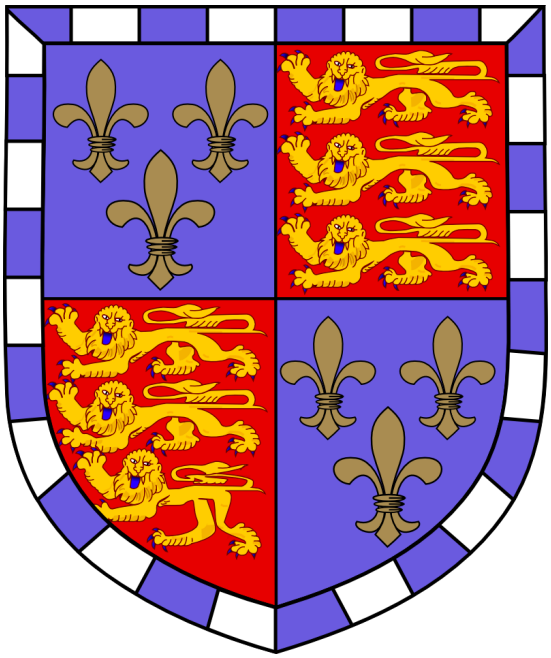
Genetic dissection of *EGFRvIII* brain and spinal mouse gliomas through whole-exome sequencing and *in vivo piggyBac* mutagenesis forward genetic screening



Imran Noorani

Christ's College, University of Cambridge

December 2018

This dissertation is submitted for the degree of Doctor of Philosophy

Preface

- This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.
- It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text
- It does not exceed the prescribed word limit for the relevant Degree Committee (excluding Supplementary Tables and References).
- This is a resubmitted version of my Thesis, with corrections completed as previously advised by my Examiners.

Publications

Publications that have resulted from work conducted during my PhD are:

1. **Noorani I** et al. '*PiggyBac* mutagenesis and exome-sequencing identify genetic driver lanscapes and potential therapeutic targets of *EGFR*-mutant gliomas' – submitted and under peer review. Note that the majority of the Results sections for Chapters 3 and 4 of this Thesis are also presented in this paper for publication.
2. Collord G, Tarpey P, Kurbatova N, Martincorena I, Moran S, Castro M, Nagy T, Bignell G, Maura F, Young MD, Berna J, Tubio JMC, McMurran CE, Young AMH, Sanders M, **Noorani I**, Price SJ, Watts C, Leipnitz E, Kirsch M, Schackert G, Pearson D, Devadass A, Ram Z, Collins VP, Allinson K, Jenkinson MD, Zakaria R, Syed K, Hanemann CO, Dunn J, McDermott MW, Kirollos RW, Vassiliou GS, Esteller M, Behjati S, Brazma A, Santarius T, McDermott U. An integrated

genomic analysis of anaplastic meningioma identifies prognostic molecular signatures. Scientific Reports. 2018 Sep 10;8(1):13537. PubMed PMID: 30202034;

3. **Noorani I**, Sanai N. Surgical Management of Incidental Gliomas. Neurosurgical Clinics of North America. 2017 Jul;28(3):397-406. Review. PMID: 28600014.
4. de la Rosa J, Weber J, Friedrich MJ, Li Y, Rad L, Ponstingl H, Liang Q, de Quirós SB, **Noorani I**, Metzakopian E, Strong A, Li MA, Astudillo A, Fernández-García MT, Fernández-García MS, Hoffman GJ, Fuente R, Vassiliou GS, Rad R, López-Otín C, Bradley A, Cadiñanos J. A single-copy Sleeping Beauty transposon mutagenesis screen identifies new PTEN-cooperating tumor suppressor genes. Nature Genetics. 2017 May;49(5):730-741. PMID: 28319090.

Acknowledgements

I would like to thank my supervisor Professor Allan Bradley for his helpful support and guidance during my PhD, Professor Sebastian Brandner for reviewing the histopathology of my mouse tumors and for his advice on the study, Mr Thomas Santarius for enlightening discussions on this work, and for the support of other members of the lab and the Wellcome Trust Sanger Institute, in particular Alex Strong, Hannes Ponstingl and Jorge de la Rosa. I am very grateful to my parents for all of their encouragement in my efforts.

List of main corrections in this revised thesis:

