_PTS11-833	**8**
MB-P-835-T-Brain_PTS11-835	**10**
MB-P-849-T-Brain_PTS11-849	**10**
MB-P-731-T-Brain_PTS11-731	**10**
MB/PTCH	
MB-P-19-T-Brain_Ptch-19	**9.5**
MB-P-104-T-Brain_Ptch-104	**8.9**
MB-P-105-T-Brain_Ptch-105	**9.3**
MB-P-105-T-Brain_Ptch-159	**9.3**
MB-P-300-T-Brain_Ptch-300	**9.1**
MB-P-483-T-Brain_Ptch-483	**8.5**
Control /cerebellum	
PTS11-Cerebellum-908	**9.3**
PTS11-Cerebellum-910	**9.5**
PTS11-Cerebellum-1014	**8.9**
PTS11-Cerebellum-1018	**9.4**
PTS11-Cerebellum-1020	**9.1**
PTS11-Cerebellum-1012	**8.8**

After hybridisation and data generation, the gene expression data from the mutagenized mouse model tumours was first analysed by D. Williamson (NICR, Newcastle University) to assess if it was consistent with human MB expression. Briefly, Non-Negative Matrix Factorisation (NMF) (Tamayo *et al.*, 2007) was employed to define metagenes from all four clinicogenetic subgroups using the human expression data (Kool *et al.*, 2008; Fattet *et al.*, 2009). Then, these metagenes were projected across the mouse tumour expression data using all available orthologous probes. Both the predisposed only *Ptch1*+/- control tumours and the mutagenized experimental murine model tumours are seen to cluster with human SHH subgroup tumours (Figure 5.8A).

To establish more precise categorisation of the mutagenized mouse model expression data, the murine tumour subgroup identities were tested via a Support Vector Machine (SVM). This was performed using human data metagene scores as a guidance (training) set and then using the murine tumour data as the test set. All human tumours were trained correctly with zero errors, and ~96% (29/30) of the transgenic murine tumours were significantly categorised to be SHH subgroup tumours, together with 6/6 (100%) of the predisposed to Ptch+/- control murine tumours. This indicates that the gene expression within SB mutagenized murine tumours is similar to human SHH subgroup tumour expression (Figure 5.8B), confirming the appropriateness of the *Ptch1*+/- SB model.

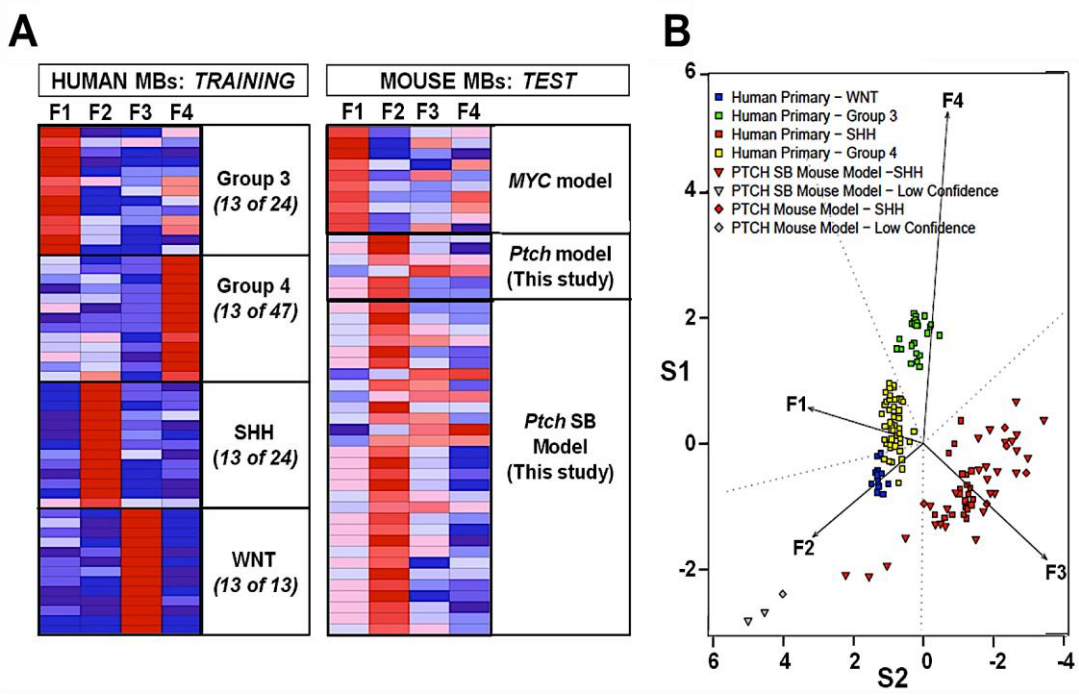

Figure 5.8: SB induced murine and Ptch$^{+/-}$ MBs are SHH subgroup tumours. Projection of four sub-group specific metagenes derived from 108 human primary MB expression profiles (Training Set) onto MB mouse model expression profiles (Test Set) using Non-negative Matrix Factorisation (NMF). **A.** Heatmaps depicting expression of metagenes within human cancers (training; left panel) and expression of metagenes throughout murine tumours (test; right panel). **B.** Pseudo-plot of NMF projections coloured according to confident MB sub-group classifications by Support Vector Machine trained on human primary medulloblastoma using metagene expression values. This unsupervised analysis shows that the majority of *Ptch* mouse model samples recapitulate the expression profiles of primary SHH MB. Figures generated by D. Williamson (NICR).

5.2.7 Microarray expression analysis of mutagenized tumours identifies *Igf2* as a key network-associated gene

Having established that the tumours generated by SB mutagenesis of *Ptch1*+/- mice are representative of the human MB SHH subgroup tumours, the next question was what impact disruption of the expression network by SB insertion in one or more network CIS gene had upon gene expression. Hence, the mouse expression data were split into network-hit and non-network-hit samples. Very few significantly differentially expressed genes were defined by this approach, with the top genes shown in Table 5.3. *Igf2* was the highest differentially expressed gene with a mean fold change of 3.58 (p=0.002). This suggests that the network activity may have an effect on some aspects of the insulin-signalling pathway.

To validate the relationship between the activity of the CISs within the network and *Igf2* expression, a quantitative RNA expression analysis using TaqMan real-time

PCR was carried out to confirm and quantify the expression of *Igf2* in all available MB mouse tumours used for the insertion site analysis, compared with the normal cerebellum.

Table 5.3: Gene expression difference between tumours with and without network hits.

Illumina Probe_ID	Gene Name	Mean Fold Change in Expression: Network vs. non-Network	*P*-value
LMN_2597769	*Igf2*	3.58	0.002
ILMN_1255871	*Loxl1*	1.64	0.029
ILMN_2757232	*Aqp4*	1.64	0.037
ILMN_1242456	*Kank1*	0.66	0.002
ILMN_2642012	*LOC100046457*	0.65	0.007
ILMN_2731265	*Hsd11b2*	0.65	0.026
ILMN_1252076	*Lyz2*	0.64	0.040
ILMN_1230157	*Rnd3*	0.61	0.038

The expression of *Igf2* was analysed in tumours with hits in the network and those without. Confirming the Illumina bead array result, a significantly higher expression of *Igf2* was detected in tumours with inserts in the network than in tumours without a network hit (p=5.2 E-06, Figure 5.9A). Moreover, *Igf2* expression was significantly higher in tumours with hits in *Nfia*, the most frequently inserted CIS gene, than tumours with no inserts in the *Nfia* gene (p=0.018, Figure 5.9B). It is also noteworthy that *Igf2* was expressed at a very high level within some network tumours, when compared to expression in the normal cerebella, with expression in some non-network tumours being very low (Figure 5.9C).

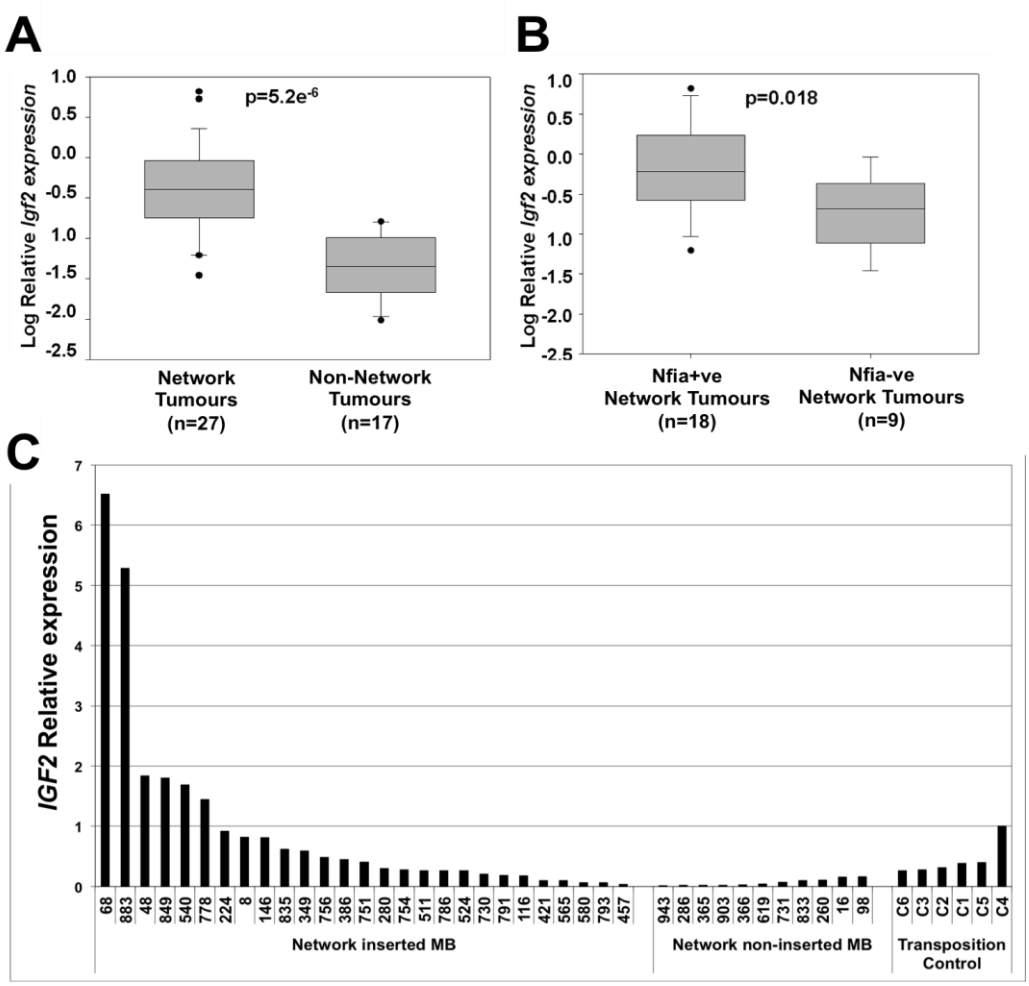

Figure 5.9: Relative *Igf2* expression in network and *Nfia* inserted tumours. **A.** Box and whisker plots showing *Igf2* expression in tumours with inserts in 1 or more network CIS versus tumours with no inserts. **B.** *Igf2* expression in tumours with 1 or more insertions in *Nfia*, versus tumours with no *Nfia* insert but with inserts in at least 1 other network CIS gene. Boxes represent the 25th to 75th percentile, with median values shown as solid lines. Whiskers represent 95th percentile, and outliers are shown individually. **C. Relative *Igf2* expression in murine tumours and cerebellum**. Relative gene expression is shown, calculated using the $2^{-\Delta\Delta Ct}$ method, using the highest cerebellar RNA expression value as reference (C4).

5.3 Discussion

Gene expression data sets obtained both from publicly available sources, and generated *de novo* from murine tumours generated during the SB screen, have been used here to investigate the possibility of interactions between CIS genes, and to check the appropriateness of the *Ptch1*+/- model for comparative studies. Reassuringly, the results confirm that the tumours induced by whole-body SB mutagenesis most closely resemble human SHH subgroup tumours in terms of gene expression patterns. This is clearly seen through the NMF and SVM results (Figure 5.8 A and B). All SB induced murine tumours and *Ptch1*+/- tumours clustered with human SHH subgroup tumours.

The network and gene ontology analysis used here provides a more in-depth look at how the CISs identified in the SB screen may interact with each other, and implies a role for the network in both neuronal development and cancer formation. The ARACNe analysis suggests that at least 7 CIS genes identified in this study are linked to each other within a single transcription factor network which does not map directly to a known signalling pathway. Interestingly, four out of the seven MB CIS genes that defined the network, namely *CREBBP*, *NF1A*, *NFIB* and *TEAD1*, as well as three other genes within the network (*ERBB4, DIP2C* and *EDIL3*), have been identified as MB CISs in a recent tissue specific SB mutagenesis screen of primary tumours within the *Ptch1* model (Wu *et al.*, 2012) shown as diamonds in Figure 5.2). It is also noteworthy that *MYT1L* is the most highly-networked gene among the CISs, indicative of an important role as a key co-ordinator. Furthermore, expression of *MYT1L* impacts upon survival.

The gene ontology results suggest that the normal biological impact of these CISs in the cerebellum is integrated toward a single output, neuronal differentiation/development, and that perturbation of this network (through SB mutation in mouse or expression changes in human) enhances tumour formation.

The significant enrichment for genes critical for neuronal architecture within the network is of particular interest, and many individual genes have important known neuronal functions. For example, one of these genes, *Pclo* (Presynaptic cytomatrix protein), is involved in synaptic formation where it can negatively regulate synaptic vesicle (SV) exocytosis (Leal-Ortiz *et al.*, 2008), and the network analysis suggest that this gene may be a target gene for *Tgif2*. *PPM1A* is known to be essential for neuroblast migration to the olfactory bulb (Khodosevich *et al.*, 2009). This gene is a member of the PP2C family, which is involved in negative regulation of the cell stress response pathway (Wu *et al.*, 2007). Additionally, this gene was shown to drive different types of

cells, such as neuronal cells, toward apoptosis (Klumpp *et al.*, 2006) suggesting its role in cancer formation needs to be assessed further. From the network analysis, this gene may be a target of *Tgif*2 and *Myt1l*. Dendrite microtubule-associated protein *MAP2* is a large neuronal protein expressed predominantly within dendrites of mammalian brain tissues. It is considered a neuronal marker as it is involved in the microtubule-associated complex that is necessary for neurogenesis (Johnson and Jope, 1992;Sanchez *et al.*, 2000;Suzuki *et al.*, 2011). Another important networked gene involved in regulating neuronal activities is Gamma-aminobutyric acid B receptor (*GABBR2*). This gene is highly expressed in the dendrites of Purkinje cells (Chung *et al.*, 2008a). It has an inhibitive neurotransmitter effect in the central nervous system. It is believed that the dysfunction of this gene governs a variety of nervous system disorders. For example, hypoactivity of this gene is associated with schizophrenia and bipolar disorder, based on analysis of post-mortem samples (Fatemi *et al.*, 2011). Furthermore, *GPHN* is a cornerstone of the inhibitory process of neurotransmitter receptors to the microtubule matrix of the postsynaptic cytoskeleton. This process is performed via highly active binding to a tubulin dimer and inhibitory glycine receptor (Gly R) domain. Missense and deletion mutations of this gene may cause hyperekplexia and MoCo deficiency-like symptoms, respectively (Reiss *et al.*, 2001;Rees *et al.*, 2003). Both disorders display severe and fatal neurological damage.

All of these genes are involved in processes, or structures, associated with differenting/differentiated neurons, and are expressed at a lower level in SHH subgroup tumours than group 3 and 4 tumours. This is consistent with a block/downregulation of expression by the SB mutations within the linked CIS genes, leading to maintenance of an immature cellular phenotype, diverting the normal direction of neuronal cells from the differentiation process toward proliferation. Also, from the function of the genes it could be inferred that the mutation of the network TF CISs work synergistically with the disrupted SHH signalling pathway in the mouse model to maintain undifferentiated, highly proliferative cells, one of the known hallmarks of tumourgenesis.

One of the key observations from this analysis is that the mutated CISs identified here appear to synergise to increase the expression of *Igf2* – a key gene already known to be critical for medulloblastoma development in the *Ptch1*+/- mouse model. The higher expression of *Igf2* within SB induced tumours with insertion in the network reported here provides further evidence for the importance of *Igf2* in terms of MB generation, and potentially identifies key genes involved in regulating the expression of this gene in developing cerebellar GNPCs.

The first clear indication for *Igf2* involvement in MB was a study conducted by Hahn *et al.*, 2000, which aimed to assess the impact of *Igf2* upon MB and rhabdomyosarcoma (RMS) tumourgenesis. The *Igf2* gene is imprinted with only the paternal allele being expressed and they found that *Igf2* was upregulated in *Ptch1*$^{+/-}$ RMSs due to high expression of the imprinted allele driven by one alternative promoter (P3). To assess the impact of *Igf2* upon MB and RMS formation, the murine *Ptch1*$^{+/-}$ allele used here was transferred into an *Igf2* null background. This resulted in the frequency of tumour formation dropping to almost zero, indicating that *Igf2* expression is essential for tumour formation. Furthermore, loss of the *Ptch1* wild type allele from embryonic tissue was accompanied by increased *Igf2* expression, suggesting that the gene may also be a target of the Shh signalling pathway.

Subsequently, Rao *et al.*, 2004, used cell-type specific postnatal gene transfer to express *Igf2* and *Shh* alone and in combination in cerebellar granule neuron precursor cells (GNPCs), to investigate the roles of *Igf2* and *Shh* in MB formation. Transfection of *Igf2* alone resulted in no tumour generation. However, co-transfection of *Igf2* with *Shh* increased the frequency of tumours relative to *Shh* alone from 15% to 39%. Immunohistochemical analysis of *Igf2/Shh* tumours established that they were classical MBs expressing neural progenitor/early differentiation markers (*β-III Tubulin* and *Neu N*), but not markers of terminal differentiation such as glial fibrillar acidic protein (*GFAP*) and neurofilament protein. Immunoperoxide staining using antibodies against components of the *Igf2* signalling pathway (*Igf1R, IRS-1, AKt*) establishing that *Igf2* signalling was activated in MBs generated by both *Shh* only and through the combination of *Igf2* and *Shh*.

Recently, Igf signalling pathways were assessed in terms of their synergistic action with Shh induced proliferative cells. Immunocytochemistry was used to establish the precise location of *Igf* signalling elements in the developing mouse cerebellum (Fernandez *et al.*, 2010). *Igfr1* (the main Igf receptor) was observed to be expressed ubiquitously at all stages and very strongly in the external granule layer (EGL). *Igf1* was expressed in Purkinje neurons from E17.5, but not uniformly, and peaked at postnatal day 5 (P5). In contrast, the expression of *Igf2* was specific to the meninges overlaying the EGL and choroid plexus. Igf binding proteins 1-6 (*Igfbps 1-6*) showed varied expression patterns, some mimicking *Igf2*, with *Igfbp5* being strongly expressed in both the EGL and GNPCs. The potency and synergy of *Shh* and *Igf* mitogens upon GNPCs proliferation was then analysed using cerebellar explants, purified GNPCs culture aggregates, and isolated GNPCs cultures, assessing proliferation using BrdU

incorporation. As expected, all showed mitogenic effects, with SHH being most strong. Synergy was clear as *Shh* plus either *Igf* increased BrdU incorporation ~3X relative to *Shh* only, and ~7X relative to either *Igf*. Inhibition of smoothened (*SMO*) using cyclopamine drastically reduced the impact of *Shh*, but had no effect upon *Igf* induced proliferation. Antagonism of *Igfr1* with a blocking antibody effectively blocked the response in *Igf* treated cells, while *Shh* induced proliferation was inhibited by >50% by blocking *Igfr1* as well. Furthermore, while *Igfbp2, 3 and 5* were all able to inhibit the impact of *Igf* upon proliferation, *Igfbp5* was also able to inhibit the impact of Shh. Collectively, these findings illustrate the significant synergistic impact of *Igfs* upon *Shh* induced proliferation, and also indicate that *Igfbps* can modulate both signalling networks in a complex manner (Fernandez *et al.*, 2010).

It is also noteworthy that several of the CIS genes identified by the SB screen have already been implicated in *Igf2* regulation. For instance, in prostate cancer *Igf2* expression is elevated due to the *EGR1* gene, a gene that is co-activated by *CREBBP*, one of our network CIS genes (Svaren *et al.*, 2000;Yu *et al.*, 2004;Lai and Wade, 2011). Furthermore, the expression of *IGF2* is downregulated by the increased expression of *PTEN* in hepatoma tumours (Kang-Park *et al.*, 2003): We would therefore predict *Igf2* expression to be induced by *PTEN* loss of function, the inferred impact of SB insertion observed here. Collectively, the high *Igf2* expression in murine tumours with network hits, the evidence that specific network CIS play an role in *Igf2* expression, and the proven synergy between *Shh* and *Igf2* in tumour development, provide strong evidence that the network identified here underpins *Igf2* upregulation, critical for MB progression.

An obvious question which remains unanswered by this analysis is why, if the *Ptch1*+/- model accurately mimics gene expression in human SHH subgroup tumours, has *Igf2* not been identified as a key SHH upregulated gene in any human microarray study to date. The Affymetrix probes for *Igf2* on human microarray chips are known to have low specificity due, at least in part, to the existence of readthrough transcripts from an upstream gene (*INS*). For example probesets 202409_at, 202410_x_at and 210881_s, which are on the U133 Plus 2 chip, hybridise to a minimum of 2 genes, raising the possibility that *Igf2* expression is not accurately assayed. This is supported by recent analyses by collaborators in Newcastle using RNAseq which have established that increased *Igf2* expression is one of the strongest SHH specific expression signals (S. Clifford and D. Williamson, personal communication). More importantly, this accounts for an apparent anomaly between the human and mouse datasets analysed here, and

further supports the conclusion that the transcription factor network defined here undperpins *Igf2* upregulation and MB progression in both mouse and human. This suggests that further analysis of the genes identified, and the processes they affect, may identify valid targets for novel therapeutic strategies. It is the prioritisation of these genes for further study, and their preliminary analysis, which is the subject of the final results chapter.

Chapter 6 (Result 4)

Identification and preliminary functional analysis of key candidate genes

6.1. Introduction

In the previous chapter, I presented evidence that a subset of TF CISs identified in the current study, namely *Nfia, Nfib, Tgif2, Tgf13, Myt1l, Tead1*, and *Crebbp*, are linked within a transcription network that is enriched for both TFs and genes involved in neuronal development. NMF analyses using metagene scores from all human MB subgroups, and expression data from *Ptch1*+/- only control and SB mutagenised murine tumours, revealed that genes in murine SHH tumours were expressed in a similar pattern to genes in human SHH subgroup tumours. Moreover, Support Vector Machine (SVM) analysis accurately grouped both SB murine tumours and *Ptch1*+/- control murine tumours with human SHH subgroup tumours, providing strong evidence that SB mutagenised murine tumours faithfully model human SHH subgroup tumours. Detailed analysis of gene and metagene expression further suggested an impact of the expression of network genes upon survival within SHH subgroup human tumours, and identified Igf2 as a key output of network activity.

All of these results are consistent with the CIS genes identified here being involved in MB development. A logical and essential extension of this work is the functional analysis of specific candidate CISs thought to have an impact on MB generation. For simplicity, initial experiments would ideally be performed in human or mouse SHH subgroup MB cell lines, and would use transient transfection of siRNAs to modulate gene expression of specific genes. However, no MB cell line exists which has been shown to model SHH subgroup tumours, making the use of cell lines (or the interpretation of the results that would be obtained) problematic. GNPCs are the cell of choice because of the difficulty of establishing and maintaining Shh cell lines. To date, no confirmed SHH subgroup MB cell line has been established, as during culture Shh signalling is switched off for reasons that remain unclear, unlike cells isolated from WNT MB subgroup which conserve WNT signalling activity (Sasai *et al.,* 2006). GNPCs are believed to be the cell type in which SHH subgroup MBs

originate, and primary GNPC cell culture is widely used to investigate Shh signalling during post natal murine development (Rao *et al.,* 2004). I therefore decided to use primary GNPC culture for downstream analyses of individual genes. As the efficiency of transient transfection is low in many primary cultures, lentiviral vectors were chosen to facilitate high efficiency stable transduction of constructs and/or shRNAs for functional analyses.

The CIS and network genes identified here include a large number with functions of potential relevance to MB, or linked to SHH signalling, making it unclear which CIS genes should be singled out for analysis: For instance, *Pten* and *Crebbp* are already candidates for involvement in disease as human orthologs are somatically mutated at low frequency in human MBs (Parsons *et al.,* 2011;Robinson *et al.,* 2012). *TEAD1* has been shown to be up-regulated and to co-activate *YAP1* in human SHH subgroup tumours (Fernandez *et al.,* 2009), and the *Nfi* family that includes *Nfia* and *Nfib* are believed to have a major role in axonal extension during cerebellar development (Wang *et al.,* 2004;Wang *et al.,* 2007). In addition, *Myt1l* is a neuronal gene which has been utilized in combination with other transcription factors to reprogramme mouse fibroblasts to a mixed neuronal lineage (glutamatergic and GABAergic neurons, (Vierbuchen *et al.,* 2010). *Tgif2* functions are also related to central nervous system development, and mutations in *Tgif* genes can cause holoprosencephaly (HPE), a developmental disorder affecting craniofacial development which can also be caused by mutations in *SHH* (Taniguchi *et al.,* 2012).