1. The title of the thesis has been changed to more closely match the content of the body of work contained herein.
2. A new chapter has been written on 'Materials and Methods' which provides methodological details from all experimental chapters in a logical order. References to published works are included as appropriate.
3. Details of somatic variant calling as part of whole-exome sequencing data analysis are described in the Methods.
4. The site at which the *EGFRvIII* transgene is inserted in mice is stated early in the Methods section.
5. RNA-sequencing data analysis to determine differentially expressed genes, as well as gene set enrichment analysis, detection of *EGFRvIII* transcripts and transposon fusion transcripts, are all described in the Methods.
6. Methods used for comparative genomic analysis with human tumor data are described in the Methods chapter, including reference to the publicly available software Cbioportal.
7. Further detail of immunohistochemical staining protocols is provided.
8. Both low as well as high power views are now provided for all pathology images of tumors shown, enabling the reader to better appreciate the details of the histology.
9. Details of how gliomas were graded based on histological assessment by a neuropathologist are fully described in the Materials and Methods section.
10. Pathology images have been aligned for appropriate presentation, and with scale bars presented.
11. The majority of figures have been re-drawn to meet high-quality publication standards.
12. Figures have been referenced in the text appropriately.
13. Figure legends have been revised throughout to give more detailed information.
14. The results of gene set enrichment analysis of RNA-seq data from our mouse tumors are now presented.
15. A completely new analysis with new data has been provided for RNA-sequencing of EGFRvIII tumors with *piggyBac* transposition, highlighting detection of fusion transcripts which further supports the data provided DNA analysis of transposon insertion sites.
16. To address concerns of both examiners regarding the demonstrating EGFR recombination in tumors of our mice, I have provided new data of immunohistochemical staining for EGFR and

EGFRvIII showing specific expression in tumors or tumor precursor lesions. As previously, these have also been reviewed by a Consultant Neuropathologist (Prof Brandner) who agrees with the pathology labelled by these stains.

17. Further consideration is given to the locations of recombination with nestin-cre with appropriate citation of the literature.

18. Tables of all mice produced in this study are listed, including age at which they were culled and histology (where histology is available).

19. The aims of Chapter Three have been altered to emphasise the focus on studying the genetics of *EGFRvIII* gliomas.

20. The Discussion has been revised to include new sections expanding on the development of therapies targeting EGFR, and the challenges faced by these for treating glioblastoma in patients.

21. Abbreviations have been explained at first usage.

22. The number of animals / samples have been stated in all experiments.

23. The statistical tests used are named next to each p-value stated in the Results sections.

24. For survival analysis of patients based on gene expression levels, the same cut-off for 'high' versus 'low' gene expression has been selected for all genes.

25. Sections of the Results describing data on known mutations in gliomas from publicly available patient databases have been removed, and the publications first presenting these data are simply referred to instead.

26. A sub-section discussing the limitations of my work has been added in the Discussion for all Chapters presenting new data.

27. The Introduction to Chapter 5 has been rewritten in regards to the background on medulloblastoma, correcting previous inaccuracies.

28. The term 'cross-species' has been corrected to 'comparative' in regards to genomics comparing mouse and humans in this work.

29. Supplementary tables for whole-exome sequencing, RNA-sequencing, *piggyBac* common integration sites, and fusion transcripts are included.

30. A new section has been added to the Discussion Chapter on novel developments in transposon mutagenesis screening.

31. The possible reasons why EGFRvIII was observed to induce gliomas in my mouse model whereas in it was not in previous publications are explained in the discussion for Chapter 3.

32. A description of limitations of the work is provided at the end of each experimental Chapter.

33. Referencing format is now consistent across the Thesis.

Table of Contents

Summary of Work

Genetic dissection of *EGFR*vIII brain and spinal gliomas through whole-exome sequencing and *PiggyBac* mutagenesis forward genetic screening

Imran Noorani

Glioma is the commonest intrinsic brain tumor, and its high-grade form has a devastating prognosis. These tumors also arise in the spinal cord, carrying significant morbidity in children; however the genetics of these spinal gliomas is poorly understood. *EGFRvIII* is a common driver mutation in brain gliomas; it is unclear when this is acquired during glioma evolution and what its cooperative genetic drivers are. Here, we show that *EGFRvIII* initiates gliomagenesis *in vivo*; *EGFRvIII* leads to glioma precursors in the subventricular zone and brain surface, and later glioma formation in the brain and spinal cord. The long latency for tumor formation implies the need for additional mutations to drive gliomagenesis. In these tumors, we detected further genetic alterations including amplification of *EGFRvIII*, mutations of *Trp53* and *Tead2*, and *Cdkn2a* deletion, through whole-exome sequencing. To shed further light on *EGFR*-cooperative genes for glioma progression, we conducted a genome-wide *piggyBac* transposon mutagenesis screen *in vivo*, which identified known glioma drivers (including *Cdkn2a*, *Pten* and *Nf1*) and novel putative partners, including genes that regulate neuronal differentiation such as *Sox6* and *Tcf12*, and a novel regulator of the Ras pathway *Spred1*. RNA-sequencing confirmed the presence of fusion transcripts (transposon mediated effects) for these genes. We demonstrate the clinical relevance of these cooperative genes through comparison with large human glioma databases, demonstrating recurrent genetic alterations of these genes are in patient tumors implicating them as putative drivers, and we highlight that expression levels of *Sox6* and *Tcf12* correlate with patient prognosis. We show that there are shared and distinct mutated genes in brain and spinal gliomas. Although *Pten* is a well-known tumor suppressor for brain gliomas, it was not previously known whether *Pten* drives spinal gliomagenesis. Given recurrent transposon insertions in *Pten* were found in both brain and spinal gliomas, we generated conditional mice with *EGFRvIII* and *Pten* loss, demonstrating *Pten* accelerates spinal glioma formation. Our work elucidates the genetic evolutionary processes behind *EGFRvIII*-driven gliomas, provides a detailed genomic comparison between brain and spinal gliomas, and provides functional genomic datasets to help decipher complex human glioma genomes.

Chapter One: Thesis Introduction

Cancer

Cancer is a disease in which abnormal cells divide in an uncontrolled way and have the ability to invade surrounding tissues and metastasise to other regions of the body. Diverse cancers have been proposed to display several 'hallmarks', and it has been suggested that as normal cells evolve towards a cancerous state they necessarily acquire these hallmarks in a stepwise fashion [1]. It is thought that a predominant mechanism enabling acquisition of these hallmarks is somatic mutation in the preneoplastic and neoplastic cells [2], although of course there are also important contributions from epigenetic changes, germline genetic variants, and effects from the tumor microenvironment. Certain cell types are more likely to become cancerous than others, and the cell-of-origin can have important consequences for a tumor. These processes are actively being explored and still not fully understood.

Hallmarks of Cancer

A key characteristic of all cancers is sustained cellular proliferation. In order to achieve this, there must be activation of cellular signalling that induces a cell to enter and progress through the cell cycle. A common mechanism to achieve this is through growth factor receptor signalling (with activation of intracellular kinase domains) that turn on proliferative signalling cascades. Several possible ways cancer cells can do this are through: production of excessive growth factors by the cancer cells themselves for autocrine signalling; production of signalling molecules by cancer cells that induce the surrounding stromal tissue to produce growth factors that reciprocally drive cancer cell proliferation [3]; increased growth factor receptor protein on the cancer cell surface; structural changes in the receptor proteins that allow signalling independent of ligand-binding. These many ways of activating cellular signalling often converge on similar pathways, such as the Ras-MAPK and PI3K-AKT pathways that activate transcriptional programs that stimulate entry and progression through the cell cycle. For example, around 40% of melanomas have activating mutations in B-raf giving increased

signalling through the Raf to MAPK pathway, and about 40% of gliomas have activating mutations in EGFR (epidermal growth factor receptor) that turn on the PI3K-AKT pathway.

Normal cells possess several negative feedback loops that act to regulate cell proliferation and suppress excessive cell divisions by diminishing activation of signalling circuitry [4]. An example of this is the PTEN phosphatase which negatively regulates PI3K signalling through degradation of PIP3; many cancers often show loss of function mutations or deletions of the *PTEN* gene leading to increased proliferative PI3K signalling [5]. Typically, early in cancer formation the cells acquire mutations in oncogenes that activate proliferative pathways such as Ras and also in tumor suppressor genes (TSGs) that normally suppress these pathways, giving synergistic effects to maximise proliferative stimuli on the cells. However, many studies suggest that excessive oncogenic signalling in normal cells may trigger processes to counteract this as a protective mechanism, processes including senescence and apoptosis [6]. It is thought that cancer cells have overcome these defence mechanisms by disabling senescence or apoptosis processes. An important example of this is the *TP53* gene, which can activate cell death in response to oncogenic signals to protect against cancer formation; p53 was one of the earliest tumor suppressor genes to be discovered. In a similar fashion, *CDKN2A* is a tumor suppressor gene that activates upon expression of oncogenic signalling to trigger senescence and help prevent cancer formation. CDKN2A functions through the Rb (retinoblastoma-associated) pathway. The Rb protein is itself a critical tumor suppressor, loss of which is found a variety of cancers. Rb is thought to transduce signals external to the cell that are growth inhibitory and acts as a cell-cycle progression gatekeeper; in contrast, TP53 senses damage from within the cell such as to the genome, and if such DNA damage is beyond a critical level then TP53 can trigger apoptosis. Knockout mice lacking *Trp53* have normal development of tissues and organs but develop cancers later on including sarcomas and leukaemias, reflecting increased cellular proliferation due to lack of this important tumor suppressor [7].

Cancers have acquired a special property called immortalisation, whereby they can undergo an unlimited number of cell divisions unlike normal cells which can only a divide a limited number of times before entering a crisis state and undergoing cell death through senescence.

A critical step in immortalisation is overcoming the natural shortening of telomeres that occurs with successive cell divisions. Telomeres are multiple tandem hexanucleotide repeats that protect chromosomal ends, and as they are destroyed cells enter a crisis state. Cancers acquire the ability to stop telomerase erosion through expression of telomerase, a DNA polymerase that can regenerate telomere DNA. This enzyme is usually absent in normal tissue but is expressed in 90% of immortalised cells, highlighting it is an important early requirement for cellular immortalisation and therefore carcinogenesis [8].

Tumors require oxygen and nutrients to survive, even more so than normal tissue because tumors are continuously trying to grow and expand. The ability to form capillaries and other blood vessels is tightly regulated in development. However, cancers have a unique ability to break this regulation and turn on an 'angiogenic switch', whereby the balance between angiogenic stimulators and inhibitors is turned heavily in favour of activation. As such, cancers continuously produce new blood vessels and microvessels to support their growth, although these vessels tend to be abnormal themselves as evidenced by their leakiness and haemorrhage which is so characteristic of cancer. It is thought that angiogenesis is a property acquired early in cancer formation, since these abnormal blood vessels are seen in preneoplastic lesions. Common examples of angiogenic factors are VEGF (vascular endothelial growth factor) and FGF (fibroblast growth factor); VEGF overexpression is often seen in cancer [9].