In addition to known functions, the consistency of expression patterns in both human MBs and the SB murine tumours can also be used to help prioritise these candidate genes for downstream functional analysis. Because GNPCs are extensively used to analyse SHH signalling, two microarray based expression datasets are publicly available and can be used to further prioritise candidate genes prior to function analysis. The first (Subkhankulova *et al.,* 2010) analysed the gene *Bmi1* to investigate its role in cerebellar GNPC development, with expression data being generated from normal *Bmi1*+/+ GNPCs samples, *Bmi1*+/+ GNPCs treated with Shh, *Bmi1*-/- GNPCs, and *Bmi1*-/- GNPCs treated with Shh. The second was a study conducted by Gibson *et al.,* (2010) in which the cellular origins of WNT and SHH MB subtypes were investigated; expression data were generated from embryonic day 16.5 dorsal brainstem (E16.5 DBS), $Ctnnb1^{+/lox(ex3)};Tp53^{flx/flx}$ MB GNPCs, normal P7 GNPCs and $Ptch1^{+/-};Tp53^{-/-}$ MB GNPCs. Analysis of the expression of network genes in these datasets was

therefore also undertaken to identify those genes whose direction of expression change is consistent with CIS mode of action and/or expression in MBs, to provide further criteria to prioritise candidate genes, prior to functional analysis in GNPCs.

6.1.1. Aims:

The aims of the work presented in this chapter are, therefore:

6.1.1.1. To analyse network gene expression in both *Ptch1*+/- tumours / wt GNPCs, and SHH treated / untreated GNPCs using the data of Subkhankulova *et al.,* (2010) and Gibson *et al.,* (2010) to further prioritise candidate genes.

6.1.1.2. To use shRNA and/or overexpression of prioritised candidate genes to modulate gene expression in GNPCs, and assess the impact of this modulation upon the expression of genes involved in GNPC development and SHH signalling.

6.2. Results

6.2.1. Network gene expression in *Ptch1*+/- MB mimics expression in human MB

To provide further evidence for the importance of network genes in SHH MB, and to define the most promising genes for downstream functional analysis, the expression of the network genes was first analysed in the dataset of Gibson *et at.*, (2010). Their data from 3 wild type GNPC samples isolated from P7 days cerebella was compared to the data from 6 *Ptch1*+/- MB samples from the same developmental stage. The raw data, generated using the Affymetrix mouse 430_2 GeneChip, was first imported into GeneSpring, and subjected to standard normalisation (see Methods). A total of 81 human network genes have orthologous Affymetrix probes, and the expression of these orthologs in the cerebellar and *Ptch1*+/- MB samples was compared using T-tests. A total of 51 of these genes (single probe per gene) showed a statistically significant expression difference between wild type GNPCs and *Ptch1*+/- tumours, providing further confirmation that the network genes identified within the human expression dataset are relevant to SHH signalling. The expression of all 51 significant probes (with p-values ranging from 0.0114 - 9E-04) is shown visually in the heatmap in Figure 6.1.

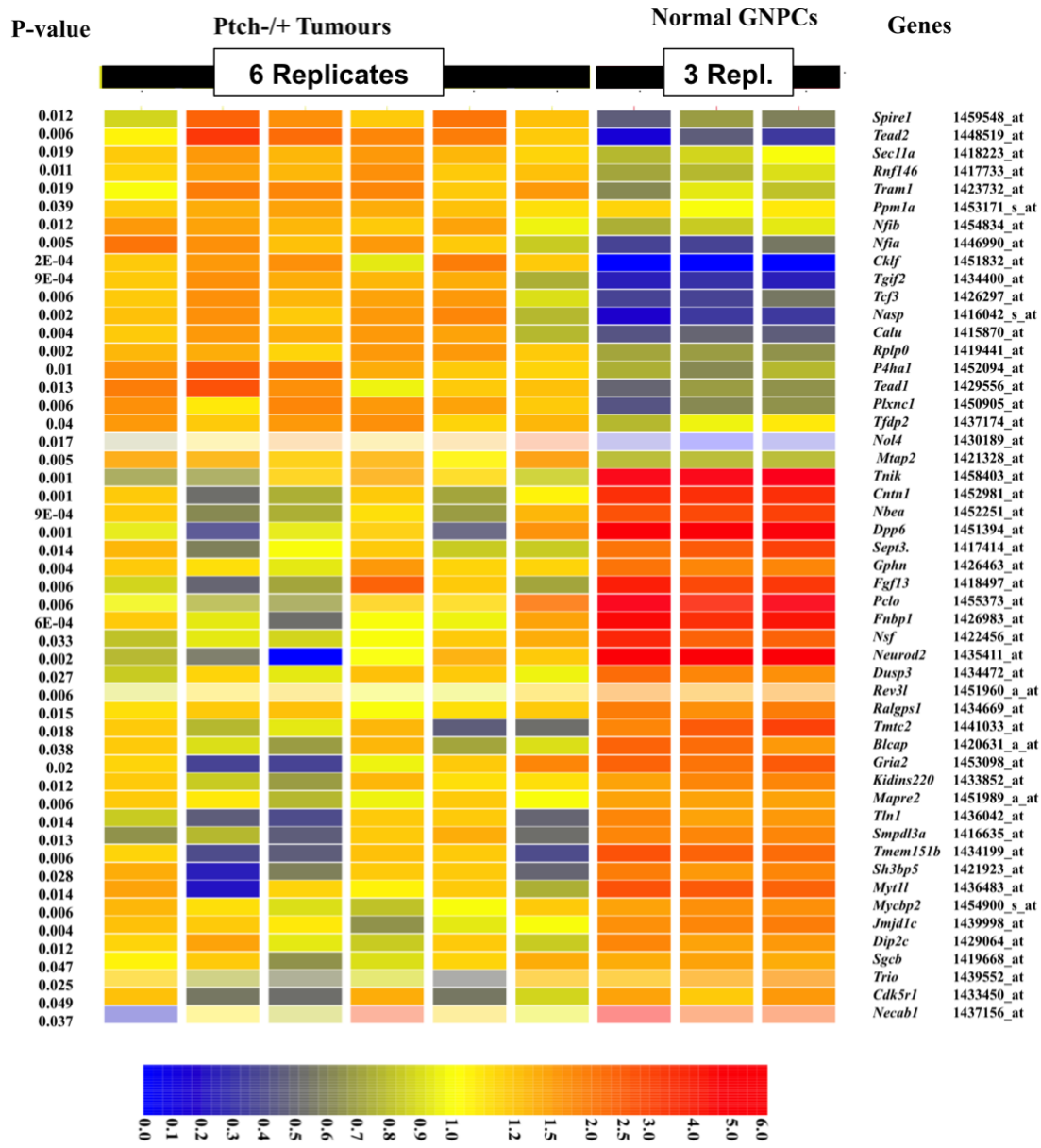

Figure 6.1: Expression of 51 networked genes showing significant differential expression. Expression heatmap of 51 genes within the CIS network which are differentially expressed between GNPCs from wild type mice and Ptch+/- MBs. One probe per gene is shown. The colour bar shows relative expression levels.

To present the expression of specific candidate genes more clearly, the expression of network CIS genes, networked transcription factors, and network genes with neuronal ontologies (which are enriched within the network) are shown separately in Figure 6.2. Amongst the 7 MB CIS genes in the network (Figure 6.2A), 3 genes (*Nfib*, *Tgif2* and *Tead1*) show a significant elevation of their expression within *Ptch1*+/- tumours compared to wild type GNPCs, consistent with their inferred mode of action from CIS data (see Table 3.6). In contrast, *Myt1l* shows downregulation within *Ptch1*+/- MB compared to wild type cells. This is also wholly consistent with the same

expression behaviour of these genes in human MB and the hypothesised mode of action within the transgenic model. Strikingly however, the expression of two genes (*Nfia* and *Fgf13*) is inconsistent with the MOA. *Nfia* is hypothesised to be down-regulated in the transgenic mice model on the basis of CIS data, but shows significant up-regulation in *Ptch1*+/- tumours, while *Fgf13* shows the opposite pattern. This inconsistency makes them less attractive for downstream analysis.

Regarding transcription factors other than CISs, the expression heatmap (Figure 6.2B) shows a significant over expression of three genes, namely *Cklf*, *Tcf3*, and *Tead2* within the *Ptch1*+/- MBs. This is also consistent with the expression of these genes within the SHH subgroup of human MB (see Figure 5.5).

Finally, the expression heatmap of neuronal genes in the network is presented in Figure 6.2C. These genes are expressed in general at low levels in *Ptch1*+/- MB compared to normal GNPCs. Specifically, *Pclo*, *Nsf*, *Neurod2* and *Nbea* show discernible lower expression within *Ptch1*+/- MB compared to wild type GNPCs. This behaviour of expression is very similar to expression in human MB. Some neuronal genes (*Gphn*, *Gria2*, *Mycbp2*, and *Cdk5r1*) do show lower expression within *Ptch1*+/- tumours compared to wild type, however, and one neuronal gene (*Ppm1a*) was expressed equally within both groups.

This general concordance of gene expression patterns within murine *Ptch1*+/- MB with the pattern of network activation seen in human MBs, provides further supporting evidence for the importance of these genes in MB development. In addition, the consistent behaviour of *Myt1L*, and *Tgif2*, together with the high p-values (*Tgif2*- p=6.83E-5 and *Myt1l*- p=0.00327), and their known functions, highlights these genes as prime candidates for downstream functional analysis.

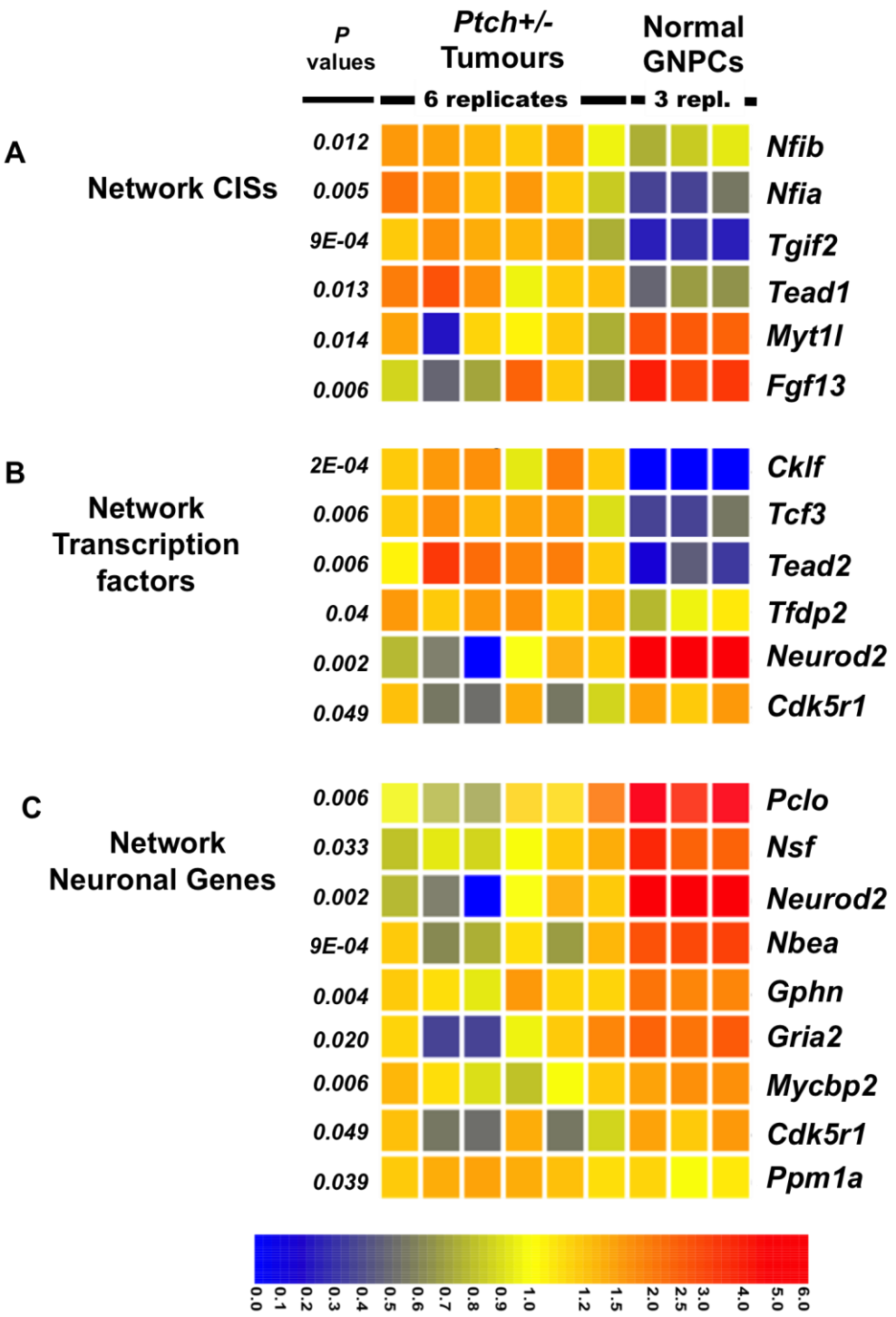

Figure 6.2: Expression heatmaps of selected network genes based on gene ontology (GO): Gene expression heatmaps of the network genes based on enriched GO specifications between normal GNPCs (3 riplicates) and Ptch+/-MB (6 riplicates) are shown. Probes with statistical significant values are shown. A. Expression of CIS genes which define the network. B. Expression of other 10 transcription factors in the network. C. Expression of 13 neuronal genes in the network.

6.2.2. Network genes exhibit differential expression in SHH treated GNPCs compared to normal GNPCs

To assess if the expression of the network genes is altered by Shh overexpression alone, as opposed to within tumours where multiple secondary alterations may have occurred, the data of Subkhankulova *et al.,* (2010) was also analysed. Specifically, gene expression within postnatal day P5-7 murine GNPCs with and without Shh treatment was compared. In this study the Affymetrix 430 murine GeneChip was used to analyse gene expression. A total of 81 human network genes have orthologous Affymetrix probes on this chip, and the expression of these was again compared between the two groups using T-tests. A total of 56 of these genes (single probe per gene) showed statistically significant expression differences between wild type GNPCs and Shh treated GNPCs (p-values ranging from 0.00214 to 2.26E-05, data not shown), confirming that the network genes identified within the human expression dataset are relevant to SHH signalling. The expression of network CIS genes, and network genes with transcription factor or neuronal ontologies (which are enriched within the network) are shown in Figure 6.3.

As expected, several CISs that define the neuronal network in human tumours show higher expression in Shh-treated GNPCs (*Tgif2*, *Tead1* Figure 6.3A), which is consistent with the inferred mode of action and qPCR (see Figure 4.11, 4.8 and 4.10 respectively). *Myt1l* expression is also consistent with the inferred mode of action from SB mutagenesis and qPCR (Figure 4.11 and 4.9), as it is lower in Shh treated GNPCs than untreated. Moreover, the expression of these genes is also consistent with human tumour data. *Fgf13*, *Nfia* and *Nfib* recorded a higher expression within Shh-GNPCs, also consistent with the inferred mode of action for these genes, with the exception of *Nfia*.

Transcription factors within the neuronal network show an obvious differential expression between the shh-GNPCs and the wild-type (Figure 6.3B). For instance, *Tead2* was expressed at a higher level within Shh-GNPCs, as it was in the *Ptch1*+/- MBs.

The expression of Neuronal genes in the network is presented in Figure 6.3C. As shown there are several genes have shown higher expression within Shh-GNPCs, such as *Cdk5r1*, *Nsf*, *Dlg1*, *Dcx* and *Gphn*. However, many neuronal genes shows downregulated expression within Shh-GNPCs, e.g. *Neurod2*, *Ppm1a*, *Mycbp2* and *Pclo*.

These results mimic those obtained from the network genes expression in human medulloblastoma data presented in Chapter 5. Moreover, many of these genes have shown a similar expression change in the mouse study where *Ptch1*+/- tumours were analysed (Gibson *et al.*, 2010, Figure 6.2C).

The above results indicate that the CISs identified from the SB whole body mutagenized mice, and the network genes they have defined, show broadly consistent differential expression between *Ptch1*+/- MBs and GNPCs, and between Shh-treated GNPCs and untreated GNPCs isolated from independent studies (Gibson *et al.*, 2010;Subkhankulova *et al.*, 2010). This suggests that not only are the network genes involved in MB development, their expression is also influenced by Shh signalling, providing further evidence for the importance of these genes specifically in SHH subgroup tumours.

Importantly, the results also help to define the most appropriate candidates for downstream analysis. While the expression of most CIS genes analysed are consistent with the inferred CIS mode of action, the data for *Nfia* and *Fgf13* are not consistent. This could complicate downstream analysis of this gene, and effectively eliminate it as a straightforward candidate for analysis. Of the remaining CISs, two stand out as showing consistent expression, relevant function, but having not previously been implicated in MB development. *Tgif2* (up-regulated) and *Myt1l* (down-regulated) exhibit consistent expression in all investigated datasets. Their expression is also consistent with the inferred mode of action caused by transposons insertions, with qPCR results from SB whole body mutagenised medulloblastoma samples, and with microarray expression data extracted from human MBs. Given the known function of these genes, both being involved in neuronal development, they are strong candidate MB genes. They were therefore chosen for downstream analysis within GNPCs through overexpression using ORF-cDNA (*Tgif2*), and knockout using shRNA (*Myt1l*).

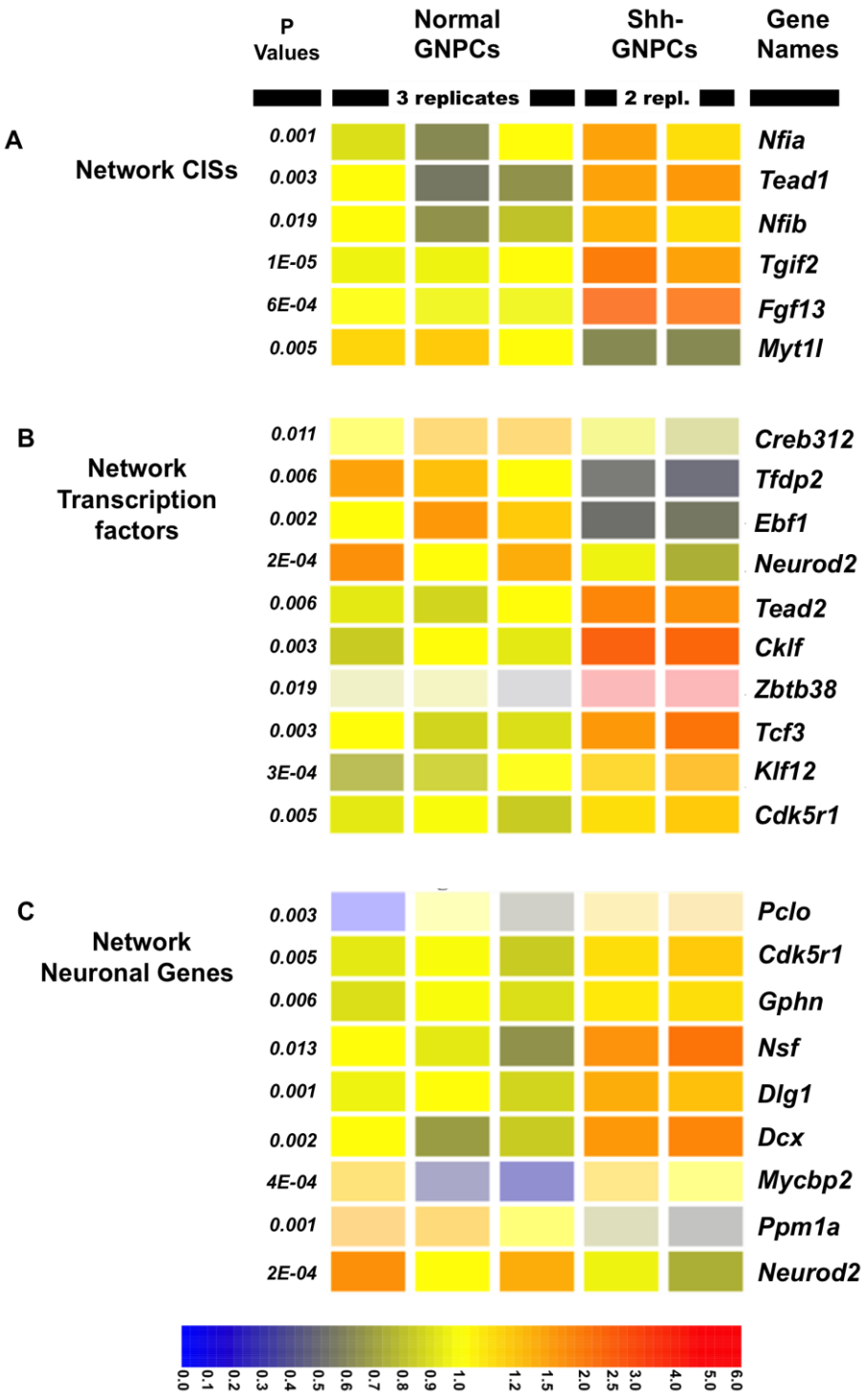

Figure 6.3: expression of CISs TF and Neuronal network genes in GNPCs. network genes expression according to GO analysis in Normal GNPCs (3 replicates samples) and shh treated GNPC (shh-GNPC) (2 replicates samples). P=values ranged from 0.011 – 3E-04 for TF genes and 0.013 – 4E-04 for Neuronal genes.

6.2.3. Isolation and preliminary analysis of GNPCs:

GNPCs were isolated from CBA wild type postnatal (P) day 4-7 mice, and cultured according to established protocols (see Methods). To confirm their ability to respond to Shh induction, cell counts were applied to GNPCs isolated from normal littermates at postnatal (P) day 4-7 in the presence or absence of Shh (3µg/ml) after 72hrs. This time point was chosen as it is appropriate to assess induction of proliferation upon Shh stimulation of these cultures (Wechsler-Rreya and Scott, 1999). Differentiation was also assessed visually. To account for biological sample variability, all experiments were conducted in triplicate on independently derived cultures of two groups –cerebellar GNPCs treated with Shh (Shh+/+) and cerebellar GNPCs without (Shh-/-). The results are shown in Figure 6.4.

Initially, the cells were seeded at a density of $3.65X10^5$ cell/cm^2/24 well plate. Shh was supplied at 3µg/ml at the seeding time where applicable. After 72hrs the cells were harvested and quantified automatically. As shown in Figure 6.4, the number of Shh+/+ GNPCs was significantly higher than the number of control Shh-/- GNPCs ($1.39X10^6$ and $0.65X10^6$, respectively p = 0.0007) after 72 hours. This suggested that with the addition of recombinant Shh protein , GNPC proliferation was induced significantly.

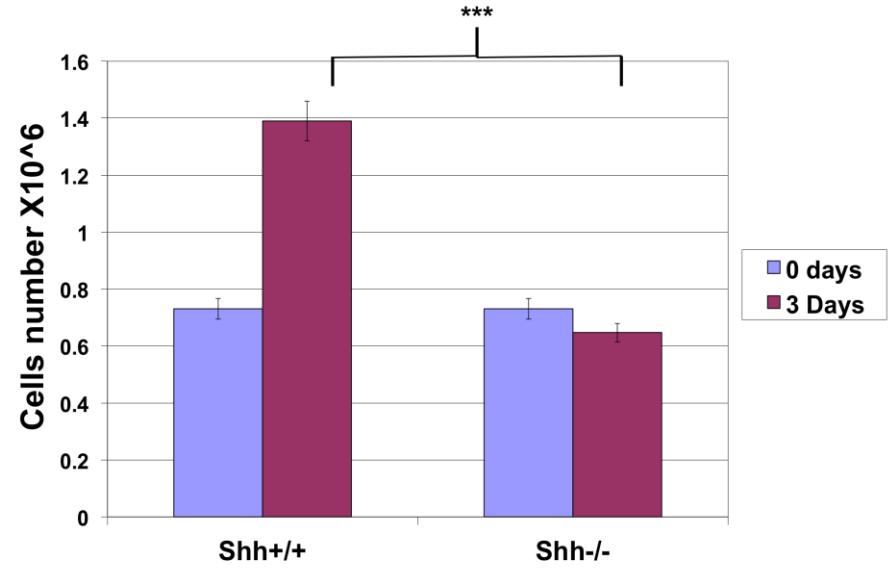

Figure 6.4: Assessment of Shh signaling pathway induction in GNPCs. Normal cerebellar GNPCs were cultured in presence of Shh (Shh+/+) or absence of Shh (Shh-/-) for 3 days. 0 day = normal GNPCs numbers at seeding time. (***p=0.00074).

6.2.4. Investigation of candidate gene expression in GNPCs

The expression of the two candidate CIS genes (*Myt1l* and *Tgif2*) was then investigated in cerebellar GNPCs treated with Shh or without, to establish the impact of the addition of Shhrecombinant Shh protein . In each case, expression was assessed in normal cerebellar GNPCs treated with Shh (Shh+/+ GNPCs), control normal cerebellar GNPCs treated with no Shh (Shh-/- GNPCs), and normal cerebellar GNPCs (0 day) that had been isolated and RNA extracted without culturing. The cells were cultured for 3 days and supplied with 3μg/ml Shh protein where applicable. For each sample, triplicate reactions were carried out to measure the average cycle threshold (Ct) value per sample. The average relative expression for each gene, relative to the 0-day control sample, was determined using the ΔΔCt method and expression was normalised to the reference gene *β-Actin*.