Immune surveillance programs are an important checkpoint to halt early neoplastic formation – preneoplastic cells express novel antigens on their cell surface that can potentially be recognised by immune cells, which can then potentially kill them before they progress. Cancer formation thus requires the ability to escape detection by the immune system and / or mitigate destruction by it. This hypothesis is supported by experimental work in mice showing tumor incidence is increased when tumors are transplanted into immunocompromised mice compared with immunocompetent ones; in particular, mice with low numbers of natural killer (NK) cells, CD8+ T cells or CD4+ T-cells generate more tumors in this fashion suggesting these cells are important for tumor immunity [10]. Clinical observations also support these ideas given that, for example, patients with colon cancer and high numbers of T cells in their tumor

tend to have a better prognosis most likely because there is some degree of tumor immunity associated with infiltration of these T cells [11]. PD-L1 (programmed death ligand 1) inhibitors have proven successful in targeting interactions between cancer cells and T-cells. PD-L1 on some tumor cells interacts with PD-1 (programmed cell death protein 1) on T cells to suppress T cells from destroying cancer cells. PD-L1 inhibitors have proven successful in activating T cell immune destruction of tumor cells, improving patient survival in certain cases of advanced melanoma, renal cell carcinoma and non-small cell lung cancer [12, 13].

As implied from the descriptions above, a combination of mutations is required for cancer initiation and progression. When a cell acquires a mutation driving cellular proliferation, this mutation becomes selected for as the cell population carrying it expands as part of a Darwinian natural selection process. A subpopulation of these cells may then acquire an additional mutation in an oncogene or tumor suppressor gene that gives an extra growth advantage, generating further clonal selection for cells carrying this combination of mutations. This process of cancer gene mutation and clonal selection can occur repeatedly, and this is the basis of the multistep tumor progression model. Analysis of cancer genomes reveals gains and losses of many chromosomal regions; this was noticed even before modern next generation sequencing (NGS) technologies were available, for example with the use of comparative genome hybridisation (CGH). Now, with NGS it is even more apparent that cancer genomes have widespread mutations and copy number changes. These findings all imply that an elevated mutational rate is important in carcinogenesis, since this will allow preneoplastic cells to gain the required mutations in necessary cancer genes [14, 15]. Defects in cellular DNA repair machinery and / or in the mechanisms that normally protect the cell from the damaging effects of external DNA mutagens can lead to the undesirable effect of increased mutational rate, predisposing to cancer formation [16]. Once again, TP53 is an example of such a molecule, which can sense DNA damage and initiate repair mechanisms as a ‘guardian of the genome’. Similarly, caretaker genes that participate in DNA mismatch repair, for instance *MSH2*, are often tumor suppressor genes.

Over the last few decades, it has been increasingly recognised that there is a great deal of cellular intratumor heterogeneity, with different cell subpopulations in a tumor having

different phenotypes and genetics. Moreover, these different cell populations can have varying degrees of tumorigenicity – not all cancer cells are born equal. A prominent hypothesis that has emerged from these observations is the cancer stem cell (CSC) hypothesis which states that cancers consist of a hierarchy of cells with varying degrees of differentiation, with cancer stem cells sitting at the top of this hierarchy and able to generate all of the more differentiated progeny [17]. These cancer stem cells often express the same markers of stemness that are used to distinguish their normal stem cell counterparts, and their transcriptomic profiles often overlap with normal stem cells from the same tissue of origin. Importantly, cancer stem cells have the ability to efficiently initiate tumor formation in mice, in contrast to more differentiated cancer cells which are much less efficient at doing so. These findings have prompted many to speculate that cancers themselves originate from normal stem cells that acquire mutations in oncogenes or TSGs that endow these cells with proliferative abilities and thus cancer stem cell properties. However, it is still unclear whether CSCs originate from aberrations in normal stem cells or in progenitor cells, or more radically yet through mutations in differentiated cells that then acquire stem cell-like characteristics. Indeed, it has been recently shown that even in fully formed cancers there is plasticity in phenotype of differentiated and cancer stem cells, with some differentiated cancer cells having the ability to form CSCs in the right conditions[18]. Given the clinical findings that large numbers of CSCs are associated with shorter time to recurrence, poorer prognosis and resistance to chemo- and radiotherapy [19], it will be essential to improve our understanding of these cells in order to design better targeted therapeutics.