As shown in Figure 6.5, the expression levels of *Tgif2* was significantly elevated in Shh+/+ GNPCs (p= 0.037) compared to both Shh-/- GNPCs and 0-day GNPCs. This result is in agreement with the expression levels that had been recorded in both murine and human tumours (and the Shh treated murine GNPCs, Figure 6.3A), and suggests that activation of the Shh signalling pathway has the ability to upregulate *Tgif2* expression.

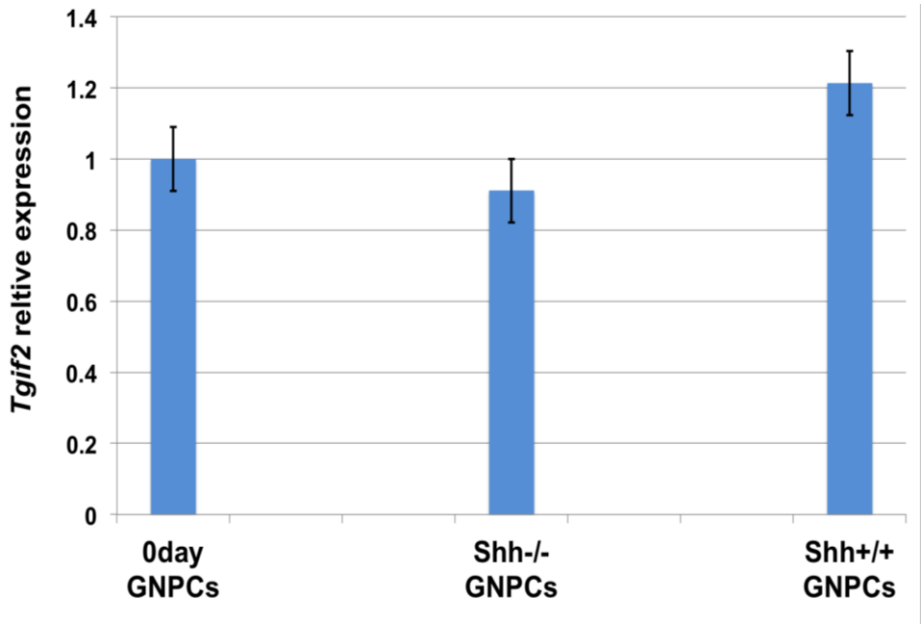

Figure 6.5: Expression of *Tgif2* in different cerebellar GNPC samples. qPCR results showing relative expression of *Tgif2* in isolated but not cultured GNPCs (0-day GNPCs), GNPCs cultured for 3 days without Shh recombinant (Shh-/-GNPCs), and GNPCs cultured for 3 days with the addition of 3µg/ml Shh recombinant are shown. The expression levels of *Tgif2* was significantly elevated in Shh+/+ GNPCs (p= 0.037) compared to both Shh-/- GNPCs and 0-day GNPCs The ΔΔCt method was applied to calculate the expression relative to 0-day cultures.

The expression of *Myt1l*, the other candidate CIS gene analysed here, is shown in Figure 6.6. Expression is slightly lower in Shh+/+ GNPCs compared to 0-day GNPCs but the difference is not statistically significant (p=0.45). The expression of *Myt1l* was, however, significantly higher in the Shh-/-GNPCs (p= 0.007). As Shh will be active in 0-day cells, this suggests that the effective withdrawal of Shh results in the upregulation of the pro-neural *Myt1l* gene. This is also in agreement with the *Myt1l* expression in murine and human medulloblastoma, and the murine GNPCs (Figure 6.3A). Moreover, these results reveal the impact of the Shh signalling pathway upon *Myt1l* and may reflect the maintenance of an undifferentiated phenotype, as *Myt1l* is known to function by driving neuronal cells toward differentiation.

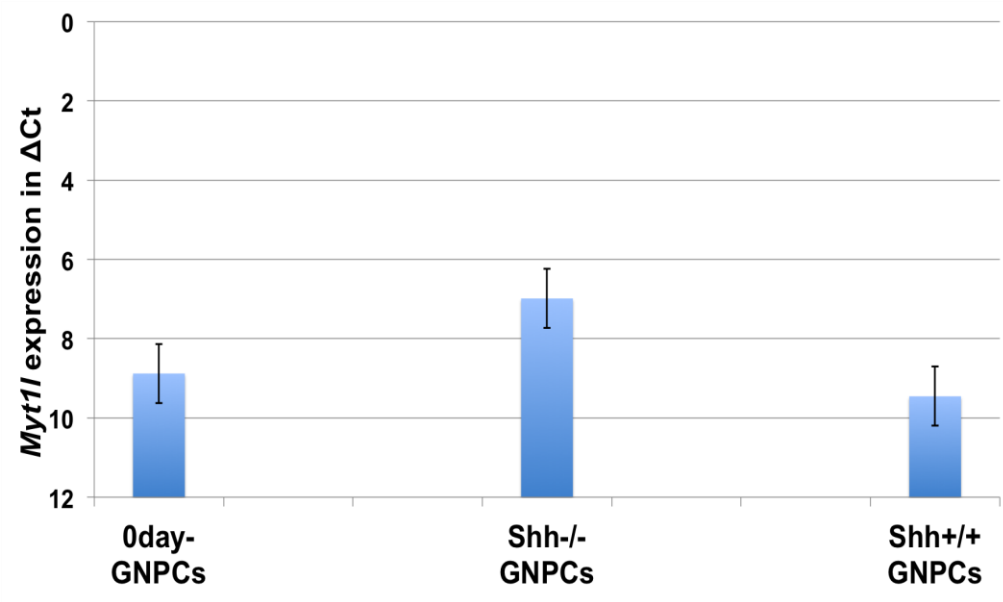

Figure 6.6: Expression of *Myt1l* in different cerebellar GNPCs groups. qPCR results showing ΔCt expression of *Myt1l* in isolated but not cultured GNPCs (0-day GNPCs), GNPCs cultured for 3 days without Shh recombinant (Shh-/- GNPCs) and GNPCs cultured for 3 days with the addition of 3µg/ml Shh recombinant (Shh+/+ GNPCs). The expression of *Myt1l* was, however, significantly higher in the Shh-/-GNPCs (p= 0.007).

In summary, the expression of both *Tgif2* and *Myt1l* in the Shh-treated GNPCs here is consistant with the expression data from Gibson *et al.*, (2010), and indicates that the expression of both *Myt1L* and *Tgif2* are influenced by Shh signalling. This is consistent with mutagenesis of these genes by SB enhancing or altering the timing of existing Shh transcription circuitry to affect tumour penetrance.

6.2.5 Expression of *Gli1* and *Math1* in Shh-treated CGPCs.

To prove that the Shh signalling pathway is stimulated byrecombinant Shh protein , expression of the transcription factor *Gli1* (a major effector of the Shh pathway – Wechsler-Reya and Scott, 1999) was also analysed in cerebellar Shh+/+ GNPCs, Shh-/- GNPCs and 0-day GNPCs (Figure 6.7). Consistent with Shh pathway activation, the expression of *Gli1* was significantly elevated in Shh+/+GNPCs relative to both controls Shh-/-GNPCs and 0-day GNPCs. Furthermore, the expression of *Gli1* was completely absent in Shh-/- GNPCs, consistent with downregulation of the Shh pathway relative to 0-day cells. These results confirm that the Shh pathway is efficiently stimulated in this GNPC culture model upon addition of recombinant Shh protein.

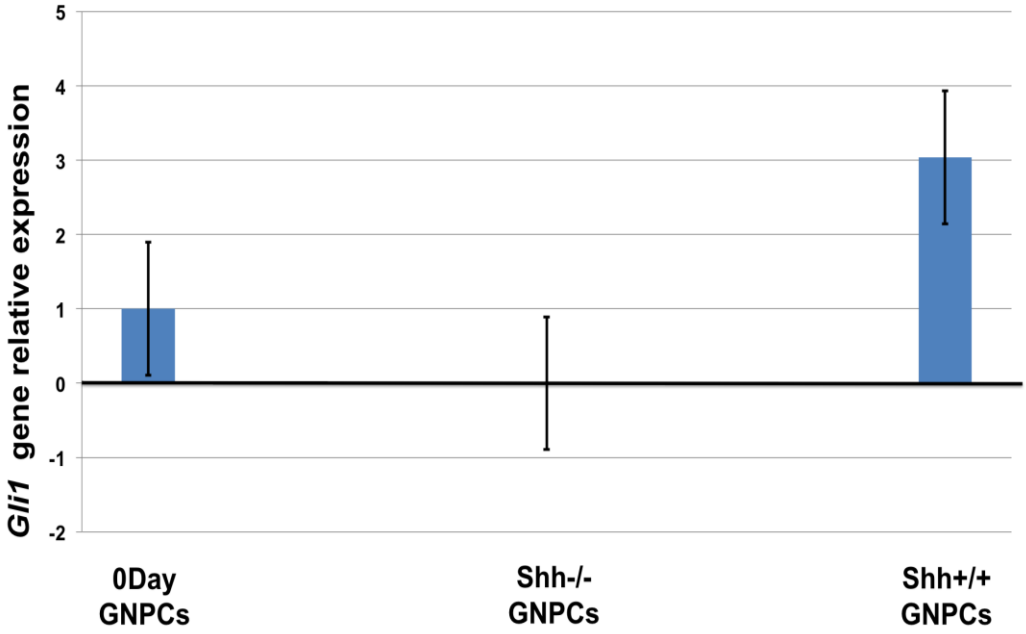

Figure 6.7: Expression of *Gli1* in different cerebellar GNPCs groups. qPCR results of relative expression of *Gli1* in isolated but not cultured GNPCs (0day-GNPCs), GNPCs cultured for 3 days without Shh recombinant (Shh-/-GNPCs) and GNPCs cultured for 3 days with the addition of 3µg/ml Shh recombinant. The ΔΔCt method was applied to calculate the relative expression.

Finally, as a marker of proliferation and an undifferentiated phenotype in cerebellar GNPCs, expression of the early neuronal marker *Math1* was also analysed (Wechsler-Reya and Scott, 1999). *Math1*, as a neural specific TF, is expressed predominantly in the external germinal layer of the developing cerebellum (Alder *et al.*, 1996). Helms and Johnson, (1998) provided evidence that *Math1* was expressed exclusively in proliferating and undifferentiated GNPCs. Thus, *Math1* expression should be high in undifferentiated neuronal cells. As shown in Figure 6.8, the expression of *Math1* was significantly upregulated in Shh+/+ GNPCs relative to both controls Shh-/- GNPCs and 0-day GNPCs. Furthermore, *Math1* expression was observed to be absent in Shh-/- GNPCs. These results confirm that after 72 hours of culture in the presence of Shh, the GNPCs are still expressing this marker of undifferentiated neuronal cells, consistent with the expected impact of Shh signalling upon cerebellar GNPCs. It also confirms that withdrawal of Shh leads to loss of *Math1* expression.

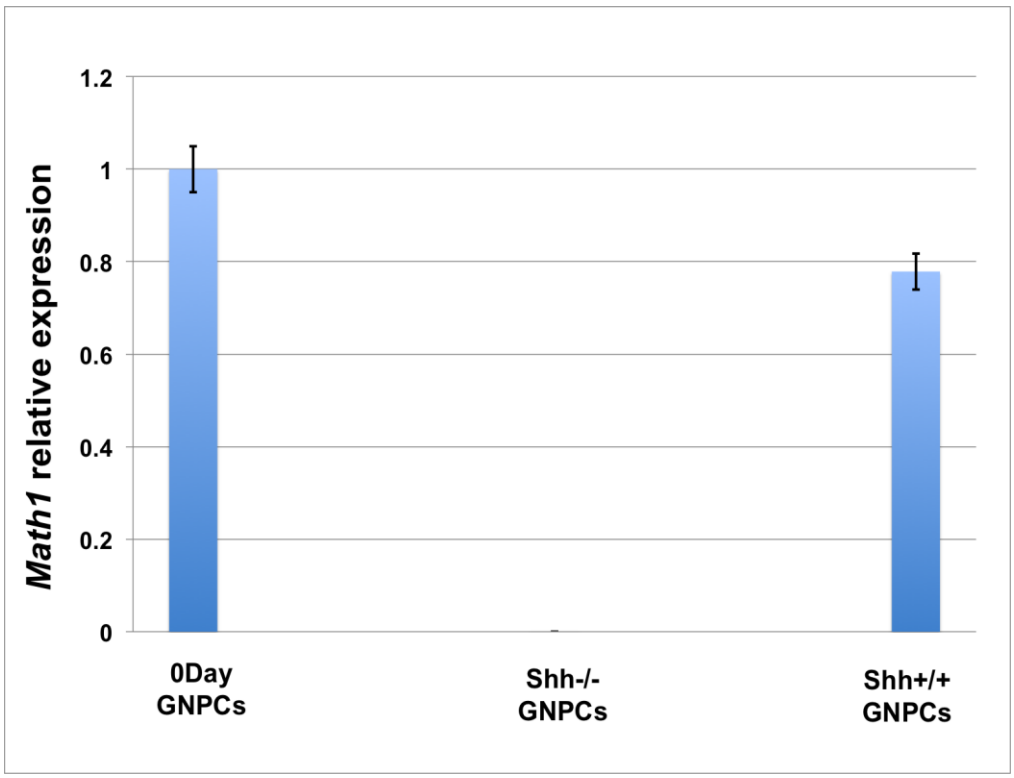

Figure 6.8: Expression of *Math1* in different GNPCs groups. qPCR results of relative expression of *Math1* in isolated but not cultured GNPCs (0 day-GNPCs), GNPCs cultured for 3 days without Shh recombinant (Shh-/-GNPCs) and GNPCs cultured for 3 days with the addition of 3µg/ml Shh recombinant. The ΔΔCt method was applied to calculate the relative expression.

All of the results obtained with the GNPC cultures, both in terms of proliferation, expression of candidate genes, and expression of *Gli1* and *Math1*, are consistent with these primary cells being an appropriate model in which to examine the role of *Tgif2* and *Myt1l* in both GNPC development and *Shh* signalling using lentiviral transduction techniques. Appropriate lentiviral clones, utilising a Puromycin selection system, were therefore obtained from Gencopia (http://www.genecopoeia.com) and used for all subsequent experiments.

6. 2.6. Puromycin kill curve to optimise transduction selection conditions.

Having confirmed that the cultured GNPCs were reacting to Shh as anticipated, I then performed a kill curve to identify the optimal puromycin concentration for use during transduction experiments. This step is required because cell lines/types can vary extensively in their sensitivity to puromycin.

The cerebellar GNPCs were seeded in a 24 well plate at $2.3X10^5$ cell/well, and supplied with 3µg/ml of recombinant Shh protein as the puromycin works most

efficiently when the cells are highly proliferative. The concentrations of puromycin used were 0µg/ml, 0.5µg/ml, 1µg/ml, 2µg/ml, 4µg/ml and 8µg/ml. The cells were harvested and quantified at day 2, 4 and 6 after the seeding time. As shown in Figure 6.9, GNPS are very sensitive to puromycin, as concentrations of puromycin as low as 0.5µg/ml are sufficient to kill cerebellar GNPCs after 4 days. As expected, control GNPCs with no puromycin showed significant growth and proliferation over the same period.

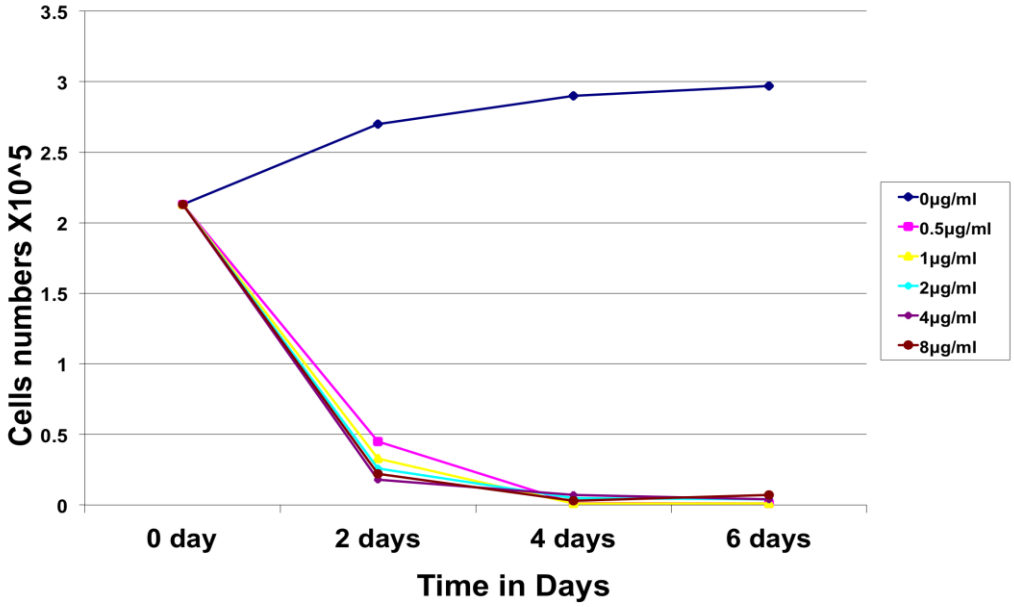

Figure 6.9: Puromycin kill curve for GNPCs. Wild type cerebellar GNPCs were isolated and cultured for 6 days. All groups were supplied with 3µg/ml Shh recombinant. Puromycin concentration is colour coded (see key).

6.2.7. Over-expression of *Tgif2* in GNPCs

To artificially upregulate *Tgif2* in CGNPs, an expression clone containing a CMV promotor and puromycin cassette was obtained from Genecopia (http://www.genecopoeia.com). This ORF clone was constructed to include an N-terminal FLAG-epitope tag in order to facilitate the detection of the protein produced by western blot (WB) and immunocytochemistry (IM) techniques. As a preliminary analysis to establish titre and transduction conditions, the ORF-*Tgif2* clone, and a negative control empty vector, were separately transduced into 3T3 cells (see Methods).

6.2.7.1. Preliminary analysis of ORF-Tgif2 transduced cells

The preliminary experiment was performed in 3T3 cells to check the quality and integrity of the system and to reduce both cost and time required. 3T3 cells were seeded and transduced with ORF-*Tgif2* lentiviral stocks at a range of Multiplicity of infection (MOI) as follows: 1 MOI, 5 MOI and 10 MOI, in order to check system efficiency. As a negative control wt-3T3 cells were cultured at the same seeding density until the end of the experiment. Then the transduced cells were selected using puromycin at a concentration of 2μg/ml, administered 24 hour after transduction. One 3T3 cell line culture was used to check the puromycin kill efficiency. After three days post transduction, cells were harvested and mRNA was isolated.

To establish if the FLAG-epitope-tag was expressed along with ORF-*Tgif2*, a standard PCR assay was conducted to analyse mRNA in all experimental transduced 3T3 cell lines and the wt-3T3 cell line. Vector isolated from transformed bacteria was also analysed. This was performed using primers spanning the FLAG-tag and ORF-*Tgif2*. As expected, amplified products were obtained from all experimental transduced 3T3 cell line with ORF-*Tgif2* and the plasmid (Figure 6.10). The expected amplicon size is ~199bp. There was no product from the wt-3T3 cell line negative control. β-Actin gene expression was used as an endogenous control, and all samples gave the anticipated amplification products of ~133bp, except for the *Tgif2* plasmid which gave a product size of ~228bp derived from the bacterial genomic DNA. The explanation for this is the primers were designed to span an exon-exon boundary. Therefore, the product is larger from plasmid DNA isolated from the bacteria, as the intronic sequence is still between the two exons. However, in the mouse cDNA the intronic sequence is absent. Collectively, these results indicate that the ORF-*Tgif2* is being transcribed within the cell line system.

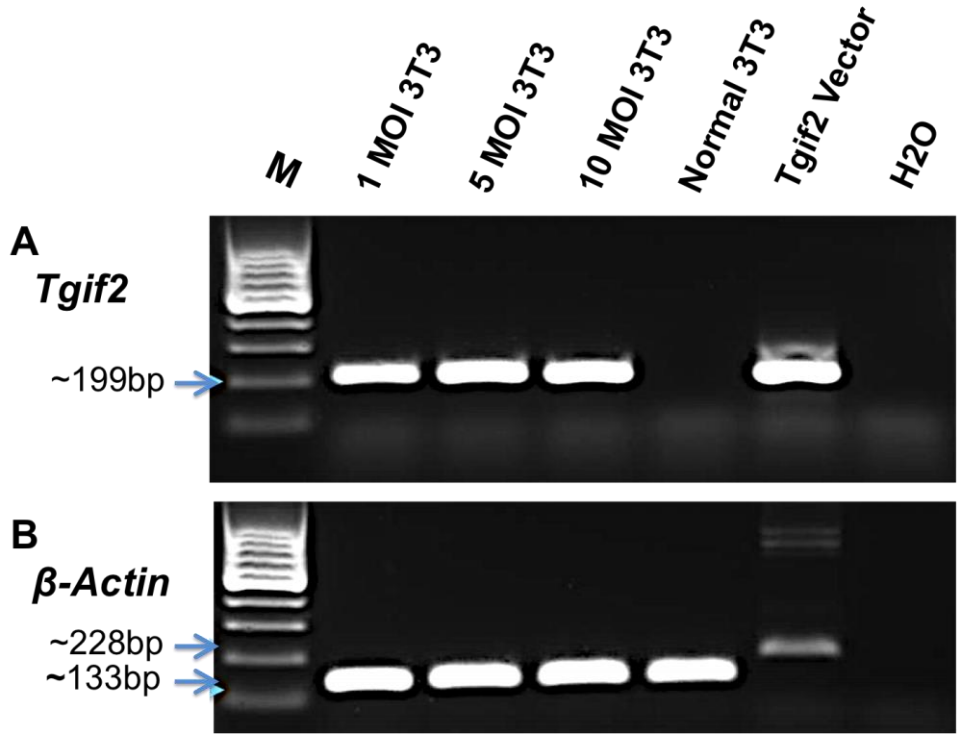

Figure 6.10: Expression analysis of ORF-*Tgif2* and FLAG-epitope tag in the transduced 3T3 cell line. Standard PCR results using cDNA templates generated from RNA isolated from 3T3 cell line transduced with different MOI as follows: 1, 5 and 10 MOI. (A) Panel are primers used span part of ORF-*Tgif2* and the FLAG-epitope tag to give an amplicon size of ~199bp. Untransduced 3T3 cells cDNA was used as a negative control, lentiviral ORF-*Tgif2*-FLAG-epitop tag as a positive control. (B) *β-Actin* (133bp/228bp) was included as an endogenous control.

To quantify expression of the ORF-*Tgif2* at different MOIs, Real Time PCR expression analysis of all transduced 3T3 cell cultures and the negative control were also performed, and the results of these are shown in Figure 6.11. As anticipated, the *Tgif2* transcript was expressed in transduced cells at a level dependent upon MOI , with the fold changes relative to the negative control being 91.9X, 198.1X and 572.8X for 1 MOI, 5 MOI and 10 MOI respectively. This indicates that the vector expressing the cDNA for *Tgif2* is working appropriately at the transcription level for tranduction into GNPCs.

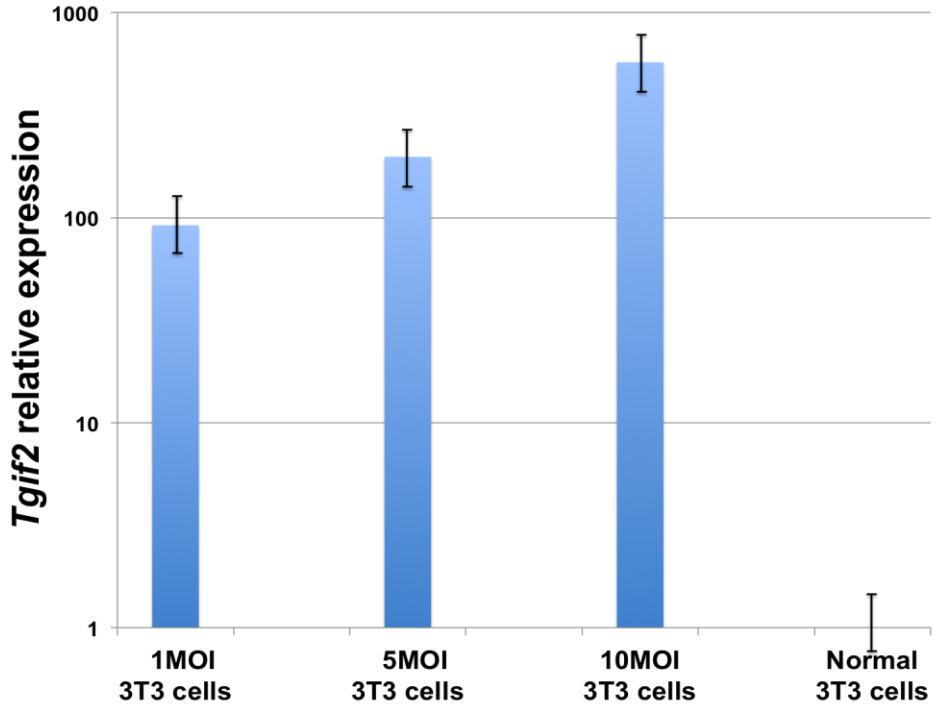

Figure 6.11: Relative expression of *Tgif2* in transduced 3T3 cells. The cells were harvested 3 days after transduction with Lentivirus expressing the cDNA for *Tgif2*. Samples are shown according to different viral MOI (1 MOI, 5 MOI and 10 MOI). qPCR expression was calculated using *β-Acin* as a reference gene and is shown relative to control samples with no viral transduction. The expression axis is shown in a log scale, with the normalized untransduced control sample set to 1.