Sequencing Human Tumors

With the advent of next-generation sequencing technologies and the continually diminishing costs of sequencing genomes, large-scale studies involving sequencing of hundreds of tumor genomes or exomes have become increasingly attractive and feasible. The Cancer Genome Atlas (TCGA) has published several major studies sequencing several hundred human low-grade gliomas and glioblastomas, which have revealed the mutational landscapes of these

tumors. These are primarily observational studies looking for associations between certain types of cancer and specific mutations. With increasingly advanced statistical methods however, the ability of such studies to differentiate true cancer 'driver' mutations (those that confer a selective growth advantage on a tumor) from the background 'passenger' ones (mutations that do not confer a growth advantage) is continually improving. In order to prove that these putative drivers are indeed real drivers and to understand the underlying biology, it is necessary to functionally validate them in model systems, for example by inducing the same mutations in mice to determine if they generate the intended cancers. This move to model organisms from human sequencing data will also more readily enable investigation of the molecular mechanisms underpinning how genetic mutations drive cancer forward.

Gliomas

There is a diverse spectrum of brain tumors, reflecting the presence of different tissues in the central nervous system (CNS). With the exception of metastases from primary sites other than the brain, gliomas and meningiomas are the commonest types of brain tumor. Meningiomas are typically benign tumors that arise from the meningeal coverings of the brain. Other types of tumor include medulloblastomas (which are commoner in children), ependymomas, pilocytic astrocytomas and primitive neuroectodermal tumors.

Low grade gliomas (LGGs, World Health Organisation, grade II) are a heterogeneous population of intrinsic brain tumors whose natural history is to evolve to higher grade tumors. These tumors histologically contain cells with similar appearances to glial cells (neuronal support cells), including astrocytes and oligodendrocytes. LGGs constitute 15% of all adult brain tumors, and they most commonly present with seizures (in 80% of cases) [20]. A model for the natural history of gliomas posits four phases: 1) the occult stage, in which tumor-initiating cells proliferate but there is no detectable tumor on MRI; 2) the clinically silent stage, in which tumor mass becomes apparent on MRI but the patient does not have any symptoms (incidental glioma)[21]; 3) the symptomatic stage, in which the tumor elicits symptoms such as seizures or weakness; 4) malignant transformation, in which the low grade glioma switches to a more biologically aggressive high grade glioma [22, 23]. Upon malignant transformation, the tumor is termed a secondary glioblastoma, Fig 1.

Glioblastoma multiforme (GBM), or high-grade glioma, is the most common intrinsic brain tumor, and characteristically invades surrounding brain aggressively, making complete surgical resection unachievable. It tends to affect middle-aged to elderly people, and can either arise de novo (primary glioblastoma) or by transformation from an LGG (secondary glioblastoma). Its prognosis is therefore poor, with a median survival of only 14 months despite maximal therapy with surgery and chemo- or radiotherapy [24]. These survival times have not substantially changed over the last few decades despite improvements in treatment, and clearly further understanding of the biology of this cancer is needed before significant advances in treating it are made. The current chemotherapeutic standard of care for GBM is

temozolomide, a DNA alkylating agent that improves prognosis by around 2 months when given with radiotherapy [24, 25]. Given the limited impact this has on survival, there is a strong need for more molecularly targeted treatments for this cancer that will improve prognosis further. A recent example of such a therapy that has entered clinical practice for recurrent GBM is bevacizumab (Avastin): a monoclonal antibody targeting vascular endothelial growth factor (VEGF) and thus aims to reduce angiogenesis within tumors. Although bevacizumab slows down glioma growth, it does not improve overall survival in GBM; there is a benefit on progression-free survival of around 2 – 3 months in a randomised controlled study on patients with newly diagnosed GBM [26]. Nevertheless, all patients succumb to the disease after developing resistance to treatment. Surgery, although beneficial, is not curative because glioma cells tend to invade well beyond the visible margins of resection, and such cells (possibly glioma stem cells) trigger disease recurrence. A number of genes are implicated in resistance to temozolomide, most are involved in DNA repair, for example *MGMT*, *MSH1* and *MSH2*. However, much is still to be learnt in this field because tumors even in those patients without such resistance mutations become resistant to chemotherapy, suggesting other unknown mechanisms are involved.

Recent genome-wide sequencing studies of GBMs have provided insight into common genetic drivers of this tumor and have highlighted genetic differences between primary and secondary GBMs. Primary GBMs usually have one or more mutations in three main molecular pathways: Ras/RTK pathway, p53 pathway, and the Rb pathway [27, 28]. Within the Ras/RTK pathway, *EGFR* (30-50% of tumors) and *PTEN* (30%) are the most commonly mutated in GBM, although mutations in *NF1* and *RAS* have also been documented. Mutations within this pathway tend to enhance cellular proliferation. Of the p53 pathway, *TP53* (25%) itself is most commonly mutated in GBM. The *TP53* gene is normally activated following DNA damage to cells, inducing transcription of genes whose ultimate effects include apoptosis. Mutations in *TP53* are thought to have effects such as inhibition of apoptosis, stimulation of cell proliferation and neovascularisation, which are hallmarks of cancer [1]. Although a mutation in the Rb pathway is present in most GBMs, the *RB* gene itself is infrequently mutated and instead mutations in *CDKN2A* (50% of tumors) are particularly common. *CDKN2A* is the locus

for two tumor suppressor genes – *INK4A* and *P19-ARF*. *In vitro* and *in vivo* models have validated a number of such mutations as driving tumor growth and invasion. Mouse models have been particularly helpful in demonstrating how mutations in multiple pathways can co-operate together to accelerate tumorigenesis [29].

The *IDH1* (isocitrate dehydrogenase 1) mutation is characteristically found more commonly in secondary GBMs and also in LGGs [30], and although the mechanism by which this mutation contributes to carcinogenesis is still unclear it is thought to act epigenetically through abnormal methylation of DNA, Fig 1.1. Although *IDH1* is mutated in the majority of LGGs and secondary GBMs, it is still not clear how and at what stage this mutation may contribute to carcinogenesis: *IDH1* mutations (of which the R132H mutation is the most frequent) predict a better prognosis, and the mutation in conditional transgenic mice has not been found to accelerate established gliomas although there is some evidence suggesting it initiates gliomagenesis and also induces acute myeloid leukaemia in mice [31-33]. The recent pathological classification of gliomas has been changed to take into account key genetic changes, including not only *IDH1* which predicts a better prognosis, but also the presence or absence of 1p/19q co-deletions and *TERT* promoter mutations. Gliomas can thus be classified into five types depending on the presence or absence of these three genetic alterations, with *TERT* promoter mutations alone signifying the worst prognosis [34].

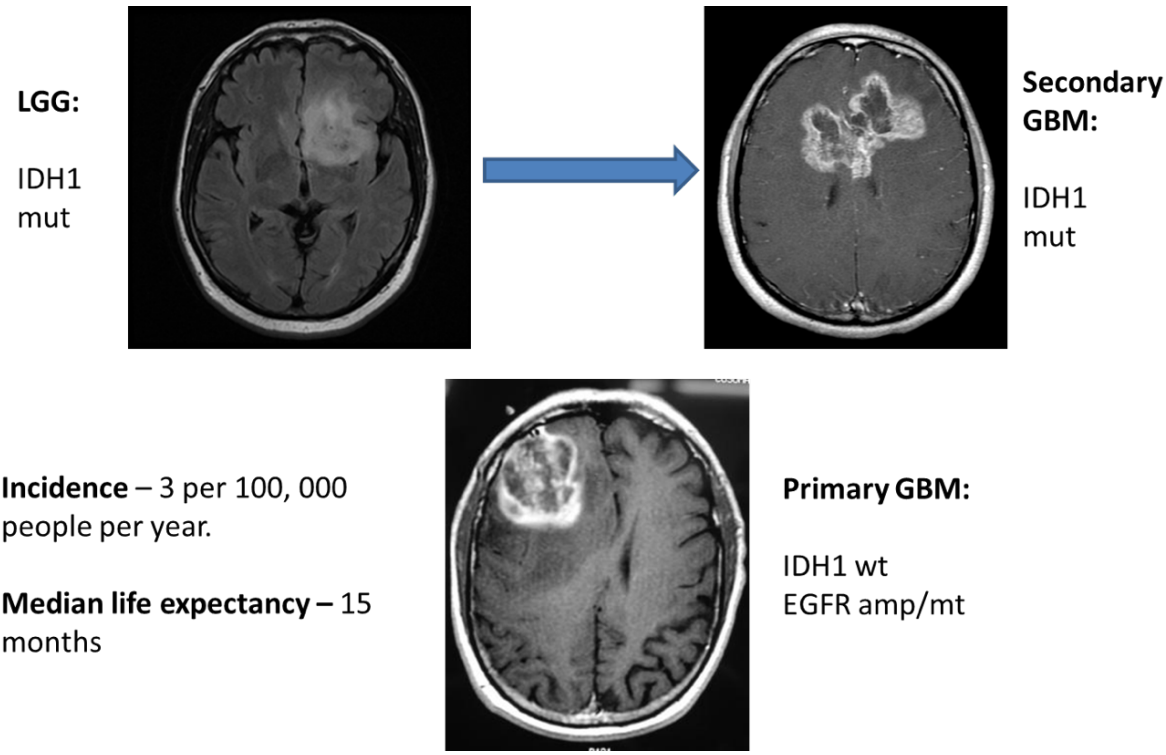


Figure 1.1. Glioma is classified into low-grade glioma (LGG) and high-grade glioma or glioblastoma (GBM). These subtypes have different clinical properties, genetics and prognoses. Only the key mutations described from the literature are illustrated here, although the reality is that these tumors have complex genomes.

Spinal Gliomas

Spinal tumors that arise from the spinal cord itself, so called intramedullary spinal cord tumors (IMSTs), are rare tumors – they make up 2% of all tumors in the CNS [35]. Of these IMSTs, spinal ependymomas are the commonest in adults, whereas spinal astrocytomas are the commonest in children and adolescents. Although these are rare tumors, spinal gliomas are a significant clinical problem – it is often difficult or not possible to achieve complete surgical resection of the tumor, and adjuvant chemotherapy and radiotherapy often only have minimal benefit but significant adverse effects; therefore, novel treatments based on improved understanding of the molecular biology and genetics of these tumors are in great need.