6.2.7.2 investigation of *Tgif2* protein expression in transduced 3T3 cells

Next, Western blotting was performed in order to establish if the *Tgif2* transcript was being translated, and to establish the level of *Tgif2* protein in the 3T3 transduced cell line. Whole cell protein lysates were isolated from all transduced cell groups and the wt-3T3 control cell line. An independent FLAG-tag protein was utilised as a positive control. Different quantities of the protein, depending on availability, were utilised. The primary antibody (Ab) was applied at a concentration of 1:500. As depicted in Figure 6.12, no *Tgif2* protein could be detected in any of the transduced 3T3 cell lines, regardless of the MOI; the expected protein size is 28Kd. Moreover, there was no expression detection from wt-3T3, with the only expression observed from the independent positive control. *β–Actin* expression was, however, detected as an endogenous control. Importantly, these results revealed that the primary Ab was working efficiently, and that the protein lysates were of sufficient quality for Western analysis.

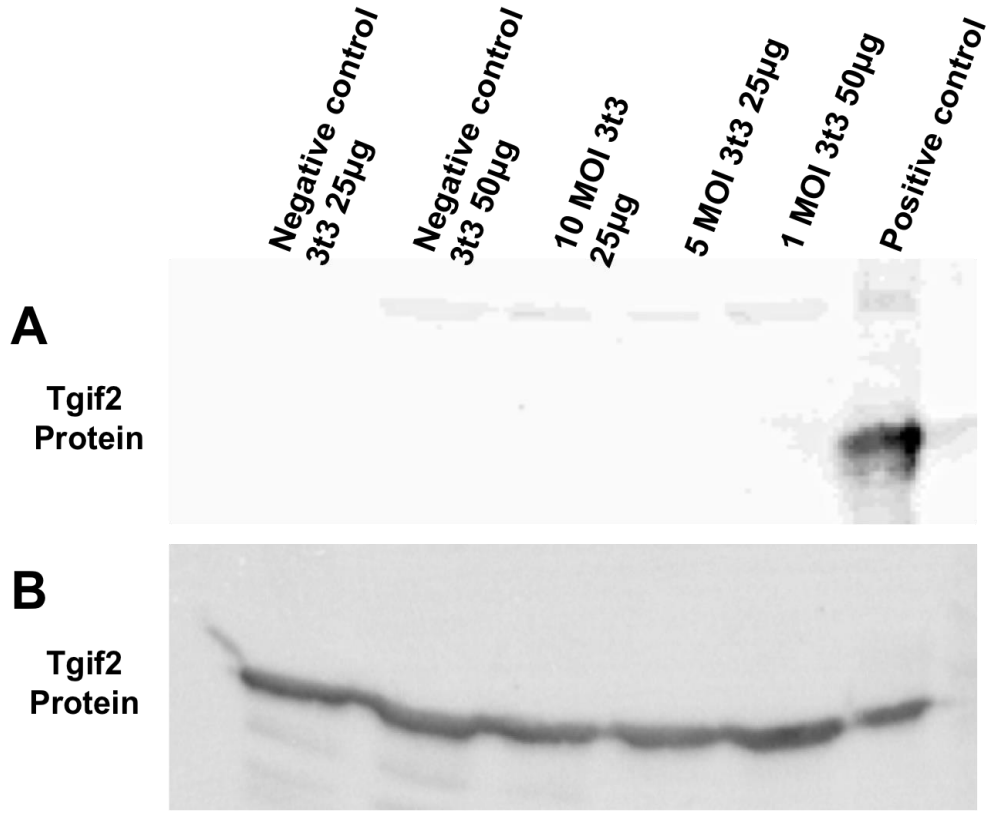

Figure 6.12: Western blot analysis of *Tgif2* in transduced 3T3 cells. Whole protein lysate isolated from transduced 3T3 cells were used (1 MOI; 50µg, 5 MOI; 25µg and 10 MOI; 25µg), wild type 3T3 cell were used as a negative control (25µg and 50µg) and ~15µg of a lysate from a cell line expressing an independent FLAG-epitope tag protein was used as a positive control. β-Actin was used as an endogenous control.

These results established that the ORF-*Tgif2* construct was efficiently transduced into 3T3 cells, was expressed at a high level at the transcriptional level, but that no FLAG tag could be detected. This indicated that further analysis of the system, both in terms of further blotting experiments and sequence analysis of the vector, would be required to establish the reason for the negative result with the anti-FLAG antibody. However, due to time limitations the analysis of *Tgif2* was suspended at this stage to concentrate on the analysis of the other prime candidate, *Myt1l*.

6.2.8. *Myt1l* knockdown in GNPCs

To investigate the functional impact of *Myt1l* upon medulloblastoma generation, lentiviral short hairpin RNA (shRNA) gene silencing technology was used. GNPCs were transduced with lentiviral vector expressing short hairpin RNAs for *Myt1l* (shRNA-*Myt1l*). Two shRNA-*Myt1l*s were utilised for transduction, namely shRNA-

Myt1l-8 and shRNA-*Myt1l*-9. Moreover, a mock knockdown lentiviral empty vector (shRNA-Mock) was utilised as a knockdown control. GNPCs were isolated from P6 mice, and seeded at a density of $5X10^5$cells/cm^2 in 96-well plates. The experimental groups were as follows; shRNA-*Myt1l*-8, shRNA-*Myt1l*-9, shRNA-Mock and Shh+/+GNPCs with no transduction. The viral stocks were applied 24 hours post seeding time at a MOI of 3 where applicable. Shh+/+GNPCs were also cultured to assess puromycin selection efficiency. All groups were supplied with Shh recombinant Shh protein at 3μg/ml upon seeding. mRNA was isolated from the transduced puromycin resistant GNPCs 3 days post transduction.

The relative expression of Shh+/+GNPCs, shRNA-Mock, shRNA-*Myt1l*-8 and shRNA-*Myt1l*-9 was quantified using the ΔΔCt methods, and the results are shown in Figure 6.13. *Myt1l* mRNA expression in shRNA-*Myt1l*-9 GNPCs was decreased by about 66% compared to Shh+/+GNPCs. A modest and non significant ~6% downregulation was also observed with shRNA-*Myt1l*-8. However, the shRNA-Mock samples showed significantly higher expression of *Myt1l* compared to Shh+/+GNPCs. This contrasts with the earlier analysis of GNPCs which indicated no significant impact on *Myt1l* expression after 3 days of Shh culture, and suggests that transduction *per se* may impact upon *Myt1l* gene expression. Irrespective of the underlying cause of the increase observed with the empty vector, these results indicate that efficient knockdown is obtained with shRNA-*Myt1l*-9.

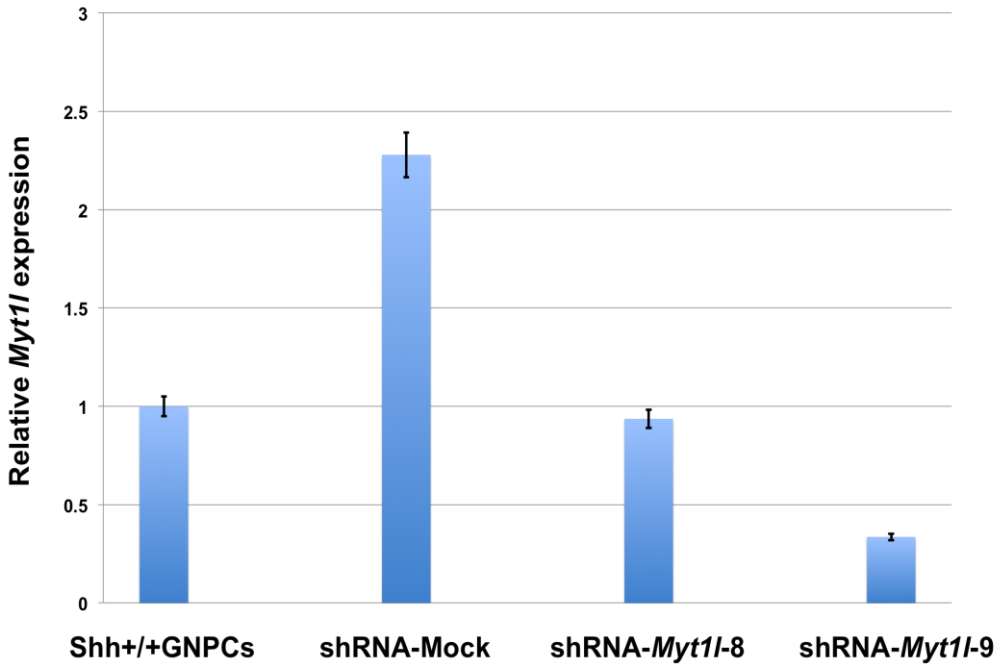

Figure 6.13: Expression of *Myt1l* in different transduced GNPCs groups. qPCR results showing relative expression of *Myt1l* in mRNA isolated from 0-day Shh+/+GNPCs, transduced GNPCs with virus expressing the empty vector (shRNA-Mock), and transduced GNPCs with lentivirus expressing shRNA-*Myt1l*-8 (ShRNA-*Myt1l*-8) or shRNA-Myt1l-9 (shRNA-*Myt1l*-9). All transduced samples were cultured in the presence of 3µg/ml Shh recombinant. The cells were harvested after 3 days of transduction based on 1µg/ml of puromycin selection. Expression was calculated using the ΔΔCt method relative to 0-day Shh+/+ GNPcs. β-Actin was used as an endogenous reference gene.

6.2.8.1 Impact of Myt1l knock down on Gli1 and Math1 expression

To begin to investigate the impact of *Myt1l* knock down on Shh signalling and neuronal differentiation, the mRNA expression levels of *Gli1* and *Math1* were analysed in all GNPCs groups. As shown in figure 6.14, the expression of *Gli1* was significantly (P=0.00018) higher (~2.4X) in shRNA-*Myt1l*-9 GNPCs relative to Shh+/+GNPCs. This result suggests that *Myt1l* knock down may further stimulate the Shh signalling pathway. Furthermore, the level of *Gli1* was reduced in the mock transduction where *Myt1l* expression was increased relative to 0-day cells (Figure 6.14), further suggesting that *Myt1l* expression affects the expression of Shh pathway genes.

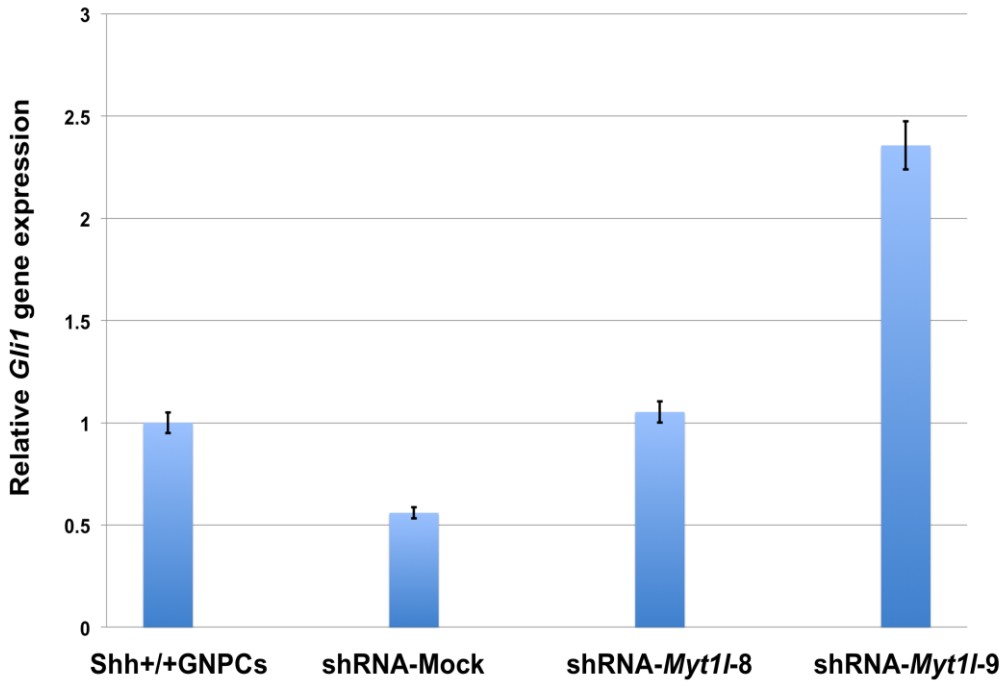

Figure 6.14: Expression of *Gli1* in different transduced GNPCs groups. qPCR results showing expression of *Gli1* in mRNA isolated from tranduced GNPCs with viral expressing empty vector (shRNA-Mock), and transduced GNPCs with lentiviral expressing shRNA-Myt1l-8 (ShRNA-Myt1l-8) or shRNA-Myt1l-9 (shRNA-Myt1l-9). Expression is shown relative to 0-day Shh+/+GNPCs. The expression of *Gli1* was significant (P=0.00018) in shRNA-*Myt1l*-9 GNPCs relative to Shh+/+GNPCs. All samples were cultured in the presence of 3µg/ml Shh recombinant protein. The transduced cells were harvested after 3 days based on 1µg/ml of puromycin selection. Relative expression was calculated using the ΔΔCt method. ß-Actin was used as an endogenous reference gene.

The expression of *Math1* (a marker for undifferentiated neuronal cells) was also found to be significantly different between the transduced groups, as shown in figure 6.15. The expression of *Math1* was significantly (P= 3.39752E-05) higher in shRNA-*Myt1l*-9 (~2.5X) relative to 0-day Shh+/+GNPCs, while the expression of *Math1* was modestly increased in shRNA-*Myt1l*-8 (~1.5X). This result is consistent with an inverse relationship between the expression of the 2 genes; the expression of *Math1* is increased as *Myt1l* expression is decreased. However, there was no significant difference between *Math1* expression in the shRNA-Mock compared to Shh+/+GNPCs.

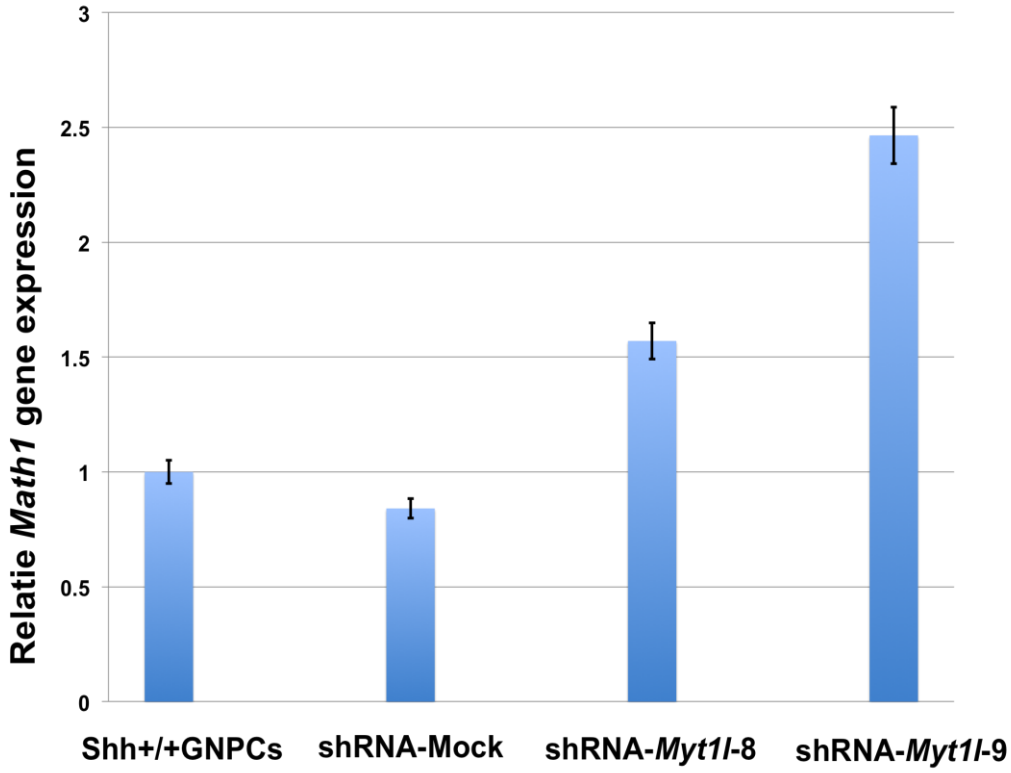

Figure 6.15: Expression of *Math1* in different transduced GNPCs groups. QPCR results of relative expression of *Math1* gene in mRNA isolated from Shh+/+GNPCs, tranduced GNPCs with viral expressing empty vector (shRNA-Mock), transduced GNPCs with lentiviral expressing shRNA-*Myt1l*-8 (ShRNA-*Myt1l*-8) or shRNA-Myt1l-9 (shRNA-*Myt1l*-9). The expression of *Math1* was significant (P= 3.39752E-05) in shRNA-*Myt1l*-9 relative to 0-day Shh+/+GNPCs. All samples exemplify cultured cells in the presence of 3µg/ml Shh recombinant. The cells were harvested after 3 days of transduction based on 1µg/ml of puromycin selection. Relative expression calculated based on ΔΔCt method. β-Actin was used as a an endogenous reference gene.

6.2.8.2 Investigation of Myt1l expression in Brain and GNPCs protein by WB

To confirm that *Myt1l* silencing was successful at the protein level, investigation of *Myt1l* protein is required in both mouse brain and cerebellar GNPCs. Initially, the primary Ab was optimised in P4-6 mouse brain and cerebellar GNPCs. Firstly, whole tissue protein lysate was isolated using urea buffer (see Methods). Isolated protein concentration was established via a BCA protein assay (Thermo Scientific). Aproximately 50µg of whole protein lysate was examined with different primary Ab concentrations (1:200, 1:500, 1:1000 and 1:2000) (Figure 6.16). Whole protein lysate isolated from 3T3 cell line (30µg) was used as a negative control, as *Myt1l* is not expressed in this cell line. As shown in Figure 6.16, *Myt1l* protein was detected in

whole brain tissue at all the primary Ab concentrations used. In addition, there was no expression observed in the 3T3 cell line.

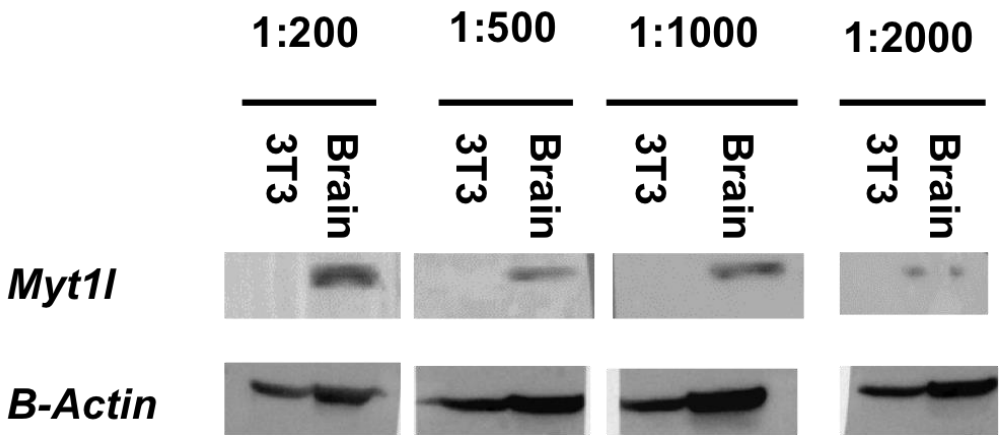

Figure 6.16: Preliminary investigation of *Myt1l* protein expression in whole mouse brain tissue. Whole mouse brain tissue was investigated to detect *Myt1l* protein expression using different anti-*Myt1l* antibody (1:200, 1:500, 1:1000 and 1:2000) with 50μg of brain tissue. β-Actin was used as endogenous control. *Myt1l* protein was also investigated in 3T3 cells as a negative control.

Since the primary Ab successfully detected the expression of *Myt1l* in mouse brain, an investigation of *Myt1l* protein expression in GNPCs was undertaken. Whole cell protein lysate was isolated from 0-day GNPCs, Shh+/+ GNPCs and Shh-/- GNPCs, using urea buffer. Moreover, protein from P4-6 day mouse brain tissue was used as a positive control, and the 3T3 cell line was used as a negative control. Western blot results are shown in Figure 6.17. No expression of *Myt1l* was detected in 0-day GNPCs, Shh-/- GNPCs or Shh+/+ GNPCs, although expression was clearly observed in brain tissue.

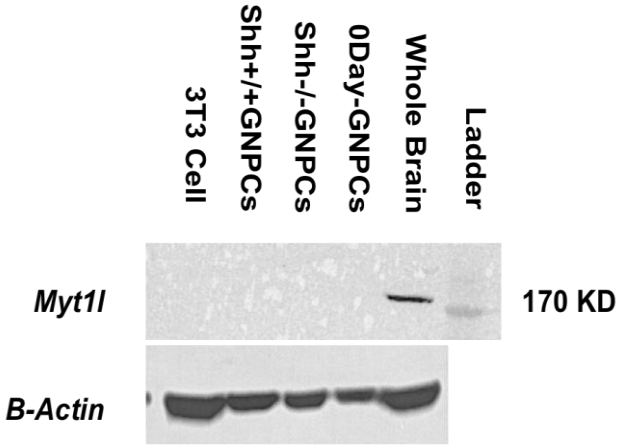

Figure 6.17: Western blot analysis of *Myt1l* gene in mouse brain and GNPCs. Whole protein lysates isolated from P6 mouse brain, 0 Day-GNPCs, Shh-/-GNPCs, Shh+/+GNPCs and the 3T3 cell line were examined for *Myt1l* expression by WB. 1:1000 dilution of *Myt1l* Ab was applied. The product size is ~170KD. β-Actin was examined as an endogenous control.

These results suggest either that *Myt1l* is expressed at a very low level in GNPCs relative to brain, that the protein is not accessible to the antibody in a whole cell lysate of these cells, or that GNPCs express an isoform which does not react with the specific antibody used. Despite the inability to detect *Myt1l* protein in GNPCs here, the significant changes in expression of both *Gli1* and *Math1* observed upon *Myt1l* shRNA knockdown strongly suggest that the knockdown is effective at the protein level. Confirmation of this will, however, require further work.

6.3. Discussion

In this chapter, I have outlined preliminary analysis of two candidate genes within murine GNPCs. As an initial step, the expression of neuronal network genes identified in the current study was analysed within two publicly available and independent datasets; one generated from wild type GNPCs and MBs from $Ptch1^{+/-}$ mice (Gibson *et al.*, 2010), the other generated from wild type GNPCs cultured with and without shh (Subkhankulova *et al.*, 2010). This step was performed to further prioritise the most promising candidate genes for functional analysis in terms of consistency of expression behaviour, in addition to novelty and known functional considerations.

The majority of the network genes showed significant differential expression in these datasets, providing further evidence for the connectivity of the network. Interestingly, the CISs which define the network, and the neuronal genes within it, showed consistent expression changes within the two datasets; genes expressed more highly in *Ptch1*+/- tumours relative to GNPCs were also more highly expressed in shh treated GNPCs than untreated. Furthermore, the expression of most CIS genes was consistent with the inferred mode of action (MOA) from insertion site data. This provides further evidence that the network genes are relevant to both MB development, and more specifically to SHH signalling. A noteworthy exception to this is *Nfia*, the gene most frequently targeted by SB inserts identified here. The inferred mode of action is loss of function, due to the mixed orientation of insertions and their wide distribution within the gene. However, in both murine datasets analysed here *Nfia* was more highly expressed when shh signalling was active (*Ptch1*+/- tumours or shh treated GNPCs). *Nfia*, although potentially an important gene in SHH subgroup MB, was not chosen for downstream analysis because of this inconsistency which could complicate the interpretation of any results obtained. However, based on consistency of expression patterns in human and mouse, clarity of mode of action, novelty, and known function, two genes were chosen for downstream analysis; *Tgif2* and *Myt1l*.

Tgif2 is a member of *Tgif* family, known to repress neuronal genes and play an important role in neuronal development. Mutation of *TGIF1*, a paralog of *TGIF2* have been shown to cause some forms of holoprosencephaly (HPE) (Gripp *et al.*, 2000), a serious genetic disease affecting human craniofacial development which can also be caused by mutations in SHH. Recently, Taniguchi *et al.*, (2012) showed that mice lacking both *Tgif1* and *Tgif2* showed HPE-like phenotypes which mimic those seen in

Shh-/- mice embryos. The authors showed that growth factor β/Nodal signalling was disrupted in the TGIF double mutant. Tgif2 overexpression has also been found to repress a subset of neuronal genes such as *Cath1, Msx2, Pax6, and Wnt1* (Knepper *et al.*, 2006).

In the analyses presented here, *Tgif2* was expressed persistently at high levels in all tumour or SHH treated datasets, suggesting that this gene may have a role in MB generation. Consistent with this, SB insertions in *Tgif2* resulted in upregulation of the gene, expression being driven by the SB enhancer (see Figure 4.8), and its expression was significantly higher in the murine samples where shh is active when analysed in this chapter (see Figure 6.4 and 6.2).

Collectively, these results raise the possibility that *Tgif2* is a direct or indirect target of Shh, and that genes downstream of *Tgif2* promote tumorigenesis by repressing neuronal specific genes and maintaining an undifferentiated state. SB mutagenesis of *Tgif2* may therefore overexpress a gene normally upregulated during Shh signalling pathway. Finally, to my knowledge there has been no study that has investigated the role of *Tgif2* in MB generation, making it a prime candidate for downstream analysis.

The second gene I analysed, *Myt1l*, was already a good candidate gene prior to the analyses of murine data for several reasons; because of its position as the most highly networked gene (Figure 5.2), because it showed consistently lower expression within tumour generated by whole-body SB mutagenised mice and human MB data (Figures 4.9 and 5.5, respectively), because low expression was associated with poorer patients survival (Figure 5.3), and because of its known function as a pan neural differentiation gene (Pang *et al.*, 2011). Its candidature was further supported by the analysis of both murine datasets, as its expression was lower within *Ptch1*+/- tumours and Shh treated GNPCs compared to normal cerebellar GNPCs (Gibson *et al.*, 2010; Subkhankulova *et al.*, 2010).

Myt1l, myelin transcription 1-like, is a zinc finger gene (also called *Myt1* and *NZF3*) and belongs to the *Myt1* family. It is widely expressed in the murine developing brain. In one study, the highest expression of this gene was obseved in the adult murine brain and spinal cord, whereas no expression was observed in glial cells, suggesting that *Myt1l* is strictly expressed in neurons (Kim *et al.*, 1997). In terms of the role *Myt1l* plays in neuronal development, Vierbuchen *et al.*, (2010) found that *Myt1l* in combination with 2 other transcriptional factors, *Ascl1* and *Brn2*, can accelerate the transformation of mouse embryonic fibroblasts into functional neurons *in vitro*.

Additionally, *Myt1l* has the ability to enhance the conversion of human pluripotent stem cells into functional neurons in the presence of *POU3F2* and *Ascl1* (Pang *et al.*, 2011).

Prior to lentiviral transduction experiments, I first assessed the suitability of GNPC cultures for analysis of these genes by checking their response to shh signalling. This was achieved by culturing them with and without 3μg/ml of recombinant Shh protein for 3 days and assessing proliferation and expression of key reporter genes. This steps was essential to assess the ability of these cells to be maintained *in vitro*. I measured the expression of both candidate CISs *(Tgif2 and Myt1l)*, an effector of the Shh signalling pathway *(Gli1)* and a marker for neuron progenitor cells *(Math1)*. This analysis revealed that GNPCs proliferation was significantly higher within cells treated with Shh compared to wild type GNPCs, congruent with the known proliferative impact of Shh upon cultured GNPCs (Dahmane and Altaba, 1999). Additionally, initial qPCR expression analysis of *Tgif2* and *Myt1l* was consistent with the mode of action of SB insertion, with these genes showing up-regulation and down regulation respectively in Shh treated GNPCs compared to wild type (see Figure 4.7 and 4.8). Moreover, both *Gli1 and Math1* showed higher expression in Shh treated GNPCs compared to wild types, and this was expected as both are markers for Shh induction in GNPCs (Wechsler-Reya and Scott, 1999).

Having established the suitability of the primary cell culture model, I went forward to lentiviral based analysis of both *Tgif2* and *Myt1l*. I confirmed the elevated expression of *Tgif2* in Shh+/+GNPCs, and successfully transduced the lentiviral ORF-*Tgif2* clone into 3T3 cells, confirming gene expression at the mRNA level in a MOI dependent fashion. However I could not confirm expression at the protein level. This could be due to the translation period of the protein not being sufficient to be detected by WB. Moreover, I did not have time to check the viability of the FLAG-tag expression, or check the reading frame of the vector (although it was purchased as a validated intact ORF). It is also possible that the FLAG tag is not accessible to the antibody, due to specificities of the protein conformation. Owing to time constraints, troubleshooting of this issue had to be postponed for future work. Despite this, the results presented here suggest that *Tgif2* is a valid target for further analysis in SHH subgroup tumours.

My analysis of *Myt1l*, however, was more informative, as I successfully knocked down expression and showed that this increased *Gli1* and *Math1* expression, providing confirmatory evidence that *Myt1l* expression does affect both SHH signalling and neuronal differentiation. This suggests that GNPCs where *Myt1l* is knocked down

are highly proliferative and maintained in an undifferentiated state. It is also of interest that *Myt1l* expression and *Math1* expression are reduced as a result of Shh signalling, while both Shh signalling (as assayed by *Gli1* expression) and *Math1* expression are induced or enhanced as a result of *Myt1l* knock down. These results are consistent with these genes being functionally linked, and further validate the relevance of the gene expression network identified here to Shh signalling and MB development.

Additional work remains to be done in the analysis of *Myt11*, however, as I was unable to demonstrate knockdown at the protein level. The failure to detect protein expression by WB could be due to several reasons. First, *Myt1l* protein is localised in the nucleus, and the current analysis was performed upon whole cell protein lysates, in which dilution of *Myt1l* may take place. Furthermore, the time of the protein isolation may not have given enough time for high protein expression, as it was 3 days after transduction. In addition, the Ab used in these analyses was monoclonal, therefore the application of a polyclonal Ab would be worth trying as *Myt1l* has several protein isoforms.

Further experiments involving *Myt1l* silencing are desirable as the impact upon phenotype/proliferation remains to be assessed. Despite the need for further analyses, the results presented here provide initial validation of *Myt1l* as a key gene involved in the regulation of SHH signalling and GNPC development, and suggest that further analysis of this gene may identify targets for therapeutic assessment within SHH subgroup MB.

Chapter 7

General Discussion

In the work presented here I have demonstrated that MB development was enhanced significantly by the application of whole-body SB mutagenesis of the *Ptch*$^{+/-}$ murine model, and this was achieved without affecting tumour latency. Furthermore, this system has no tumour induction within wild type mice. However, for the RMSs that developed within this model there was no significant impact on mortality relative to predisposition control animals. This suggests that the mutagenesis had no direct impact upon MRSs development. The whole-body SB mutagenesis system utilised in the current study also produced a high frequency of haematological tumours, consistent with one other screen using this combination of transposase and mutagen (Collier *et al.*, 2009;Bender *et al.*, 2010), in addition to several solid malignancies at lower frequency. The generation of 19 liver adenomas in this screen was unexpected, as this tumour type has only been observed previously in a tissue specific screen using conditioned mobilisation of SB within the liver by using Cre under the control of an albumin enhancer (Keng *et al.,* 2009). The data here indicated that this tumour type can be induced by SB mutagenesis alone. The most significant CIS within the adenomas was in the 24th intron of *Egfr*, with inserts showing a virtually identical pattern to that observed by Keng *et al.*, (2009) where *Egfr* was also the top CIS. This demonstrates the specificity and power of SB mutagenesis, and the importance of this gene in terms of hepatocellular carcinoma.

The CISs identified within SB induced MB primary tumours, 11 of which had not been implicated in MB before, were found to be enriched for transcription factors and to have functions involved in neuronal development, differentiation and/or migration and cell death. The majority of the candidate genes also showed significant expression differences between the human MB subgroups. Low expression levels of one gene – *Myt1l* – were significantly associated with poor outcome within human MBs. Collectively, these data indicated that the MB CISs identified here are promising candidates for involvement in the disease process, and can be considered for further investigation to identify new therapeutic targets or MB markers that may be of use in the treatment or management of this disease.

Because of the significant excess of transcription factors in the MB CIS list, ARACNe expression network analysis was applied to human MB expression data to identify networks of transcription factors and their targets of relevance to MB. This is the first time, to my knowledge, that ARACNe expression network analysis has been applied to the results of a mutagenesis screen to gain insight into a disease process. Interestingly, 7 CISs were tightly linked within a single network. This network consists mainly of transcription factors and their target genes critical for neuronal differentiation and proliferation.

Further investigation of the network components via human heatmaps of gene expression revealed that network genes are differentially expressed in human MB subgroups and that expression of two of the CIS genes, *TGIF2* and *MYT1L*, showed strong correlation with network activity. These two CIS genes were of particular interest because of their impact upon the SHH signalling pathway and neuronal differentiation respectively. Both the inferred MOA of transposon inserts and qPCR validations revealed that *Tgif2* was showing oncogene-like behaviour. This is supported by a recent study where the double knock out of *Tgif1* and *Tgif2* showed the inhibitory effect upon Shh expression in the developing brain (Taniguchi *et al.,* 2012). In contrast, the inferred MOA and qPCR data suggested that *Myt1l* is likely to act as a tumour suppressor. Consistent with this, low expression of *MYT1L* was associated with poor survival rates within MB patients. Furthermore, this gene has been shown recently to be involved in the reprogramming of mouse and human fibroblasts into neuron-like cells (Ambasudhan *et al.,* 2011;Pang *et al.,* 2011), suggesting that reduced expression may help to maintain the undifferentiated cellular phenotype of MBs. To my knowledge neither of the above genes has been implicated in MB before, and both are valid targets for further functional analysis.

Importantly, analysis of network activity as a whole, both within human and mouse expression data, suggests that one of the major outputs of network perturbation during tumourigenesis is upregulation of the *Igf2* gene. Investigation of the genes differentially expressed between tumours with transposons hits in the CISs and those without hits in the CISs within the putative neuronal network, identified the most differentially expressed gene as *Ifg2*. My quantitative analysis of *Igf2* expression also revealed that *Igf2* expression was significantly higher within tumours with insertions within network CIS genes than those without. A previous study has concluded that *Igf2* expression is an absolute requirement for MB generation, as *Igf2*$^{-/-}$*; Ptch1*$^{+/-}$ mice do not develop tumours (Rao *et al.,* 2004). Foci of proliferating GNPCs are, however, still

observed on the surface of the cerebellum in juveniles mice of this genotype, suggesting that *Igf2* expression is required for transformation of pre-cancerous lesions. The identification of a neuronal gene network underpinning the expression of *Igf2* in GNPCs therefore elaborates upon the existing model of tumour formation within *Ptch1*+/- mice.

The summary above indicates that the integration of expression networks and mutagenesis data has been very informative here. It may, therefore, also be a useful tool in other mutagenesis studies, particularly in screens where modest numbers of CISs are identified. In large screens, it has been possible to infer gene networks directly from co-occurrence of CISs within tumours, For instance, a recent SB mutagenesis forward screen study to identify candidate genes involved in intestinal polyposis. March *et al.*, (2011) identified co-occurring networks, but this involved pairwise comparisons of >900 CISs.

Because of the strong candidature of both *Tgif2* and *Myt1L* I also initiated functional analysis and downstream investigation of these two CISs to provide further information concerning their role in MB formation. I over-expressed *Tgif2* within 3T3 cells by transducing retroviral costructs carrying ORF-*Tgif2*. The overexpression of *Tgif2* was quantified by qPCR, but expression at the protein level could not be confirmed. *Myt1l* was, however, successfully knocked down in GNPCs treated with recombinant Shh protein in culture. Two neuronal progenitor cell markers, *Gli1* and *Math1* were shown by qPCR to be upregulated in these cells. This is consistent with reduced expression of this gene maintaining the GNPCs progenitor characteristics and reducing neuronal differentiation. As far as I am aware this is the first demonstration of a link between *Myt1l* and the *Shh* signalling pahway. Unfortunately, *Myt1l* protein expression could not be detected within GNPCs.

It is also worth noting that the bioinformatic analysis of subgroup specific metagenes within human and mouse MB data, and especially the NMF projection, has provided the first comprehensive evidence that the *Ptch1*+/- mouse is an appropriate model of SHH subgroup MB at the level of gene expression. Although the tumours generated by this model resemble human SHH MBs in terms of pathology (Goodrich *et al.*, 1997), confirmation that gene expression is comparable indicate that this model can be usefully examined and manipulated to search for both tumour biomarkers and therapeutic targets for human MB.

Recently, an independent SB mutagenesis screen of the *Ptch1*+/- model has been conducted using a tissue specific SB mutagenesis screen system (Wu *et al.*, 2012). This was performed by expressing SB11 under the control of the *Math1* promoter in the

GNPCs of transgenic *Ptch1*$^{+/-}$ mice. This was a highly penetrant screen, with mutagenesis enhancing MB penetrance from ~39% (54/139) in no predisposition controls up to ~97% (271/279) in experimental (*Ptch1*$^{+/-}$*;Math1;SB11;T2/Onc*) mice. This contrasts with the screen reported here where the frequency of MB was only 6% in controls and ~23% in the experimental group. However, because of the high penetrance, 139 primary mice tumours were available for analysis resulting in the identification of 350 CISs, compared to 18 CISs from 41 primary MB generated here.

In addition to affecting penetrance, the tissue specific screen had a significant impact upon latency, reducing it from 8 to 2.5 months. This suggests that some of the mutations affected the aggressiveness/development of disease, as opposed to being transformative. Consistent with this interpretation, the *Math*1 conditional SB screen study generated metastatic disease via the cerebrospinal fluid pathways in approximately 80% of cases (Wu *et al.*, 2012), whereas there was no macroscopic evidence of metastatic disease within the tumours analysed here. This is important because metastases are uncommon both within the *Ptch1*+/- model (Goodrich *et al.*, 1997) and within human SHH tumours (Kool *et al.*, 2008).

The large number of CISs identified, and the large number of metastases, enabled (Wu *et al.*, 2012) to establish that the spectra of alterations in primary and metastatic tumours are distinct, highlighting the difference of the metastatic niche and the potential problems for therapeutic intervention based on analysis of primary tumours alone. Although the two screens are not comparable, Wu *et al*., 2012, identified a subset of 17 CISs with mutations in both alleles, suggesting that they were key genes in tumour development. Interestingly, 4 of the network CIS genes here were identified by Wu et al. (*Crebbp*, *Nfia*, *Nfib* and *Tead1*) and a further 3 are within the network (*Dip2c*, *Edil3* and *Erbb4*): All but one of these genes (*Tead1*) are biallelically mutated in the (Wu *et al.*, 2012) data, consistent with the network being critical for tumour formation rather than progression. It is also noteworthy that *Igfr1* and *Ifg2* were both identified as CIS genes within the SB conditioned mutagenesis screen of (Wu *et al.*, 2012), *Igfr1* being the main receptor that transfers *Ifg2* gene signals within GNPCs (Corcoran *et al.*, 2008). Thus, while the penetrance of the model and aggressiveness of the mutagen were much higher in the (Wu *et al.*, 2012) study, the footprint of the network identified here is present within their data.

The idea that GNPCs are the only known source of neuron granule cells and the cells of origin of MB has dominated thinking for nearly a decade, and has only recently been revised as other, distinct, cells of origin have been identified. Li *et al.* (2013)

identified a rare population of neuronal progenitor cells that, surprisingly, also generate granule cells and MB in mouse cerebellum. Although Nestin expression has been identified in the cerebellum [neural stem cells (NSCs) and radial glia specifically], its expression in GNPCs has remained debatable (Sotele *et al.,* 1994; Lee *et al.,* 2005). Some reports have suggested that Nestin is downregulated in these cells prior to commitment to the granule lineage (Alder *et al.,* 1996), although some studies have reported that GNPCs can be Nestin positive (Tanori *at al.,* 2010). This is complicated by the fact that it is difficult to identify Nestin-positive cells from Nestin-positive fibers. Several studies have targeted GNPCs through engineered animals that expressed Cre recombinase under the control of the Nestin promoter (Milles *et al.,* 2006; Frappart *et al.,* 2007).

The study by Li *et al.,* (2013) aimed to identify the cells of origin of Nestin expression and its developmental role within the murine cerebellum. The authors, primarily, determined that Nestin protein (a marker of multipotent stem cells that have committed to the granule neuron lineage) is expressed in the cerebellum of mice at postnatal day 4 (P) in *Nestin-CFP* transgenic mice. The reason behind the usage of cyan fluorescent protein (CFP) was to allow this reporter to fuse to nuclear localized Nestin Protein to discriminate the Nestin positive cell bodies from glial fibres which also express Nestin (Shi *et al.,* 2008; Vukojevic *et al.,* 2010). Surprisingly, these CFP positive cells (*Nestin*-expressing progenitors, NEPs) did not express *Math1*, a well-recognised marker for GNPCs (Ben-Arie *et al.*, 1997).

For further purification and cell quantification in the cerebellar EGL, *Nestin-CFP* mice were crossed with *Math1-GFP* mice in order to create *Nestin-CFP-Math1-GFP* animals. This allowed NEPs to be distinguished from GNPCs within the cerebellum of single animals so that GFP could be used to exclusively label GNPCs for fate mapping. Use of this approach to analyse microdissected cerebellar EGL from P4 *Math1-GFP-Nestin-CFP* mice established that approximately 3-5% of the labelled cells expressed *Nestin*-CFP. Furthermore, no cells expressed both GFP and CFP, as recorded by flow cytometric analysis, suggesting that these cells are distinct populations. Contrary to the prevailing assumption, that GNPCs give arise to granule neurons exclusively, Li *et al.* (2013) demonstrated that NEPs also have the ability to generate granule neurons during cerebellar development.

In this study, the lineage relationship between NEPs and GNPCs was then analysed by crossing *Math1-CreR26R-GFP* mice with *Nestin*-CFP mice to generate a Math1-CreR26R-GFP-Nestin-CFP model. The flow cytometry results indicated that

about 58% of cells in the EGL were GFP positive, while only ~4% were CFP positive, and none of the cells expressed both at the same time. On the other hand, to establish if GNPCs may develop from NEPs, Nestin-Cre^{T2}-R26R-GFP mice were developed and then treated with tamoxifen. The immunostaining of cerebella from P8 mice indicated that Math1^{+}GNPCs were preferentially stationed in the outer tissue of cerebellar EGL. This confirmed that none of GNPCs or NEP give rise to, or were derived from, each other. Furthermore, NEP cells showed a capacity to respond to recombinant Shh protein , as both GNPC and NEP cell populations were significantly increased in the presence of Shh protein (P=0.00077 for NEP), while they became quiescent in the absence of Shh *in vitro*.

More evidence for the distinctive behaviour of NEPs from either GNPCs or neuronal stem cells (NSCs) came from microarray analyses of RNAs from each cell population purified from the cerebellum of P4 *Nestin-CFP-Math1-GFP* mice. The two populations differed in gene expression patterns, as established by principal component analysis. Approximately, 10.87% genes (4902 out of 45,101 probes) were significantly differentially expressed (increased or decreased by ≥2-fold, false discovery rate <0.01 by paired *t* test). Interestingly, 179 genes involved in DNA damage and repair, 62 (34.64%) were differentially expressed between NEPs and GNPCs (P<0.001). These 62 genes were downregulated in NEPs compared to GNPCs. Among these, *Chek1, Lig3* and *Parp1* expression was validated utilising qPCR.

These two cell population were also found to differ in terms of genomic instability, in particular chromosomal breaks, and centromere separation. *Ptch1* was deleted in the experimental mice to continuously stimulate the Shh pathway, by crossing *Nestin-CFP-Math1-GFP* and *Ptch1$^{C/C}$* to generate *Nestin-CFP-Math1-GFP-Ptch1$^{C/C}$* mice. Upon purification of NEPs and GNPCs from P4 *Nestin-CFP-Math1-GFP-Ptch1$^{C/C}$* cerebellar mice, metaphase spread chromosome analysis using Giemsa-banding established that NEPs have significantly higher levels of genomic instability than GNPCs, ~62% of cells compared to ~20.0%. This led the authors to postulate that NEPs might drive tumourigenicity more aggressively that GNPCs, based on genomic instability due to *Ptch1* deletion. Based on this, *Ptch1* was deleted from NEPs and GNPCs, derived from *Nestin-CreER^{T2}-Ptch1$^{C/C}$* and *Math1-CreER^{T2}-Ptch1$^{C/C}$* mice respectively. Animals of both genotypes developed tumours resembling those generated in human medulloblastoma based on histopathological appearance. In regards to molecular comparisons, cells (NEPs or GNPCs) isolated from tumours within these experimental animals showed very similar gene expression patterns compared to human

MB tumours, as assessed by principal component analysis. The tumour penetrance recorded approached 100% in both models.

Interestingly, as few as, 5000 transplanted NEPs produced medulloblastoma in 50% of Nude mice recipients compared to 0% within the same number of GNPCs recipients. However, 20000 cells of each NEPs and GNPCs gave rise to medulloblastoma in 100% and 60% of transplanted recipients, respectively. With the number of transplanted NEPs and GNPCs increased to ~50000, medulloblastoma generation was 100% within recipient's animals. This suggests that NEPs have enhanced tumourigenic potential compared to GNPCs. This study raises the possibility that Shh subgroup tumours could arise in *Nestin* expressing cells and could be distinct in terms of disease course. It is interesting in this regard that analysis of Nestin expression within the murine Illumina beads microarray data generated here established that Nestin is only expressed in tumours generated within the experimental SB model and the *Ptch1* predisposition model, and not within $P53^{+/-}$ and wild type cerebella controls (data not shown). Further analysis of Nestin expression within Shh subgroup tumours may therefore be of interest. The use of SB to target Nestin expressing cells via the generation of $Ptch1^{+/-}$/*Nestin*-SB/T2Onc murine models could also identify candidate CISs involved in *Nestin* derived MB.

Future Work

The results presented here define a novel neuronal transcription factor network involved in MB development, and highlight genes of particular interest. My preliminary experiments where *Tgif2* was overexpressed through retroviral transduction require further work to establish why no protein could be detected despite successful expression at the RNA level. Although the ORF clone was purchased as "sequence verified" it will be necessary to first confirm the existence of an intact open reading frame. It is also possible that the FLAG tag may not be accessible in the protein produced, so an N terminal flag may need to be tried. Alternatively, currently available antibodies raised against human *Igf2* should cross react with the mouse protein, so it may be possible to avoid the use of tagged ORFs should it be necessary.

The *Myt1l* work is very encouraging as the preliminary data suggests that knockdown is having the predicted effect in terms of the expression of the neuronal marker *Math1*, and that this gene affects SHH signalling as assessed by the expression of the effector *Gli1*. However, the negative result obtained with the Western indicated that the protein detection of this gene also requires more troubleshooting. Expression in

GNPCs is low, and sensitivity may be an issue. In addition, the antibody utilised here is a monoclonal version. There are 3 different isoforms of the *Myt1l* protein, and this antibody will not cross react with all of the isoforms. A polyclonal antibody may prove more robust.

Once these issues with protein detection are resolved, the transduced cGNPCs can be analysed to further analyse these genes in particular, and the network in general. First, microarray expression analyses or RNAseq could be used to confirm the activity of the network, and the role of these genes within it. In addition, RNAseq data is currently being generated from several hundred primary tumours in Newcastle, and re-generation of the network in this data will be useful as the accuracy of ARACNe network is affected by sample size and dynamic range of expression. Finally, the impact of gene knockdown/overexpression on cellular phenotype in terms of proliferation, differentiation and invasiveness etc. can all be analysed, and it is also possible to assess the cancer forming potential of transduced cGNPCs in nude mice.

The link between insulin signalling and MB has been highlighted both by the *Igf2* data presented here, but also by ontology analyses in the (Wu *et al.*, 2012) secreen. Synergy between the Shh pathway and *Igf2* upon MB development has been shown as the co-supplementation of Shh protein and Ifg family genes significantly induced BrDu uptake by GNPCs more than separate application of these substances. Furthermore, blocking the activity of *Igfr1* (the *Igf2* receptor) drastically reduced the proliferation of GNPCs, even in the presence of Shh alone (Fernandez *et al.*, 2010). The analysis applied here suggests that the mutations disrupting the network should inhibit neuronal differentiation and lead to proliferation induction through *Igf2* stimulation. Consistent with this, several CISs defining the neuronal network established here have been shown to affect *Ifg2* expression such as *TEAD1*, the co-activator of *YAP* (Fernandez *et al.*, 2012). The same thing is possible for NFI family elements (*NFIA*, *NFIB*) as these have been shown to modulate the expression of *IGFBR5* in a human osteoblastoma cell line (Perez-Casellas *et al.*, 2009). There is also evidence that Crebbp mutation may activate *Ifg2* via induction of *EGR1* (Svaren *et al.*, 2000).

The identification of these genes which may modulate *Igf2* gene expression is of particular interest. Inhibitors of SHH signalling are being used therapeutically to treat SHH subgroup MB, but acquired resistance has been documented (Yauch *et al.*, 2009), making the identification of further agents to treat SHH subgroup tumours desireable. Targetting of *Igf2*, perhaps in conjunction with SHH inhibitors, is therefore a plausible avenue for further therapeutic development.

Finally, two studies have recently successfully generated highly aggressive MB models using MYC overexpression. Human Group 3 MBs frequently overexpress MYC, tend to present a very poor prognosis, and are aggressive and invasive tumours with very large cell and anaplastic histological features (Cho *et al.*, 2011;Northcott *et al.*, 2011). MYC is usually not expressed in GNPCs (Zindy *et al.*, 2006), suggesting that this is unlikely to be the cell of origin. *MYC* is overexpressed in human MB subgroups 3 and 4, with the highest levels of MYC expression detected within Group 3 (Swartling *et al.*, 2010;Northcott *et al.*, 2011). To model these *MYC*-amplified tumours t Myc was -overexpressed in *Cdkn2c*$^{-/-}$*; Trp53*$^{-/-}$ or *Trp53*-/- mutant mice (Kawauchi *et al.*, 2012;Pei *et al.*, 2012). In one study cerebellar stem cells (CD133+) with a stabilised form of *Myc (Myc*^{T5BA}*)* and a highly proliferative activity in culture were transplanted into immunedeficient mice, and small foci of these cells was detected within 2 weeks. However, when these *Myc* expressing cells were infected with a dominant-negative *p53 (DNp53*) vector and transplanted into the cerebellum of mice, highly aggressive tumour accumulations were seen within 6-12 weeks compared to *Myc* only mice. The immunohistochemical assay revealed a larger tumour cells and more necrotic features within *Myc* tumours compared to those raised from *Ptch1*$^{+/-}$ mice. The availability of new models for highly aggressive forms of the disease open up new avenues for research, The results here suggest that utilising SB mutagenesis with these MB models could identify new biomarkers and therapeutic targets in this aggressive disease subset.