Spinal astrocytomas account for 60% of spinal tumors in children and adolescents. They often present with progressive back pain, often waking patients at night due to severity; due to their proximity to the spinal cord, expansion of these tumors gives symptoms of numbness or tingling as well as motor weakness, which can be misconstrued as clumsiness in children [36]. Most spinal astrocytomas are low grade (grade I or II), although around 20% are high grade (III or IV) and these have a poor prognosis associated with increased invasiveness of the tumor: the mean survival is 15.5 months [37]. It is typically difficult to achieve complete surgical resection of spinal astrocytomas because these tumors have cells that infiltrate into the spinal cord, and the surgeon is unable to resect beyond the margin of the spinal cord for concerns over inducing very disabling paralysis. Nevertheless, leaving behind groups of tumor cells in the spinal cord is likely to lead to recurrence of the tumor, which is often the case. Following recurrence, the treatment options are limited – radiotherapy is controversial for these tumors in children as it can cause radiation necrosis, radiation myelopathy and reduced spinal growth [38]; chemotherapy is therefore favoured but itself is of limited benefit. Given that temozolomide has been showed to provide a limited survival benefit for brain gliomas in adult patients, it can be used as a chemotherapeutic agent in spinal gliomas but has associated systemic toxicities such as neutropenia and lymphopenia [39].

The genetic basis of spinal astrocytomas is poorly understood. The main reason behind this, apart from the rarity of the tumor, is the lack of fresh tissue available for performing next generation sequencing. As explained, it is challenging to achieve complete resection of the tumor due to its dangerous location, so the amount of tumor resected is limited and used for pathological purposes only. Further efforts are needed to prospectively collect suitable material for deep sequencing to yield some insight into the biology of spinal astrocytomas; appropriate mouse models are also required to complement observations in genetic sequencing of these tumors in patients with functional validation and molecular characterisation.

In vitro Models of Cancer

An alternative model for studying cancer is *in vitro* cancer cell lines, which are available for human glioblastomas. Such cell lines have advantages in that they are generally derived from patient tumors, they proliferate well *in vitro* and can therefore be expanded relatively quickly, they are amenable to genetic manipulation such as knockout of individual genes, and the efficacy of drugs as potential therapeutic agents can be tested on these lines. However, there are also several disadvantages for using cell lines as models of human cancer in comparison to using mice for *in vivo* study. It is known that human GBM cell lines acquire multiple additional genetic lesions in culture that were not originally present in the parental tumor, including further chromosomal copies and translocations in addition to mutations. In this way, the genetics of these cell lines do not necessarily reflect that of the patient tumor, which can lead to false conclusions drawn from genetic experiments performed in such cell lines alone. One method to help overcome this problem is to culture cells derived from patient GBMs in neural stem cell conditions either as 'gliomaspheres' or as adherent cell lines, which has been shown to help maintain the original tumor genetics [40], at least in short-term culture.

Another limitation of cancer cell lines is that they do not model more complex features of carcinogenesis such as the tumor microenvironment. As discussed previously, rather than simply being collections of cancer cells, malignant tumors contain complex interactions between cancer cells and other cells such as macrophages and T cells, and these interactions are very difficult to model accurately *in vitro*. In comparison, the use of transgenic mice that develop cancer has the potential to model such interactions between the tumor and its microenvironment. Other features of cancer, including invasion, angiogenesis and metastasis, are also more readily demonstrable *in vivo* than *in vitro*. Furthermore, cancer cell lines are derived from fully formed tumors which are typically late in cancer evolution; this precludes study of the processes initiating tumor development that may differ greatly from those driving progression. Mouse models on the other hand have the potential for investigation of both early and late evolutionary forces in cancer.

The Mouse as a Model Organism

There are several reasons why the mouse is a useful model organism for studying human diseases, particularly with genetic approaches. Firstly, being a mammal the mouse shares many similarities with humans in its organ systems, allowing a diverse range of diseases affecting different organs to be studied. In particular, the anatomy of its central nervous system is not too dissimilar from humans, containing major features of the brain and spinal cord that are found in man, making the investigation of certain neurological disorders possible in mice. Mice have relatively short breeding times, with pregnancy only lasting around two weeks, and typical litters contain 5 – 10 mice, which allow for large numbers of mice to be generated within a short period of time. Given their small size, it is also practical to house even up to thousands of mice in a single room. These latter features are particularly important for testing whether experimental results gained from a few mice can be replicated in larger cohorts and of course for performing large-scale *in vivo* genetic screens.

In order to carry out genetic studies in a particular organism, its genome must be mapped and well-understood. The mouse genome was fully sequenced and published not long after that of the human[41]. This revealed a great deal of homology in gene sequences between mouse and human genomes, a crucial feature for enabling investigation of human genetic diseases in mice.

Reverse Genetics in Mice

Reverse genetics involves inserting a known mutation into mice (or any other model organism) and observing the induced phenotype. In the context of cancer genes, this entails selecting a gene that has been associated with a specific cancer, for example from a cancer sequencing study, and then determining whether it can produce the intended cancer phenotype in mice. Of course, this ideally (but not necessarily) requires some hypothesis about the gene itself. Examples of this technique are that introducing *Tp53* or *Pten* mutations in mice lead to generation of cancers, as will be discussed in depth later [42, 43].