References:

ALDER, J., CHO, N. K.andHATTEN, M. E. (1996). Embryonic precursor cells from the rhombic lip are specified to a cerebellar granule neuron identity. *Neuron,* 17**,** 389-99.

AMBASUDHAN, R., TALANTOVA, M., COLEMAN, R., YUAN, X., ZHU, S., LIPTON, S. A.andDING, S. (2011). Direct reprogramming of adult human fibroblasts to functional neurons under defined conditions. *Cell Stem Cell,* 9**,** 113-8.

ANBANANDAM, A., ALBARADO, D. C., NGUYEN, C. T., HALDER, G., GAO, X.andVEERARAGHAVAN, S. (2006). Insights into transcription enhancer factor 1 (TEF-1) activity from the solution structure of the TEA domain. *Proc Natl Acad Sci U S A,* 103**,** 17225-30.

BAO, S., WU, Q., MCLENDON, R. E., HAO, Y., SHI, Q., HJELMELAND, A. B., DEWHIRST, M. W., BIGNER, D. D.andRICH, J. N. (2006). Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature,* 444**,** 756-60.

BARRESI, M. J., HUTSON, L. D., CHIEN, C. B.andKARLSTROM, R. O. (2005). Hedgehog regulated Slit expression determines commissure and glial cell position in the zebrafish forebrain. *Development,* 132**,** 3643-56.

BATRA, S. K., MCLENDON, R. E., KOO, J. S., CASTELINO-PRABHU, S., FUCHS, H. E., KRISCHER, J. P., FRIEDMAN, H. S., BIGNER, D. D.andBIGNER, S. H. (1995). Prognostic implications of chromosome 17p deletions in human medulloblastomas. *J Neurooncol,* 24**,** 39-45.

BENDER, A. M., COLLIER, L. S., RODRIGUEZ, F. J., TIEU, C., LARSON, J. D., HALDER, C., MAHLUM, E., KOLLMEYER, T. M., AKAGI, K., SARKAR, G., LARGAESPADA, D. A.andJENKINS, R. B. (2010). Sleeping beauty-mediated somatic mutagenesis implicates CSF1 in the formation of high-grade astrocytomas. *Cancer Res,* 70**,** 3557-65.

BERMAN, D. M., KARHADKAR, S. S., MAITRA, A., MONTES DE OCA, R., GERSTENBLITH, M. R., BRIGGS, K., PARKER, A. R., SHIMADA, Y., ESHLEMAN, J. R., WATKINS, D. N.andBEACHY, P. A. (2003). Widespread requirement for Hedgehog ligand stimulation in growth of digestive tract tumours. *Nature,* 425**,** 846-51.

BERQUAM-VRIEZE, K. E., NANNAPANENI, K., BRETT, B. T., HOLMFELDT, L., MA, J., ZAGORODNA, O., JENKINS, N. A., COPELAND, N. G., MEYERHOLZ, D. K., KNUDSON, C. M., MULLIGHAN, C. G., SCHEETZ, T. E.andDUPUY, A. J. (2011). Cell of origin strongly influences genetic selection in a mouse model of T-ALL. *Blood,* 118**,** 4646-56.

BERTOLINO, E., WILDT, S., RICHARDS, G.andCLERC, R. G. (1996). Expression of a novel murine homeobox gene in the developing cerebellar external granular layer during its proliferation. *Dev Dyn,* 205**,** 410-20.

BOYLE, E. I., WENG, S., GOLLUB, J., JIN, H., BOTSTEIN, D., CHERRY, J. M.andSHERLOCK, G. (2004). GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics,* 20**,** 3710-5.

BRETT, B. T., BERQUAM-VRIEZE, K. E., NANNAPANENI, K., HUANG, J., SCHEETZ, T. E.andDUPUY, A. J. (2011). Novel molecular and computational methods improve the accuracy of insertion site analysis in Sleeping Beauty-induced tumors. *PLoS One,* 6**,** e24668.

CALABRESE, C., POPPLETON, H., KOCAK, M., HOGG, T. L., FULLER, C., HAMNER, B., OH, E. Y., GABER, M. W., FINKLESTEIN, D., ALLEN, M., FRANK, A., BAYAZITOV, I. T., ZAKHARENKO, S. S., GAJJAR, A., DAVIDOFF, A.andGILBERTSON, R. J. (2007). A perivascular niche for brain tumor stem cells. *Cancer Cell,* 11**,** 69-82.

CALLAHAN, R.andSMITH, G. H. (2000). MMTV-induced mammary tumorigenesis: gene discovery, progression to malignancy and cellular pathways. *Oncogene,* 19**,** 992-1001.

CANCER, R. U. (a). *All cancer combined key facts* [Online]. Available: http://www.cancerresearchuk.org/cancer-info/cancerstats/keyfacts/Allcancerscombined/ [Accessed 25/11/2013.

CANCER, R. U. (b). *Cancer incidence for common cancer* [Online]. Available: http://www.cancerresearchuk.org/cancer-info/cancerstats/incidence/commoncancers/ - Twenty [Accessed 25/11/2013.

CANCER, R. U. (C). *Childhood cancer mortality statistic* [Online]. Available: http://www.cancerresearchuk.org/cancer-info/cancerstats/childhoodcancer/mortality/ [Accessed 25/10/2013.

CAO, X., PFAFF, S. L.andGAGE, F. H. (2008). YAP regulates neural progenitor cell number via the TEA domain transcription factor. *Genes Dev,* 22**,** 3320-34.

CARRO, M. S., LIM, W. K., ALVAREZ, M. J., BOLLO, R. J., ZHAO, X., SNYDER, E. Y., SULMAN, E. P., ANNE, S. L., DOETSCH, F., COLMAN, H., LASORELLA, A., ALDAPE, K., CALIFANO, A.andIAVARONE, A. (2010). The transcriptional network for mesenchymal transformation of brain tumours. *Nature,* 463**,** 318-25.

CHO, Y. J., TSHERNIAK, A., TAMAYO, P., SANTAGATA, S., LIGON, A., GREULICH, H., BERHOUKIM, R., AMANI, V., GOUMNEROVA, L., EBERHART, C. G., LAU, C. C., OLSON, J. M., GILBERTSON, R. J., GAJJAR, A., DELATTRE, O., KOOL, M., LIGON, K., MEYERSON, M., MESIROV, J. P.andPOMEROY, S. L. (2011). Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. *J Clin Oncol,* 29**,** 1424-30.

CHUNG, S. H., KIM, C. T.andHAWKES, R. (2008a). Compartmentation of GABA B receptor2 expression in the mouse cerebellar cortex. *Cerebellum,* 7**,** 295-303.

CHUNG, W. J., OKAMURA, K., MARTIN, R.andLAI, E. C. (2008b). Endogenous RNA interference provides a somatic Defense against Drosophila transposons. *Current Biology,* 18**,** 795-802.

CLARK, K. J., GEURTS, A. M., BELL, J. B.andHACKETT, P. B. (2004). Transposon vectors for gene-trap insertional mutagenesis in vertebrates. *Genesis,* 39**,** 225-33.

COLLIER, L. S., ADAMS, D. J., HACKETT, C. S., BENDZICK, L. E., AKAGI, K., DAVIES, M. N., DIERS, M. D., RODRIGUEZ, F. J., BENDER, A. M., TIEU, C., MATISE, I., DUPUY, A. J., COPELAND, N. G., JENKINS, N. A., HODGSON, J. G., WEISS, W. A., JENKINS, R. B.andLARGAESPADA, D. A. (2009). Whole-body sleeping beauty mutagenesis can cause penetrant leukemia/lymphoma and rare high-grade glioma without associated embryonic lethality. *Cancer Res,* 69**,** 8429-37.

COLLIER, L. S., CARLSON, C. M., RAVIMOHAN, S., DUPUY, A. J.andLARGAESPADA, D. A. (2005). Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature,* 436**,** 272-6.

COLLIER, L. S.andLARGAESPADA, D. A. (2005). Hopping around the tumor genome: transposons for cancer gene discovery. *Cancer Res,* 65**,** 9607-10.

COLLIER, L. S.andLARGAESPADA, D. A. (2006). Transforming science: cancer gene identification. *Curr Opin Genet Dev,* 16**,** 23-9.
COLLIER, L. S.andLARGAESPADA, D. A. (2007a). Transposable elements and the dynamic somatic genome. *Genome Biol,* 8 Suppl 1**,** S5.
COLLIER, L. S.andLARGAESPADA, D. A. (2007b). Transposons for cancer gene discovery: Sleeping Beauty and beyond. *Genome Biol,* 8 Suppl 1**,** S15.
COPELAND, N. G.andJENKINS, N. A. (2010). Harnessing transposons for cancer gene discovery. *Nat Rev Cancer,* 10**,** 696-706.
CORCORAN, R. B., BACHAR RAVEH, T., BARAKAT, M. T., LEE, E. Y.andSCOTT, M. P. (2008). Insulin-like growth factor 2 is required for progression to advanced medulloblastoma in patched1 heterozygous mice. *Cancer Res,* 68**,** 8788-95.
CORDERO, F., BOTTA, M.andCALOGERO, R. A. (2007). Microarray data analysis and mining approaches. *Brief Funct Genomic Proteomic,* 6**,** 265-81.
DAHMANE, N.andRUIZ I ALTABA, A. (1999). Sonic hedgehog regulates the growth and patterning of the cerebellum. *Development,* 126**,** 3089-100.
DE BONT, J. M., PACKER, R. J., MICHIELS, E. M., DEN BOER, M. L.andPIETERS, R. (2008). Biological background of pediatric medulloblastoma and ependymoma: a review from a translational research perspective. *Neuro Oncol,* 10**,** 1040-60.
DE LA CRUZ, F.andDAVIES, J. (2000). Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol,* 8**,** 128-33.
DE RIDDER, J., UREN, A., KOOL, J., REINDERS, M.andWESSELS, L. (2006). Detecting statistically significant common insertion sites in retroviral insertional mutagenesis screens. *PLoS Comput Biol,* 2**,** e166.
DENEEN, B., HO, R., LUKASZEWICZ, A., HOCHSTIM, C. J., GRONOSTAJSKI, R. M.andANDERSON, D. J. (2006). The transcription factor NFIA controls the onset of gliogenesis in the developing spinal cord. *Neuron,* 52**,** 953-68.
DHALL, G. (2009). Medulloblastoma. *J Child Neurol,* 24**,** 1418-30.
DICKINSON, R. E., DALLOL, A., BIECHE, I., KREX, D., MORTON, D., MAHER, E. R.andLATIF, F. (2004). Epigenetic inactivation of SLIT3 and SLIT1 genes in human cancers. *Br J Cancer,* 91**,** 2071-8.
DUPUY, A. J. (2010). Transposon-based screens for cancer gene discovery in mouse models. *Semin Cancer Biol,* 20**,** 261-8.
DUPUY, A. J., AKAGI, K., LARGAESPADA, D. A., COPELAND, N. G.andJENKINS, N. A. (2005). Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature,* 436**,** 221-6.
DUPUY, A. J., FRITZ, S.andLARGAESPADA, D. A. (2001). Transposition and gene disruption in the male germline of the mouse. *Genesis,* 30**,** 82-8.
DUPUY, A. J., ROGERS, L. M., KIM, J., NANNAPANENI, K., STARR, T. K., LIU, P., LARGAESPADA, D. A., SCHEETZ, T. E., JENKINS, N. A.andCOPELAND, N. G. (2009). A modified sleeping beauty transposon system that can be used to model a wide variety of human cancers in mice. *Cancer Res,* 69**,** 8150-6.
ELKON, R., LINHART, C., SHARAN, R., SHAMIR, R.andSHILOH, Y. (2003). Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Res,* 13**,** 773-80.
ELLISON, D. W. (2010). Childhood medulloblastoma: novel approaches to the classification of a heterogeneous disease. *Acta Neuropathol,* 120**,** 305-16.
ELLISON, D. W., CLIFFORD, S. C., GAJJAR, A.andGILBERTSON, R. J. (2003). What's new in neuro-oncology? Recent advances in medulloblastoma. *Eur J Paediatr Neurol,* 7**,** 53-66.

ELLISON, D. W., KOCAK, M., DALTON, J., MEGAHED, H., LUSHER, M. E., RYAN, S. L., ZHAO, W., NICHOLSON, S. L., TAYLOR, R. E., BAILEY, S.andCLIFFORD, S. C. (2011). Definition of disease-risk stratification groups in childhood medulloblastoma using combined clinical, pathologic, and molecular variables. *J Clin Oncol,* 29**,** 1400-7.

EREZ, A., ILAN, T., AMARIGLIO, N., MULER, I., BROK-SIMONI, F., RECHAVI, G.andIZRAELI, S. (2002). GLI3 is not mutated commonly in sporadic medulloblastomas. *Cancer,* 95**,** 28-31.

ESSNER, J. J., MCIVOR, R. S.andHACKETT, P. B. (2005). Awakening gene therapy with Sleeping Beauty transposons. *Curr Opin Pharmacol,* 5**,** 513-9.

FAN, X., MATSUI, W., KHAKI, L., STEARNS, D., CHUN, J., LI, Y. M.andEBERHART, C. G. (2006). Notch pathway inhibition depletes stem-like cells and blocks engraftment in embryonal brain tumors. *Cancer Res,* 66**,** 7445-52.

FAN, X., MIKOLAENKO, I., ELHASSAN, I., NI, X., WANG, Y., BALL, D., BRAT, D. J., PERRY, A.andEBERHART, C. G. (2004). Notch1 and notch2 have opposite effects on embryonal brain tumor growth. *Cancer Res,* 64**,** 7787-93.

FATEMI, S. H., FOLSOM, T. D.andTHURAS, P. D. (2011). Deficits in GABA(B) receptor system in schizophrenia and mood disorders: a postmortem study. *Schizophr Res,* 128**,** 37-43.

FATTET, S., HABERLER, C., LEGOIX, P., VARLET, P., LELLOUCH-TUBIANA, A., LAIR, S., MANIE, E., RAQUIN, M. A., BOURS, D., CARPENTIER, S., BARILLOT, E., GRILL, J., DOZ, F., PUGET, S., JANOUEIX-LEROSEY, I.andDELATTRE, O. (2009). Beta-catenin status in paediatric medulloblastomas: correlation of immunohistochemical expression with mutational status, genetic profiles, and clinical characteristics. *J Pathol,* 218**,** 86-94.

FERNANDEZ, C., TATARD, V. M., BERTRAND, N.andDAHMANE, N. (2010). Differential modulation of Sonic-hedgehog-induced cerebellar granule cell precursor proliferation by the IGF signaling network. *Dev Neurosci,* 32**,** 59-70.

FERNANDEZ, L. A., NORTHCOTT, P. A., DALTON, J., FRAGA, C., ELLISON, D., ANGERS, S., TAYLOR, M. D.andKENNEY, A. M. (2009). YAP1 is amplified and up-regulated in hedgehog-associated medulloblastomas and mediates Sonic hedgehog-driven neural precursor proliferation. *Genes Dev,* 23**,** 2729-41.

FERNANDEZ, L. A., SQUATRITO, M., NORTHCOTT, P., AWAN, A., HOLLAND, E. C., TAYLOR, M. D., NAHLE, Z.andKENNEY, A. M. (2012). Oncogenic YAP promotes radioresistance and genomic instability in medulloblastoma through IGF2-mediated Akt activation. *Oncogene,* 31**,** 1923-37.

FESCHOTTE, C., JIANG, N.andWESSLER, S. R. (2002). Plant transposable elements: Where genetics meets genomics. *Nature Reviews Genetics,* 3**,** 329-341.

FINK, A. J., ENGLUND, C., DAZA, R. A., PHAM, D., LAU, C., NIVISON, M., KOWALCZYK, T.andHEVNER, R. F. (2006). Development of the deep cerebellar nuclei: transcription factors and cell migration from the rhombic lip. *J Neurosci,* 26**,** 3066-76.

FOGARTY, M., GRIST, M., GELMAN, D., MARIN, O., PACHNIS, V.andKESSARIS, N. (2007). Spatial genetic patterning of the embryonic neuroepithelium generates GABAergic interneuron diversity in the adult cortex. *J Neurosci,* 27**,** 10935-46.

FULTS, D. W. (2005). Modeling medulloblastoma with genetically engineered mice. *Neurosurg Focus,* 19**,** E7.

GAILANI, M. R., BALE, S. J., LEFFELL, D. J., DIGIOVANNA, J. J., PECK, G. L., POLIAK, S., DRUM, M. A., PASTAKIA, B., MCBRIDE, O. W., KASE, R.andET AL. (1992).

Developmental defects in Gorlin syndrome related to a putative tumor suppressor gene on chromosome 9. *Cell,* 69**,** 111-7.

GAJJAR, A., CHINTAGUMPALA, M., ASHLEY, D., KELLIE, S., KUN, L. E., MERCHANT, T. E., WOO, S., WHEELER, G., AHERN, V., KRASIN, M. J., FOULADI, M., BRONISCER, A., KRANCE, R., HALE, G. A., STEWART, C. F., DAUSER, R., SANFORD, R. A., FULLER, C., LAU, C., BOYETT, J. M., WALLACE, D.andGILBERTSON, R. J. (2006). Risk-adapted craniospinal radiotherapy followed by high-dose chemotherapy and stem-cell rescue in children with newly diagnosed medulloblastoma (St Jude Medulloblastoma-96): long-term results from a prospective, multicentre trial. *Lancet Oncol,* 7**,** 813-20.

GEURTS, A. M., YANG, Y., CLARK, K. J., LIU, G., CUI, Z., DUPUY, A. J., BELL, J. B., LARGAESPADA, D. A.andHACKETT, P. B. (2003). Gene transfer into genomes of human cells by the sleeping beauty transposon system. *Mol Ther,* 8**,** 108-17.

GIBSON, P., TONG, Y., ROBINSON, G., THOMPSON, M. C., CURRLE, D. S., EDEN, C., KRANENBURG, T. A., HOGG, T., POPPLETON, H., MARTIN, J., FINKELSTEIN, D., POUNDS, S., WEISS, A., PATAY, Z., SCOGGINS, M., OGG, R., PEI, Y., YANG, Z. J., BRUN, S., LEE, Y., ZINDY, F., LINDSEY, J. C., TAKETO, M. M., BOOP, F. A., SANFORD, R. A., GAJJAR, A., CLIFFORD, S. C., ROUSSEL, M. F., MCKINNON, P. J., GUTMANN, D. H., ELLISON, D. W., WECHSLER-REYA, R.andGILBERTSON, R. J. (2010). Subtypes of medulloblastoma have distinct developmental origins. *Nature,* 468**,** 1095-9.

GILBERTSON, R. J.andELLISON, D. W. (2008). The origins of medulloblastoma subtypes. *Annu Rev Pathol,* 3**,** 341-65.

GONG, S., ZHENG, C., DOUGHTY, M. L., LOSOS, K., DIDKOVSKY, N., SCHAMBRA, U. B., NOWAK, N. J., JOYNER, A., LEBLANC, G., HATTEN, M. E.andHEINTZ, N. (2003). A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature,* 425**,** 917-25.

GOODRICH, L. V., MILENKOVIC, L., HIGGINS, K. M.andSCOTT, M. P. (1997). Altered neural cell fates and medulloblastoma in mouse patched mutants. *Science,* 277**,** 1109-13.

GRABHER, C.andWITTBRODT, J. (2007). Meganuclease and transposon mediated transgenesis in medaka. *Genome Biol,* 8 Suppl 1**,** S10.

GRIPP, K. W., WOTTON, D., EDWARDS, M. C., ROESSLER, E., ADES, L., MEINECKE, P., RICHIERI-COSTA, A., ZACKAI, E. H., MASSAGUE, J., MUENKE, M.andELLEDGE, S. J. (2000). Mutations in TGIF cause holoprosencephaly and link NODAL signalling to human neural axis determination. *Nat Genet,* 25**,** 205-8.

GUICHARD, C., AMADDEO, G., IMBEAUD, S., LADEIRO, Y., PELLETIER, L., MAAD, I. B., CALDERARO, J., BIOULAC-SAGE, P., LETEXIER, M., DEGOS, F., CLEMENT, B., BALABAUD, C., CHEVET, E., LAURENT, A., COUCHY, G., LETOUZE, E., CALVO, F.andZUCMAN-ROSSI, J. (2012). Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat Genet*.

GUO, L., SHI, Q., DIAL, S., XIA, Q., MEI, N., LI, Q. Z., CHAN, P. C.andFU, P. (2010). Gene expression profiling in male B6C3F1 mouse livers exposed to kava identifies--changes in drug metabolizing genes and potential mechanisms linked to kava toxicity. *Food Chem Toxicol,* 48**,** 686-96.

HABER, D.andHARLOW, E. (1997). Tumour-suppressor genes: evolving definitions in the genomic age. *Nat Genet,* 16**,** 320-2.

HADJIPANAYIS, C. G.andVAN MEIR, E. G. (2009). Brain cancer propagating cells: biology, genetics and targeted therapies. *Trends Mol Med,* 15**,** 519-30.

HAHN, H., WOJNOWSKI, L., SPECHT, K., KAPPLER, R., CALZADA-WACK, J., POTTER, D., ZIMMER, A., MULLER, U., SAMSON, E., QUINTANILLA-MARTINEZ, L.andZIMMER, A. (2000). Patched target Igf2 is indispensable for the formation of medulloblastoma and rhabdomyosarcoma. *J Biol Chem,* 275**,** 28341-4.

HALLAHAN, A. R., PRITCHARD, J. I., HANSEN, S., BENSON, M., STOECK, J., HATTON, B. A., RUSSELL, T. L., ELLENBOGEN, R. G., BERNSTEIN, I. D., BEACHY, P. A.andOLSON, J. M. (2004). The SmoA1 mouse model reveals that notch signaling is critical for the growth and survival of sonic hedgehog-induced medulloblastomas. *Cancer Res,* 64**,** 7794-800.

HAMILTON, S. R., LIU, B., PARSONS, R. E., PAPADOPOULOS, N., JEN, J., POWELL, S. M., KRUSH, A. J., BERK, T., COHEN, Z., TETU, B.andET AL. (1995). The molecular basis of Turcot's syndrome. *N Engl J Med,* 332**,** 839-47.

HANAHAN, D.andWEINBERG, R. A. (2011). Hallmarks of cancer: the next generation. *Cell,* 144**,** 646-74.

HARTMANN, W., DIGON-SONTGERATH, B., KOCH, A., WAHA, A., ENDL, E., DANI, I., DENKHAUS, D., GOODYER, C. G., SORENSEN, N., WIESTLER, O. D.andPIETSCH, T. (2006). Phosphatidylinositol 3'-kinase/AKT signaling is activated in medulloblastoma cell proliferation and is associated with reduced expression of PTEN. *Clin Cancer Res,* 12**,** 3019-27.

HATTEN, M. E.andROUSSEL, M. F. (2011). Development and cancer of the cerebellum. *Trends Neurosci,* 34**,** 134-42.

HATTON, B. A., VILLAVICENCIO, E. H., PRITCHARD, J., LEBLANC, M., HANSEN, S., ULRICH, M., DITZLER, S., PULLAR, B., STROUD, M. R.andOLSON, J. M. (2010). Notch signaling is not essential in sonic hedgehog-activated medulloblastoma. *Oncogene,* 29**,** 3865-72.

HATTON, B. A., VILLAVICENCIO, E. H., TSUCHIYA, K. D., PRITCHARD, J. I., DITZLER, S., PULLAR, B., HANSEN, S., KNOBLAUGH, S. E., LEE, D., EBERHART, C. G., HALLAHAN, A. R.andOLSON, J. M. (2008). The Smo/Smo model: hedgehog-induced medulloblastoma with 90% incidence and leptomeningeal spread. *Cancer Res,* 68**,** 1768-76.

HELMS, A. W.andJOHNSON, J. E. (1998). Progenitors of dorsal commissural interneurons are defined by MATH1 expression. *Development,* 125**,** 919-28.

HUI, C. C., SLUSARSKI, D., PLATT, K. A., HOLMGREN, R.andJOYNER, A. L. (1994). Expression of three mouse homologs of the Drosophila segment polarity gene cubitus interruptus, Gli, Gli-2, and Gli-3, in ectoderm- and mesoderm-derived tissues suggests multiple roles during postimplantation development. *Dev Biol,* 162**,** 402-13.

HUSE, J. T.andHOLLAND, E. C. (2010). Targeting brain cancer: advances in the molecular pathology of malignant glioma and medulloblastoma. *Nat Rev Cancer,* 10**,** 319-31.

IMOTO, I., PIMKHAOKHAM, A., WATANABE, T., SAITO-OHARA, F., SOEDA, E.andINAZAWA, J. (2000). Amplification and overexpression of TGIF2, a novel homeobox gene of the TALE superclass, in ovarian cancer cell lines. *Biochem Biophys Res Commun,* 276**,** 264-70.

IRIZARRY, R. A., HOBBS, B., COLLIN, F., BEAZER-BARCLAY, Y. D., ANTONELLIS, K. J., SCHERF, U.andSPEED, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics,* 4**,** 249-64.

IVICS, Z., HACKETT, P. B., PLASTERK, R. H.andIZSVAK, Z. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell,* 91**,** 501-10.

IZSVAK, Z., CHUAH, M. K., VANDENDRIESSCHE, T.andIVICS, Z. (2009). Efficient stable gene transfer into human cells by the Sleeping Beauty transposon vectors. *Methods,* 49**,** 287-97.

IZSVAK, Z., IVICS, Z.andHACKETT, P. B. (1997). Repetitive elements and their genetic applications in zebrafish. *Biochem Cell Biol,* 75**,** 507-23.

JOHNSON, G. V.andJOPE, R. S. (1992). The role of microtubule-associated protein 2 (MAP-2) in neuronal growth, plasticity, and degeneration. *J Neurosci Res,* 33**,** 505-12.

JOHNSON, R. L., ROTHMAN, A. L., XIE, J., GOODRICH, L. V., BARE, J. W., BONIFAS, J. M., QUINN, A. G., MYERS, R. M., COX, D. R., EPSTEIN, E. H., JR.andSCOTT, M. P. (1996). Human homolog of patched, a candidate gene for the basal cell nevus syndrome. *Science,* 272**,** 1668-71.

JONES, D. T., JAGER, N., KOOL, M., ZICHNER, T., HUTTER, B., SULTAN, M., CHO, Y. J., PUGH, T. J., HOVESTADT, V., STUTZ, A. M., RAUSCH, T., WARNATZ, H. J., RYZHOVA, M., BENDER, S., STURM, D., PLEIER, S., CIN, H., PFAFF, E., SIEBER, L., WITTMANN, A., REMKE, M., WITT, H., HUTTER, S., TZARIDIS, T., WEISCHENFELDT, J., RAEDER, B., AVCI, M., AMSTISLAVSKIY, V., ZAPATKA, M., WEBER, U. D., WANG, Q., LASITSCHKA, B., BARTHOLOMAE, C. C., SCHMIDT, M., VON KALLE, C., AST, V., LAWERENZ, C., EILS, J., KABBE, R., BENES, V., VAN SLUIS, P., KOSTER, J., VOLCKMANN, R., SHIH, D., BETTS, M. J., RUSSELL, R. B., COCO, S., TONINI, G. P., SCHULLER, U., HANS, V., GRAF, N., KIM, Y. J., MONORANU, C., ROGGENDORF, W., UNTERBERG, A., HEROLD-MENDE, C., MILDE, T., KULOZIK, A. E., VON DEIMLING, A., WITT, O., MAASS, E., ROSSLER, J., EBINGER, M., SCHUHMANN, M. U., FRUHWALD, M. C., HASSELBLATT, M., JABADO, N., RUTKOWSKI, S., VON BUEREN, A. O., WILLIAMSON, D., CLIFFORD, S. C., MCCABE, M. G., COLLINS, V. P., WOLF, S., WIEMANN, S., LEHRACH, H., BRORS, B., SCHEURLEN, W., FELSBERG, J., REIFENBERGER, G., NORTHCOTT, P. A., TAYLOR, M. D., MEYERSON, M., POMEROY, S. L., YASPO, M. L., KORBEL, J. O., KORSHUNOV, A., EILS, R., PFISTER, S. M.andLICHTER, P. (2012). Dissecting the genomic complexity underlying medulloblastoma. *Nature,* 488**,** 100-5.

KALENDAR, R., GROB, T., REGINA, M., SUONIEMI, A.andSCHULMAN, A. (1999). IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theoretical and Applied Genetics,* 98**,** 704-711.

KANEKO, Y., SAKAKIBARA, S., IMAI, T., SUZUKI, A., NAKAMURA, Y., SAWAMOTO, K., OGAWA, Y., TOYAMA, Y., MIYATA, T.andOKANO, H. (2000). Musashi1: an evolutionally conserved marker for CNS progenitor cells including neural stem cells. *Dev Neurosci,* 22**,** 139-53.

KANG-PARK, S., LEE, Y. I.andLEE, Y. I. (2003). PTEN modulates insulin-like growth factor II (IGF-II)-mediated signaling; the protein phosphatase activity of PTEN downregulates IGF-II expression in hepatoma cells. *FEBS Lett,* 545**,** 203-8.

KARHADKAR, S. S., BOVA, G. S., ABDALLAH, N., DHARA, S., GARDNER, D., MAITRA, A., ISAACS, J. T., BERMAN, D. M.andBEACHY, P. A. (2004). Hedgehog signalling in prostate regeneration, neoplasia and metastasis. *Nature,* 431**,** 707-12.

KAWAUCHI, D., ROBINSON, G., UZIEL, T., GIBSON, P., REHG, J., GAO, C., FINKELSTEIN, D., QU, C., POUNDS, S., ELLISON, D. W., GILBERTSON, R.

J.andROUSSEL, M. F. (2012). A mouse model of the most aggressive subgroup of human medulloblastoma. *Cancer Cell,* 21**,** 168-80.

KAZAZIAN, H. H., JR. (2004). Mobile elements: drivers of genome evolution. *Science,* 303**,** 1626-32.

KEFAS, B., COMEAU, L., FLOYD, D. H., SELEVERSTOV, O., GODLEWSKI, J., SCHMITTGEN, T., JIANG, J., DIPIERRO, C. G., LI, Y., CHIOCCA, E. A., LEE, J., FINE, H., ABOUNADER, R., LAWLER, S.andPUROW, B. (2009). The neuronal microRNA miR-326 acts in a feedback loop with notch and has therapeutic potential against brain tumors. *J Neurosci,* 29**,** 15161-8.

KENG, V. W., VILLANUEVA, A., CHIANG, D. Y., DUPUY, A. J., RYAN, B. J., MATISE, I., SILVERSTEIN, K. A., SARVER, A., STARR, T. K., AKAGI, K., TESSAROLLO, L., COLLIER, L. S., POWERS, S., LOWE, S. W., JENKINS, N. A., COPELAND, N. G., LLOVET, J. M.andLARGAESPADA, D. A. (2009). A conditional transposon-based insertional mutagenesis screen for genes associated with mouse hepatocellular carcinoma. *Nat Biotechnol,* 27**,** 264-74.

KENNEY, A. M.andROWITCH, D. H. (2000). Sonic hedgehog promotes G(1) cyclin expression and sustained cell cycle progression in mammalian neuronal precursors. *Mol Cell Biol,* 20**,** 9055-67.

KHODOSEVICH, K., SEEBURG, P. H.andMONYER, H. (2009). Major signaling pathways in migrating neuroblasts. *Front Mol Neurosci,* 2**,** 7.

KIM, J. G., ARMSTRONG, R. C., V AGOSTON, D., ROBINSKY, A., WIESE, C., NAGLE, J.andHUDSON, L. D. (1997). Myelin transcription factor 1 (Myt1) of the oligodendrocyte lineage, along with a closely related CCHC zinc finger, is expressed in developing neurons in the mammalian central nervous system. *J Neurosci Res,* 50**,** 272-90.

KIM, T. H., LEE, D. K., FRANCO, H. L., LYDON, J. P.andJEONG, J. W. (2010). ERBB receptor feedback inhibitor 1 regulation of estrogen receptor activity is critical for uterine implantation in mice. *Biol Reprod,* 82**,** 706-13.

KLUMPP, S., THISSEN, M. C.andKRIEGLSTEIN, J. (2006). Protein phosphatases types 2Calpha and 2Cbeta in apoptosis. *Biochem Soc Trans,* 34**,** 1370-5.

KNEPPER, J. L., JAMES, A. C.andMING, J. E. (2006). TGIF, a gene associated with human brain defects, regulates neuronal development. *Dev Dyn,* 235**,** 1482-90.

KOOL, J.andBERNS, A. (2009). High-throughput insertional mutagenesis screens in mice to identify oncogenic networks. *Nat Rev Cancer,* 9**,** 389-99.

KOOL, M., KORSHUNOV, A., REMKE, M., JONES, D. T., SCHLANSTEIN, M., NORTHCOTT, P. A., CHO, Y. J., KOSTER, J., SCHOUTEN-VAN MEETEREN, A., VAN VUURDEN, D., CLIFFORD, S. C., PIETSCH, T., VON BUEREN, A. O., RUTKOWSKI, S., MCCABE, M., COLLINS, V. P., BACKLUND, M. L., HABERLER, C., BOURDEAUT, F., DELATTRE, O., DOZ, F., ELLISON, D. W., GILBERTSON, R. J., POMEROY, S. L., TAYLOR, M. D., LICHTER, P.andPFISTER, S. M. (2012). Molecular subgroups of medulloblastoma: an international meta-analysis of transcriptome, genetic aberrations, and clinical data of WNT, SHH, Group 3, and Group 4 medulloblastomas. *Acta Neuropathol,* 123**,** 473-84.

KOOL, M., KOSTER, J., BUNT, J., HASSELT, N. E., LAKEMAN, A., VAN SLUIS, P., TROOST, D., MEETEREN, N. S., CARON, H. N., CLOOS, J., MRSIC, A., YLSTRA, B., GRAJKOWSKA, W., HARTMANN, W., PIETSCH, T., ELLISON, D., CLIFFORD, S. C.andVERSTEEG, R. (2008). Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. *PLoS One,* 3**,** e3088.

KUBO, M., NAKAMURA, M., TASAKI, A., YAMANAKA, N., NAKASHIMA, H., NOMURA, M., KUROKI, S.andKATANO, M. (2004). Hedgehog signaling pathway is a new therapeutic target for patients with breast cancer. *Cancer Res,* 64**,** 6071-4.

LAI, A. Y.andWADE, P. A. (2011). Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat Rev Cancer,* 11**,** 588-96.

LANDSBERG, R. L., AWATRAMANI, R. B., HUNTER, N. L., FARAGO, A. F., DIPIETRANTONIO, H. J., RODRIGUEZ, C. I.andDYMECKI, S. M. (2005). Hindbrain rhombic lip is comprised of discrete progenitor cell populations allocated by Pax6. *Neuron,* 48**,** 933-47.

LARGAESPADA, D. A. (2009). Transposon-mediated mutagenesis of somatic cells in the mouse for cancer gene identification. *Methods,* 49**,** 282-6.

LEAL-ORTIZ, S., WAITES, C. L., TERRY-LORENZO, R., ZAMORANO, P., GUNDELFINGER, E. D.andGARNER, C. C. (2008). Piccolo modulation of Synapsin1a dynamics regulates synaptic vesicle exocytosis. *J Cell Biol,* 181**,** 831-46.

LEE, Y., KAWAGOE, R., SASAI, K., LI, Y., RUSSELL, H. R., CURRAN, T.andMCKINNON, P. J. (2007). Loss of suppressor-of-fused function promotes tumorigenesis. *Oncogene,* 26**,** 6442-7.

LEUNG, C., LINGBEEK, M., SHAKHOVA, O., LIU, J., TANGER, E., SAREMASLANI, P., VAN LOHUIZEN, M.andMARINO, S. (2004). Bmi1 is essential for cerebellar development and is overexpressed in human medulloblastomas. *Nature,* 428**,** 337-41.

LEWIS, M. T., ROSS, S., STRICKLAND, P. A., SUGNET, C. W., JIMENEZ, E., HUI, C.andDANIEL, C. W. (2001). The Gli2 transcription factor is required for normal mouse mammary gland development. *Dev Biol,* 238**,** 133-44.

LI, K. K., LAU, K. M.andNG, H. K. (2013). Signaling pathway and molecular subgroups of medulloblastoma. *Int J Clin Exp Pathol,* 6**,** 1211-22.

LI, X., LOBO, N., BAUSER, C. A.andFRASER, M. J., JR. (2001). The minimum internal and external sequence requirements for transposition of the eukaryotic transformation vector piggyBac. *Mol Genet Genomics,* 266**,** 190-8.

LIN, J. C.andCEPKO, C. L. (1999). Biphasic dispersion of clones containing Purkinje cells and glia in the developing chick cerebellum. *Dev Biol,* 211**,** 177-97.

LIU, J., HE, Y., AMASINO, R.andCHEN, X. (2004). siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in Arabidopsis. *Genes Dev,* 18**,** 2873-8.

MAERE, S., HEYMANS, K.andKUIPER, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics,* 21**,** 3448-9.

MAGNANI, D., HASENPUSCH-THEIL, K., BENADIBA, C., YU, T., BASSON, M. A., PRICE, D. J., LEBRAND, C.andTHEIL, T. (2012). Gli3 Controls Corpus Callosum Formation by Positioning Midline Guideposts During Telencephalic Patterning. *Cereb Cortex*.

MANCUSO, M., PASQUALI, E., LEONARDI, S., TANORI, M., REBESSI, S., DI MAJO, V., PAZZAGLIA, S., TONI, M. P., PIMPINELLA, M., COVELLI, V.andSARAN, A. (2008). Oncogenic bystander radiation effects in Patched heterozygous mouse cerebellum. *Proc Natl Acad Sci U S A,* 105**,** 12445-50.

MARCH, H. N., RUST, A. G., WRIGHT, N. A., TEN HOEVE, J., DE RIDDER, J., ELDRIDGE, M., VAN DER WEYDEN, L., BERNS, A., GADIOT, J., UREN, A., KEMP, R., ARENDS, M. J., WESSELS, L. F., WINTON, D. J.andADAMS, D. J. (2011).

Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis. *Nat Genet,* 43**,** 1202-9.

MARGOLIN, A. A., WANG, K., LIM, W. K., KUSTAGI, M., NEMENMAN, I.andCALIFANO, A. (2006). Reverse engineering cellular networks. *Nat Protoc,* 1**,** 662-71.

MARILLAT, V., CASES, O., NGUYEN-BA-CHARVET, K. T., TESSIER-LAVIGNE, M., SOTELO, C.andCHEDOTAL, A. (2002). Spatiotemporal expression patterns of slit and robo genes in the rat brain. *J Comp Neurol,* 442**,** 130-55.

MARINO, S., VOOIJS, M., VAN DER GULDEN, H., JONKERS, J.andBERNS, A. (2000). Induction of medulloblastomas in p53-null mutant mice by somatic inactivation of Rb in the external granular layer cells of the cerebellum. *Genes Dev,* 14**,** 994-1004.

MARLOW, R., STRICKLAND, P., LEE, J. S., WU, X., PEBENITO, M., BINNEWIES, M., LE, E. K., MORAN, A., MACIAS, H., CARDIFF, R. D., SUKUMAR, S.andHINCK, L. (2008). SLITs suppress tumor growth in vivo by silencing Sdf1/Cxcr4 within breast epithelium. *Cancer Res,* 68**,** 7819-27.

MARTIN, J. H. 2003. *Neuroanatomy : text and atlas,* New York, N.Y. ; London, McGraw-Hill.

MASSIMINO, M., GIANGASPERO, F., GARRE, M. L., GANDOLA, L., POGGI, G., BIASSONI, V., GATTA, G.andRUTKOWSKI, S. (2011). Childhood medulloblastoma. *Crit Rev Oncol Hematol,* 79**,** 65-83.

MATTISON, J., VAN DER WEYDEN, L., HUBBARD, T.andADAMS, D. J. (2009). Cancer gene discovery in mouse and man. *Biochim Biophys Acta,* 1796**,** 140-61.

MAZLOOM, A., ZANGENEH, A. H.andPAULINO, A. C. (2010). Prognostic factors after extraneural metastasis of medulloblastoma. *Int J Radiat Oncol Biol Phys,* 78**,** 72-8.

MCINTYRE, R. E., VAN DER WEYDEN, L.andADAMS, D. J. (2012). Cancer gene discovery in the mouse. *Curr Opin Genet Dev,* 22**,** 14-20.

MELHUISH, T. A., GALLO, C. M.andWOTTON, D. (2001). TGIF2 interacts with histone deacetylase 1 and represses transcription. *J Biol Chem,* 276**,** 32109-14.

MORALES, D.andHATTEN, M. E. (2006). Molecular markers of neuronal progenitors in the embryonic cerebellar anlage. *J Neurosci,* 26**,** 12226-36.

NAKAHARA, Y., NORTHCOTT, P. A., LI, M., KONGKHAM, P. N., SMITH, C., YAN, H., CROUL, S., RA, Y. S., EBERHART, C., HUANG, A., BIGNER, D., GRAJKOWSKA, W., VAN METER, T., RUTKA, J. T.andTAYLOR, M. D. (2010). Genetic and epigenetic inactivation of Kruppel-like factor 4 in medulloblastoma. *Neoplasia,* 12**,** 20-7.

NAKANO, I., MASTERMAN-SMITH, M., SAIGUSA, K., PAUCAR, A. A., HORVATH, S., SHOEMAKER, L., WATANABE, M., NEGRO, A., BAJPAI, R., HOWES, A., LELIEVRE, V., WASCHEK, J. A., LAZAREFF, J. A., FREIJE, W. A., LIAU, L. M., GILBERTSON, R. J., CLOUGHESY, T. F., GESCHWIND, D. H., NELSON, S. F., MISCHEL, P. S., TERSKIKH, A. V.andKORNBLUM, H. I. (2008). Maternal embryonic leucine zipper kinase is a key regulator of the proliferation of malignant brain tumors, including brain tumor stem cells. *J Neurosci Res,* 86**,** 48-60.

NAVAJAS, A.andGIRALT, J. (2010). Evidence in medulloblastomas. *Clin Transl Oncol,* 12**,** 271-7.

NISHIMOTO, S.andNISHIDA, E. (2007). Fibroblast growth factor 13 is essential for neural differentiation in Xenopus early embryonic development. *J Biol Chem,* 282**,** 24255-61.

NORTHCOTT, P. A., FERNANDEZ, L. A., HAGAN, J. P., ELLISON, D. W., GRAJKOWSKA, W., GILLESPIE, Y., GRUNDY, R., VAN METER, T., RUTKA, J. T., CROCE, C. M., KENNEY, A. M.andTAYLOR, M. D. (2009). The miR-17/92 polycistron is up-regulated in sonic hedgehog-driven medulloblastomas and induced by N-myc in sonic hedgehog-treated cerebellar neural precursors. *Cancer Res,* 69**,** 3249-55.

NORTHCOTT, P. A., KORSHUNOV, A., WITT, H., HIELSCHER, T., EBERHART, C. G., MACK, S., BOUFFET, E., CLIFFORD, S. C., HAWKINS, C. E., FRENCH, P., RUTKA, J. T., PFISTER, S.andTAYLOR, M. D. (2011). Medulloblastoma comprises four distinct molecular variants. *J Clin Oncol,* 29**,** 1408-14.

NORTHCOTT, P. A., SHIH, D. J., PEACOCK, J., GARZIA, L., MORRISSY, A. S., ZICHNER, T., STUTZ, A. M., KORSHUNOV, A., REIMAND, J., SCHUMACHER, S. E., BEROUKHIM, R., ELLISON, D. W., MARSHALL, C. R., LIONEL, A. C., MACK, S., DUBUC, A., YAO, Y., RAMASWAMY, V., LUU, B., ROLIDER, A., CAVALLI, F. M., WANG, X., REMKE, M., WU, X., CHIU, R. Y., CHU, A., CHUAH, E., CORBETT, R. D., HOAD, G. R., JACKMAN, S. D., LI, Y., LO, A., MUNGALL, K. L., NIP, K. M., QIAN, J. Q., RAYMOND, A. G., THIESSEN, N. T., VARHOL, R. J., BIROL, I., MOORE, R. A., MUNGALL, A. J., HOLT, R., KAWAUCHI, D., ROUSSEL, M. F., KOOL, M., JONES, D. T., WITT, H., FERNANDEZ, L. A., KENNEY, A. M., WECHSLER-REYA, R. J., DIRKS, P., AVIV, T., GRAJKOWSKA, W. A., PEREK-POLNIK, M., HABERLER, C. C., DELATTRE, O., REYNAUD, S. S., DOZ, F. F., PERNET-FATTET, S. S., CHO, B. K., KIM, S. K., WANG, K. C., SCHEURLEN, W., EBERHART, C. G., FEVRE-MONTANGE, M., JOUVET, A., POLLACK, I. F., FAN, X., MURASZKO, K. M., GILLESPIE, G. Y., DI ROCCO, C., MASSIMI, L., MICHIELS, E. M., KLOOSTERHOF, N. K., FRENCH, P. J., KROS, J. M., OLSON, J. M., ELLENBOGEN, R. G., ZITTERBART, K., KREN, L., THOMPSON, R. C., COOPER, M. K., LACH, B., MCLENDON, R. E., BIGNER, D. D., FONTEBASSO, A., ALBRECHT, S., JABADO, N., LINDSEY, J. C., BAILEY, S., GUPTA, N., WEISS, W. A., BOGNAR, L., KLEKNER, A., VAN METER, T. E., KUMABE, T., TOMINAGA, T., ELBABAA, S. K., LEONARD, J. R., RUBIN, J. B.*, et al.* (2012). Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature,* 488**,** 49-56.

OKABE, S. (1997). Hypothesis--origin of the parietal cell: microorganism? *J Clin Gastroenterol,* 25 Suppl 1**,** S141-8.

OSBORNE, C., WILSON, P.andTRIPATHY, D. (2004). Oncogenes and tumor suppressor genes in breast cancer: potential diagnostic and therapeutic applications. *Oncologist,* 9**,** 361-77.

OSTERTAG, E. M., MADISON, B. B.andKANO, H. (2007). Mutagenesis in rodents using the L1 retrotransposon. *Genome Biol,* 8 Suppl 1**,** S16.

PACKER, R. J., GAJJAR, A., VEZINA, G., RORKE-ADAMS, L., BURGER, P. C., ROBERTSON, P. L., BAYER, L., LAFOND, D., DONAHUE, B. R., MARYMONT, M. H., MURASZKO, K., LANGSTON, J.andSPOSTO, R. (2006). Phase III study of craniospinal radiation therapy followed by adjuvant chemotherapy for newly diagnosed average-risk medulloblastoma. *J Clin Oncol,* 24**,** 4202-8.

PAJER, P., PECENKA, V., KRALOVA, J., KARAFIAT, V., PRUKOVA, D., ZEMANOVA, Z., KODET, R.andDVORAK, M. (2006). Identification of potential human oncogenes by mapping the common viral integration sites in avian nephroblastoma. *Cancer Res,* 66**,** 78-86.

PANG, Z. P., YANG, N., VIERBUCHEN, T., OSTERMEIER, A., FUENTES, D. R., YANG, T. Q., CITRI, A., SEBASTIANO, V., MARRO, S., SUDHOF, T. C.andWERNIG, M.

(2011). Induction of human neuronal cells by defined transcription factors. *Nature,* 476**,** 220-3.

PARSONS, D. W., LI, M., ZHANG, X., JONES, S., LEARY, R. J., LIN, J. C., BOCA, S. M., CARTER, H., SAMAYOA, J., BETTEGOWDA, C., GALLIA, G. L., JALLO, G. I., BINDER, Z. A., NIKOLSKY, Y., HARTIGAN, J., SMITH, D. R., GERHARD, D. S., FULTS, D. W., VANDENBERG, S., BERGER, M. S., MARIE, S. K., SHINJO, S. M., CLARA, C., PHILLIPS, P. C., MINTURN, J. E., BIEGEL, J. A., JUDKINS, A. R., RESNICK, A. C., STORM, P. B., CURRAN, T., HE, Y., RASHEED, B. A., FRIEDMAN, H. S., KEIR, S. T., MCLENDON, R., NORTHCOTT, P. A., TAYLOR, M. D., BURGER, P. C., RIGGINS, G. J., KARCHIN, R., PARMIGIANI, G., BIGNER, D. D., YAN, H., PAPADOPOULOS, N., VOGELSTEIN, B., KINZLER, K. W.andVELCULESCU, V. E. (2011). The genetic landscape of the childhood cancer medulloblastoma. *Science,* 331**,** 435-9.

PAZZAGLIA, S. (2006). Ptc1 heterozygous knockout mice as a model of multi-organ tumorigenesis. *Cancer Lett,* 234**,** 124-34.

PEI, Y., MOORE, C. E., WANG, J., TEWARI, A. K., EROSHKIN, A., CHO, Y. J., WITT, H., KORSHUNOV, A., READ, T. A., SUN, J. L., SCHMITT, E. M., MILLER, C. R., BUCKLEY, A. F., MCLENDON, R. E., WESTBROOK, T. F., NORTHCOTT, P. A., TAYLOR, M. D., PFISTER, S. M., FEBBO, P. G.andWECHSLER-REYA, R. J. (2012). An animal model of MYC-driven medulloblastoma. *Cancer Cell,* 21**,** 155-67.

PEREZ-CASELLAS, L. A., WANG, X., HOWARD, K. D., REHAGE, M. W., STRONG, D. D.andLINKHART, T. A. (2009). Nuclear factor I transcription factors regulate IGF binding protein 5 gene transcription in human osteoblasts. *Biochim Biophys Acta,* 1789**,** 78-87.

PFISTER, S., HARTMANN, C.andKORSHUNOV, A. (2009). Histology and molecular pathology of pediatric brain tumors. *J Child Neurol,* 24**,** 1375-86.

PIERCE, B. 2008. *Genetics: A conceptual Approach,* New York, W.H.Freeman and Co.

POLA, R., LING, L. E., APRAHAMIAN, T. R., BARBAN, E., BOSCH-MARCE, M., CURRY, C., CORBLEY, M., KEARNEY, M., ISNER, J. M.andLOSORDO, D. W. (2003). Postnatal recapitulation of embryonic hedgehog pathway in response to skeletal muscle ischemia. *Circulation,* 108**,** 479-85.

POMEROY, S. L., SUTTON, M. E., GOUMNEROVA, L. C.andSEGAL, R. A. (1997). Neurotrophins in cerebellar granule cell development and medulloblastoma. *J Neurooncol,* 35**,** 347-52.

PUGH, T. J., WEERARATNE, S. D., ARCHER, T. C., POMERANZ KRUMMEL, D. A., AUCLAIR, D., BOCHICCHIO, J., CARNEIRO, M. O., CARTER, S. L., CIBULSKIS, K., ERLICH, R. L., GREULICH, H., LAWRENCE, M. S., LENNON, N. J., MCKENNA, A., MELDRIM, J., RAMOS, A. H., ROSS, M. G., RUSS, C., SHEFLER, E., SIVACHENKO, A., SOGOLOFF, B., STOJANOV, P., TAMAYO, P., MESIROV, J. P., AMANI, V., TEIDER, N., SENGUPTA, S., FRANCOIS, J. P., NORTHCOTT, P. A., TAYLOR, M. D., YU, F., CRABTREE, G. R., KAUTZMAN, A. G., GABRIEL, S. B., GETZ, G., JAGER, N., JONES, D. T., LICHTER, P., PFISTER, S. M., ROBERTS, T. M., MEYERSON, M., POMEROY, S. L.andCHO, Y. J. (2012). Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature,* 488**,** 106-10.

RAHRMANN, E. P., COLLIER, L. S., KNUTSON, T. P., DOYAL, M. E., KUSLAK, S. L., GREEN, L. E., MALINOWSKI, R. L., ROETHE, L., AKAGI, K., WAKNITZ, M., HUANG, W., LARGAESPADA, D. A.andMARKER, P. C. (2009). Identification of PDE4D as a proliferation promoting factor in prostate cancer using a

Sleeping Beauty transposon-based somatic mutagenesis screen. *Cancer Res,* 69**,** 4388-97.

RAO, G., PEDONE, C. A., DEL VALLE, L., REISS, K., HOLLAND, E. C.andFULTS, D. W. (2004). Sonic hedgehog and insulin-like growth factor signaling synergize to induce medulloblastoma formation from nestin-expressing neural progenitors in mice. *Oncogene,* 23**,** 6156-62.

READ, T. A., FOGARTY, M. P., MARKANT, S. L., MCLENDON, R. E., WEI, Z., ELLISON, D. W., FEBBO, P. G.andWECHSLER-REYA, R. J. (2009). Identification of CD15 as a marker for tumor-propagating cells in a mouse model of medulloblastoma. *Cancer Cell,* 15**,** 135-47.

REES, M. I., HARVEY, K., WARD, H., WHITE, J. H., EVANS, L., DUGUID, I. C., HSU, C. C., COLEMAN, S. L., MILLER, J., BAER, K., WALDVOGEL, H. J., GIBBON, F., SMART, T. G., OWEN, M. J., HARVEY, R. J.andSNELL, R. G. (2003). Isoform heterogeneity of the human gephyrin gene (GPHN), binding domains to the glycine receptor, and mutation analysis in hyperekplexia. *J Biol Chem,* 278**,** 24688-96.

REISS, J., GROSS-HARDT, S., CHRISTENSEN, E., SCHMIDT, P., MENDEL, R. R.andSCHWARZ, G. (2001). A mutation in the gene for the neurotransmitter receptor-clustering protein gephyrin causes a novel form of molybdenum cofactor deficiency. *Am J Hum Genet,* 68**,** 208-13.

REMKE, M., HIELSCHER, T., NORTHCOTT, P. A., WITT, H., RYZHOVA, M., WITTMANN, A., BENNER, A., VON DEIMLING, A., SCHEURLEN, W., PERRY, A., CROUL, S., KULOZIK, A. E., LICHTER, P., TAYLOR, M. D., PFISTER, S. M.andKORSHUNOV, A. (2011). Adult medulloblastoma comprises three major molecular variants. *J Clin Oncol,* 29**,** 2717-23.

REYA, T., MORRISON, S. J., CLARKE, M. F.andWEISSMAN, I. L. (2001). Stem cells, cancer, and cancer stem cells. *Nature,* 414**,** 105-11.

RICH, J. N.andEYLER, C. E. (2008). Cancer stem cells in brain tumor biology. *Cold Spring Harb Symp Quant Biol,* 73**,** 411-20.

RIOS, I., ALVAREZ-RODRIGUEZ, R., MARTI, E.andPONS, S. (2004). Bmp2 antagonizes sonic hedgehog-mediated proliferation of cerebellar granule neurones through Smad5 signalling. *Development,* 131**,** 3159-68.

ROBINSON, G., PARKER, M., KRANENBURG, T. A., LU, C., CHEN, X., DING, L., PHOENIX, T. N., HEDLUND, E., WEI, L., ZHU, X., CHALHOUB, N., BAKER, S. J., HUETHER, R., KRIWACKI, R., CURLEY, N., THIRUVENKATAM, R., WANG, J., WU, G., RUSCH, M., HONG, X., BECKSFORT, J., GUPTA, P., MA, J., EASTON, J., VADODARIA, B., ONAR-THOMAS, A., LIN, T., LI, S., POUNDS, S., PAUGH, S., ZHAO, D., KAWAUCHI, D., ROUSSEL, M. F., FINKELSTEIN, D., ELLISON, D. W., LAU, C. C., BOUFFET, E., HASSALL, T., GURURANGAN, S., COHN, R., FULTON, R. S., FULTON, L. L., DOOLING, D. J., OCHOA, K., GAJJAR, A., MARDIS, E. R., WILSON, R. K., DOWNING, J. R., ZHANG, J.andGILBERTSON, R. J. (2012). Novel mutations target distinct subgroups of medulloblastoma. *Nature,* 488**,** 43-8.

ROMER, J. T., KIMURA, H., MAGDALENO, S., SASAI, K., FULLER, C., BAINES, H., CONNELLY, M., STEWART, C. F., GOULD, S., RUBIN, L. L.andCURRAN, T. (2004). Suppression of the Shh pathway using a small molecule inhibitor eliminates medulloblastoma in Ptc1(+/-)p53(-/-) mice. *Cancer Cell,* 6**,** 229-40.

ROSSI, A., CARACCIOLO, V., RUSSO, G., REISS, K.andGIORDANO, A. (2008). Medulloblastoma: from molecular pathology to therapy. *Clin Cancer Res,* 14**,** 971-6.

ROUSSEL, M. F.andHATTEN, M. E. (2011). Cerebellum development and medulloblastoma. *Curr Top Dev Biol,* 94**,** 235-82.

SALSANO, E., POLLO, B., EOLI, M., GIORDANA, M. T.andFINOCCHIARO, G. (2004). Expression of MATH1, a marker of cerebellar granule cell progenitors, identifies different medulloblastoma sub-types. *Neurosci Lett,* 370**,** 180-5.

SANCHEZ, C., DIAZ-NIDO, J.andAVILA, J. (2000). Phosphorylation of microtubule-associated protein 2 (MAP2) and its relevance for the regulation of the neuronal cytoskeleton function. *Prog Neurobiol,* 61**,** 133-68.

SANCHEZ, P., HERNANDEZ, A. M., STECCA, B., KAHLER, A. J., DEGUEME, A. M., BARRETT, A., BEYNA, M., DATTA, M. W., DATTA, S.andRUIZ I ALTABA, A. (2004). Inhibition of prostate cancer proliferation by interference with SONIC HEDGEHOG-GLI1 signaling. *Proc Natl Acad Sci U S A,* 101**,** 12561-6.

SAWADA, A., KIYONARI, H., UKITA, K., NISHIOKA, N., IMUTA, Y.andSASAKI, H. (2008). Redundant roles of Tead1 and Tead2 in notochord development and the regulation of cell proliferation and survival. *Mol Cell Biol,* 28**,** 3177-89.

SCHULLER, U., HEINE, V. M., MAO, J., KHO, A. T., DILLON, A. K., HAN, Y. G., HUILLARD, E., SUN, T., LIGON, A. H., QIAN, Y., MA, Q., ALVAREZ-BUYLLA, A., MCMAHON, A. P., ROWITCH, D. H.andLIGON, K. L. (2008). Acquisition of granule neuron precursor identity is a critical determinant of progenitor cell competence to form Shh-induced medulloblastoma. *Cancer Cell,* 14**,** 123-34.

SCHWAB, M., ALITALO, K., KLEMPNAUER, K. H., VARMUS, H. E., BISHOP, J. M., GILBERT, F., BRODEUR, G., GOLDSTEIN, M.andTRENT, J. (1983). Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature,* 305**,** 245-8.

SHAHI, M. H., AFZAL, M., SINHA, S., EBERHART, C. G., REY, J. A., FAN, X.andCASTRESANA, J. S. (2011). Human hedgehog interacting protein expression and promoter methylation in medulloblastoma cell lines and primary tumor samples. *J Neurooncol,* 103**,** 287-96.

SHENG, T., LI, C., ZHANG, X., CHI, S., HE, N., CHEN, K., MCCORMICK, F., GATALICA, Z.andXIE, J. (2004). Activation of the hedgehog pathway in advanced prostate cancer. *Mol Cancer,* 3**,** 29.

SHING, D. C., TRUBIA, M., MARCHESI, F., RADAELLI, E., BELLONI, E., TAPINASSI, C., SCANZIANI, E., MECUCCI, C., CRESCENZI, B., LAHORTIGA, I., ODERO, M. D., ZARDO, G., GRUSZKA, A., MINUCCI, S., DI FIORE, P. P.andPELICCI, P. G. (2007). Overexpression of sPRDM16 coupled with loss of p53 induces myeloid leukemias in mice. *J Clin Invest,* 117**,** 3696-707.

SINGH, S. K., HAWKINS, C., CLARKE, I. D., SQUIRE, J. A., BAYANI, J., HIDE, T., HENKELMAN, R. M., CUSIMANO, M. D.andDIRKS, P. B. (2004). Identification of human brain tumour initiating cells. *Nature,* 432**,** 396-401.

STARR, T. K., ALLAEI, R., SILVERSTEIN, K. A., STAGGS, R. A., SARVER, A. L., BERGEMANN, T. L., GUPTA, M., O'SULLIVAN, M. G., MATISE, I., DUPUY, A. J., COLLIER, L. S., POWERS, S., OBERG, A. L., ASMANN, Y. W., THIBODEAU, S. N., TESSAROLLO, L., COPELAND, N. G., JENKINS, N. A., CORMIER, R. T.andLARGAESPADA, D. A. (2009). A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. *Science,* 323**,** 1747-50.

STARR, T. K.andLARGAESPADA, D. A. (2005). Cancer gene discovery using the Sleeping Beauty transposon. *Cell Cycle,* 4**,** 1744-8.

STARR, T. K., SCOTT, P. M., MARSH, B. M., ZHAO, L., THAN, B. L., O'SULLIVAN, M. G., SARVER, A. L., DUPUY, A. J., LARGAESPADA, D. A.andCORMIER, R. T. (2011).

A Sleeping Beauty transposon-mediated screen identifies murine susceptibility genes for adenomatous polyposis coli (Apc)-dependent intestinal tumorigenesis. *Proc Natl Acad Sci U S A,* 108**,** 5765-70.

STEELE-PERKINS, G., PLACHEZ, C., BUTZ, K. G., YANG, G., BACHURSKI, C. J., KINSMAN, S. L., LITWACK, E. D., RICHARDS, L. J.andGRONOSTAJSKI, R. M. (2005). The transcription factor gene Nfib is essential for both lung maturation and brain development. *Mol Cell Biol,* 25**,** 685-98.

SU, Q., PROSSER, H. M., CAMPOS, L. S., ORTIZ, M., NAKAMURA, T., WARREN, M., DUPUY, A. J., JENKINS, N. A., COPELAND, N. G., BRADLEY, A.andLIU, P. (2008). A DNA transposon-based approach to validate oncogenic mutations in the mouse. *Proc Natl Acad Sci U S A,* 105**,** 19904-9.

SUBKHANKULOVA, T., ZHANG, X., LEUNG, C.andMARINO, S. (2010). Bmi1 directly represses p21Waf1/Cip1 in Shh-induced proliferation of cerebellar granule cell progenitors. *Mol Cell Neurosci,* 45**,** 151-62.

SUTTER, R., SHAKHOVA, O., BHAGAT, H., BEHESTI, H., SUTTER, C., PENKAR, S., SANTUCCIONE, A., BERNAYS, R., HEPPNER, F. L., SCHULLER, U., GROTZER, M., MOCH, H., SCHRAML, P.andMARINO, S. (2010). Cerebellar stem cells act as medulloblastoma-initiating cells in a mouse model and a neural stem cell signature characterizes a subset of human medulloblastomas. *Oncogene,* 29**,** 1845-56.

SUZUKI, N., ANDO, S., SUMIDA, K., HORIE, N.andSAITO, K. (2011). Analysis of altered gene expression specific to embryotoxic chemical treatment during embryonic stem cell differentiation into myocardiac and neural cells. *J Toxicol Sci,* 36**,** 569-85.

SVAREN, J., EHRIG, T., ABDULKADIR, S. A., EHRENGRUBER, M. U., WATSON, M. A.andMILBRANDT, J. (2000). EGR1 target genes in prostate carcinoma cells identified by microarray analysis. *J Biol Chem,* 275**,** 38524-31.

SWARTLING, F. J., GRIMMER, M. R., HACKETT, C. S., NORTHCOTT, P. A., FAN, Q. W., GOLDENBERG, D. D., LAU, J., MASIC, S., NGUYEN, K., YAKOVENKO, S., ZHE, X. N., GILMER, H. C., COLLINS, R., NAGAOKA, M., PHILLIPS, J. J., JENKINS, R. B., TIHAN, T., VANDENBERG, S. R., JAMES, C. D., TANAKA, K., TAYLOR, M. D., WEISS, W. A.andCHESLER, L. (2010). Pleiotropic role for MYCN in medulloblastoma. *Genes Dev,* 24**,** 1059-72.

TAMAYO, P., SCANFELD, D., EBERT, B. L., GILLETTE, M. A., ROBERTS, C. W.andMESIROV, J. P. (2007). Metagene projection for cross-platform, cross-species characterization of global transcriptional states. *Proc Natl Acad Sci U S A,* 104**,** 5959-64.

TANIGUCHI, K., ANDERSON, A. E., SUTHERLAND, A. E.andWOTTON, D. (2012). Loss of Tgif function causes holoprosencephaly by disrupting the SHH signaling pathway. *PLoS Genet,* 8**,** e1002524.

TARKKONEN, K. M., NILSSON, E. M., KAHKONEN, T. E., DEY, J. H., HEIKKILA, J. E., TUOMELA, J. M., LIU, Q., HYNES, N. E.andHARKONEN, P. L. (2012). Differential roles of fibroblast growth factor receptors (FGFR) 1, 2 and 3 in the regulation of S115 breast cancer cell growth. *PLoS One,* 7**,** e49970.

TAVAKOLI, N. P.andDERBYSHIRE, K. M. (2001). Tipping the balance between replicative and simple transposition. *Embo Journal,* 20**,** 2923-2930.

TAYLOR, M. D., NORTHCOTT, P. A., KORSHUNOV, A., REMKE, M., CHO, Y. J., CLIFFORD, S. C., EBERHART, C. G., PARSONS, D. W., RUTKOWSKI, S., GAJJAR, A., ELLISON, D. W., LICHTER, P., GILBERTSON, R. J., POMEROY, S. L., KOOL, M.andPFISTER, S. M. (2012). Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol,* 123**,** 465-72.

THAYER, S. P., DI MAGLIANO, M. P., HEISER, P. W., NIELSEN, C. M., ROBERTS, D. J., LAUWERS, G. Y., QI, Y. P., GYSIN, S., FERNANDEZ-DEL CASTILLO, C., YAJNIK, V., ANTONIU, B., MCMAHON, M., WARSHAW, A. L.andHEBROK, M. (2003). Hedgehog is an early and late mediator of pancreatic cancer tumorigenesis. *Nature,* 425**,** 851-6.

THOMPSON, M. C., FULLER, C., HOGG, T. L., DALTON, J., FINKELSTEIN, D., LAU, C. C., CHINTAGUMPALA, M., ADESINA, A., ASHLEY, D. M., KELLIE, S. J., TAYLOR, M. D., CURRAN, T., GAJJAR, A.andGILBERTSON, R. J. (2006). Genomics identifies medulloblastoma subgroups that are enriched for specific genetic alterations. *J Clin Oncol,* 24**,** 1924-31.

TODD, R.andWONG, D. T. (1999). Oncogenes. *Anticancer Res,* 19**,** 4729-46.

TUBERGEN, D. G., GILCHRIST, G. S., O'BRIEN, R. T., COCCIA, P. F., SATHER, H. N., WASKERWITZ, M. J.andHAMMOND, G. D. (1993). Improved outcome with delayed intensification for children with acute lymphoblastic leukemia and intermediate presenting features: a Childrens Cancer Group phase III trial. *J Clin Oncol,* 11**,** 527-37.

UCHIDA, N., BUCK, D. W., HE, D., REITSMA, M. J., MASEK, M., PHAN, T. V., TSUKAMOTO, A. S., GAGE, F. H.andWEISSMAN, I. L. (2000). Direct isolation of human central nervous system stem cells. *Proc Natl Acad Sci U S A,* 97**,** 14720-5.

UREN, A. G., KOOL, J., BERNS, A.andVAN LOHUIZEN, M. (2005). Retroviral insertional mutagenesis: past, present and future. *Oncogene,* 24**,** 7656-72.

UZIEL, T., ZINDY, F., XIE, S., LEE, Y., FORGET, A., MAGDALENO, S., REHG, J. E., CALABRESE, C., SOLECKI, D., EBERHART, C. G., SHERR, S. E., PLIMMER, S., CLIFFORD, S. C., HATTEN, M. E., MCKINNON, P. J., GILBERTSON, R. J., CURRAN, T., SHERR, C. J.andROUSSEL, M. F. (2005). The tumor suppressors Ink4c and p53 collaborate independently with Patched to suppress medulloblastoma formation. *Genes Dev,* 19**,** 2656-67.

VAQUERIZAS, J. M., KUMMERFELD, S. K., TEICHMANN, S. A.andLUSCOMBE, N. M. (2009). A census of human transcription factors: function, expression and evolution. *Nat Rev Genet,* 10**,** 252-63.

VENKATARAMAN, S., ALIMOVA, I., FAN, R., HARRIS, P., FOREMAN, N.andVIBHAKAR, R. (2010a). MicroRNA 128a increases intracellular ROS level by targeting Bmi-1 and inhibits medulloblastoma cancer cell growth by promoting senescence. *PLoS One,* 5**,** e10748.

VENKATARAMAN, S., DIRKS, R. M., UEDA, C. T.andPIERCE, N. A. (2010b). Selective cell death mediated by small conditional RNAs. *Proc Natl Acad Sci U S A,* 107**,** 16777-82.

VIBHAKAR, R., FOLTZ, G., YOON, J. G., FIELD, L., LEE, H., RYU, G. Y., PIERSON, J., DAVIDSON, B.andMADAN, A. (2007). Dickkopf-1 is an epigenetically silenced candidate tumor suppressor gene in medulloblastoma. *Neuro Oncol,* 9**,** 135-44.

VIERBUCHEN, T., OSTERMEIER, A., PANG, Z. P., KOKUBU, Y., SUDHOF, T. C.andWERNIG, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature,* 463**,** 1035-41.

VOGELSTEIN, B.andKINZLER, K. W. (2004). Cancer genes and the pathways they control. *Nat Med,* 10**,** 789-99.

VOOGD, J.andGLICKSTEIN, M. (1998). The anatomy of the cerebellum. *Trends Cogn Sci,* 2**,** 307-13.

WALLACE, M. R., ANDERSEN, L. B., SAULINO, A. M., GREGORY, P. E., GLOVER, T. W.andCOLLINS, F. S. (1991). A de novo Alu insertion results in neurofibromatosis type 1. *Nature,* 353**,** 864-6.

WALLACE, V. A. (1999). Purkinje-cell-derived Sonic hedgehog regulates granule neuron precursor cell proliferation in the developing mouse cerebellum. *Curr Biol,* 9**,** 445-8.

WANG, W., MULLIKIN-KILPATRICK, D., CRANDALL, J. E., GRONOSTAJSKI, R. M., LITWACK, E. D.andKILPATRICK, D. L. (2007). Nuclear factor I coordinates multiple phases of cerebellar granule cell development via regulation of cell adhesion molecules. *J Neurosci,* 27**,** 6115-27.

WANG, W., STOCK, R. E., GRONOSTAJSKI, R. M., WONG, Y. W., SCHACHNER, M.andKILPATRICK, D. L. (2004). A role for nuclear factor I in the intrinsic control of cerebellar granule neuron gene expression. *J Biol Chem,* 279**,** 53491-7.

WECHSLER-REYA, R. J.andSCOTT, M. P. (1999). Control of neuronal precursor proliferation in the cerebellum by Sonic Hedgehog. *Neuron,* 22**,** 103-14.

WETMORE, C., EBERHART, D. E.andCURRAN, T. (2001). Loss of p53 but not ARF accelerates medulloblastoma in mice heterozygous for patched. *Cancer Res,* 61**,** 513-6.

WORLD, O. H. January 2013. *cancer* [Online]. Available: http://www.who.int/mediacentre/factsheets/fs297/en/index.html [Accessed 18/11/2013 18/11/2013].

WU, X., NORTHCOTT, P. A., DUBUC, A., DUPUY, A. J., SHIH, D. J., WITT, H., CROUL, S., BOUFFET, E., FULTS, D. W., EBERHART, C. G., GARZIA, L., VAN METER, T., ZAGZAG, D., JABADO, N., SCHWARTZENTRUBER, J., MAJEWSKI, J., SCHEETZ, T. E., PFISTER, S. M., KORSHUNOV, A., LI, X. N., SCHERER, S. W., CHO, Y. J., AKAGI, K., MACDONALD, T. J., KOSTER, J., MCCABE, M. G., SARVER, A. L., COLLINS, V. P., WEISS, W. A., LARGAESPADA, D. A., COLLIER, L. S.andTAYLOR, M. D. (2012). Clonal selection drives genetic divergence of metastatic medulloblastoma. *Nature,* 482**,** 529-33.

WU, Y., SONG, P., XU, J., ZHANG, M.andZOU, M. H. (2007). Activation of protein phosphatase 2A by palmitate inhibits AMP-activated protein kinase. *J Biol Chem,* 282**,** 9777-88.

XU, P., YU, S., JIANG, R., KANG, C., WANG, G., JIANG, H.andPU, P. (2009). Differential expression of Notch family members in astrocytomas and medulloblastomas. *Pathol Oncol Res,* 15**,** 703-10.

YANG, Z. J., ELLIS, T., MARKANT, S. L., READ, T. A., KESSLER, J. D., BOURBOULAS, M., SCHULLER, U., MACHOLD, R., FISHELL, G., ROWITCH, D. H., WAINWRIGHT, B. J.andWECHSLER-REYA, R. J. (2008). Medulloblastoma can be initiated by deletion of Patched in lineage-restricted progenitors or stem cells. *Cancer Cell,* 14**,** 135-45.

YAUCH, R. L., DIJKGRAAF, G. J., ALICKE, B., JANUARIO, T., AHN, C. P., HOLCOMB, T., PUJARA, K., STINSON, J., CALLAHAN, C. A., TANG, T., BAZAN, J. F., KAN, Z., SESHAGIRI, S., HANN, C. L., GOULD, S. E., LOW, J. A., RUDIN, C. M.andDE SAUVAGE, F. J. (2009). Smoothened mutation confers resistance to a Hedgehog pathway inhibitor in medulloblastoma. *Science,* 326**,** 572-4.

YIIN, J. J., HU, B., JARZYNKA, M. J., FENG, H., LIU, K. W., WU, J. Y., MA, H. I.andCHENG, S. Y. (2009). Slit2 inhibits glioma cell invasion in the brain by suppression of Cdc42 activity. *Neuro Oncol,* 11**,** 779-89.

YU, J., DE BELLE, I., LIANG, H.andADAMSON, E. D. (2004). Coactivating factors p300 and CBP are transcriptionally crossregulated by Egr1 in prostate cells, leading to divergent responses. *Mol Cell,* 15**,** 83-94.

ZAYED, H., IZSVAK, Z., WALISKO, O.andIVICS, Z. (2004). Development of hyperactive sleeping beauty transposon vectors by mutational analysis. *Mol Ther,* 9**,** 292-304.

ZHANG, B., PAN, X., COBB, G. P.andANDERSON, T. A. (2007). microRNAs as oncogenes and tumor suppressors. *Dev Biol,* 302**,** 1-12.

ZHANG, L.andGOLDMAN, J. E. (1996). Developmental fates and migratory pathways of dividing progenitors in the postnatal rat cerebellum. *J Comp Neurol,* 370**,** 536-50.

ZHAO, H., AYRAULT, O., ZINDY, F., KIM, J. H.andROUSSEL, M. F. (2008). Post-transcriptional down-regulation of Atoh1/Math1 by bone morphogenic proteins suppresses medulloblastoma development. *Genes Dev,* 22**,** 722-7.

ZINDY, F., KNOEPFLER, P. S., XIE, S., SHERR, C. J., EISENMAN, R. N.andROUSSEL, M. F. (2006). N-Myc and the cyclin-dependent kinase inhibitors p18Ink4c and p27Kip1 coordinately regulate cerebellar development. *Proc Natl Acad Sci U S A,* 103**,** 11579-83.

ZURAWEL, R. H., ALLEN, C., WECHSLER-REYA, R., SCOTT, M. P.andRAFFEL, C. (2000). Evidence that haploinsufficiency of Ptch leads to medulloblastoma in mice. *Genes Chromosomes Cancer,* 28**,** 77-81.

(World, January 2013)World Health Organization (2013), (Cancer, (a);Cancer, (b);Cancer, (C))