Forward Genetics in Mice

Forward genetic screening aims to identify gene mutations that give rise to a specific phenotype, such as cancer. In this approach, many different mutants are generated (that are not previously known to give rise to the phenotype of interest) and then these are observed for development of the phenotype. The mutants that generate such a phenotype are then further evaluated to identify the underlying mutations. This can be a powerful approach for cancer gene discovery, for example, and can be effectively applied *in vivo* in mice. In the case of cancer, there is often a multitude of genetic and epigenetic aberrations and intratumoral heterogeneity, complicating efforts for identifying the true drivers. This problem is particularly true for human GBMs. Forward genetic screening in mice can provide a complementary approach to sequencing human tumors for the identification of these true cancer driver genes, which may be mutated or epigenetically altered in human tumors.

Inbred Mice

Genetic diversity is introduced during gamete formation in mice by recombination of chromosomes in meiosis, similar to most other sexually reproducing organisms. This can lead to confounding effects in experiments performed with different mouse strains. Therefore, pure-bred mouse strains that are homozygous at all genomic loci have been invaluable for genetic experiments. These are generated by many generations of sibling matings: 98.7% of the genome is homozygous after 20 generations of these matings [44]. In order to study a specific mutation from a different genetic background, a mouse carrying the mutation of interest can be crossed with an inbred mouse, forming a congenic strain with the mutation of interest on an invariant genetic background that mitigates confounding effects of comparing results gathered from different strains of mice. This also helps ensure that the introduced mutations are the direct cause of any phenotype observed, aiding in establishing putative cancer genes as true oncogenes or tumor suppressor genes.

Different forward genetic approaches

There are several different ways for conducting forward genetic screens in mice *in vivo*, and the most important of these will be briefly discussed here. Mutagenesis can either be conducted in the germline or somatically in tissues of interest.

Irradiation

Irradiation with gamma rays introduces double strand breaks into DNA, and these can be inaccurately repaired leading to chromosomal imbalances such as deletions that effectively knockout a gene. This makes irradiation potentially useful for forward genetic screening. However, deletions caused by irradiation can be large enough to span many genes, making identification of the gene causing the phenotype more difficult and requiring detailed sequencing approaches such as whole-exome or whole-genome sequencing, and extensive validation of identified targets to confirm their true nature as genes causing the phenotype.

ENU Mutagenesis

ENU (N-ethyl-N-nitrosourea) is a very potent mutagen, acting as a DNA alkylating agent. It can induce point mutations in genes, and can therefore be used in genome wide screens by sequencing for point mutations in tissue with a phenotype of interest, such as cancer. However, given the high mutagenic rate of ENU, it can be more difficult to identify the underlying driver genes from this approach than with others such as transposons, requiring deep sequencing and more complex bioinformatics analyses.

Insertional Mutagenesis

Insertional mutagens are mobile DNA elements that insert into the genome to act as a mutagen. These have the advantage that they effectively tag the mutated locus with their

known sequence, and linker-based PCR methods can amplify the adjacent genomic regions. Insertional mutagenesis within an exon (or even just within the gene) is likely to lead to disruption of the function of the gene.

Viral Mutagenesis

Viruses can be used as cellular mutagens for cancer screening in mice. Insertion of a proviral retrovirus into the mouse genome can lead to increased expression or disruption of a gene depending on the location of the insertion within a gene. The proviral genome contains long terminal repeats (LTRs) to control transcription not only of the retrovirus but also of the host cell genes if they are downstream of these viral LTRs. Sequencing of resulting tumors formed from insertion of the provirus into cancer genes can help identify these underlying genes by determining common insertion sites of the virus [45]. An example of such a transforming virus that can be used in screens is the murine leukaemia virus (MLV). A disadvantage of this approach is that viral mutagens have insertional biases that prevent this being a truly unbiased genome-wide screening approach, and the more time-consuming and challenging nature of insertion site cloning and mapping for viruses makes this approach less favourable than others such as transposons. Retroviral screens have led to early discoveries of genes such as Myc and some of its cooperative partners in cancer [46, 47].

Sleeping Beauty Transposons

Sleeping beauty (SB) transposons are from the Tc1-Mariner family and are naturally found in salmonid fish. This transposon was reconstructed from phylogenetic data and the investigators thus gave the name '*Sleeping beauty*' to reflect that fact that it was 'awakened from a long evolutionary sleep.' Ivics et al discovered the sequence of the ancestral transposon from this class of fish, demonstrated to be two 250bp terminal DNA sequences containing inverted repeats that flank an open reading frame that codes for the transposase enzyme [48]. The entire sequence is 1.6kbp. *Sleeping beauty* is therefore a two-component

system composed of the transposon vector and the transposase enzyme. When these are present in the same cell, the transposase recognises the inverted repeats of the transposon and excises it from the donor locus. The transposon can then insert itself at a TA dinucleotide region elsewhere in the genome. In this way, the transposase catalyses a 'cut and paste' reaction of the transposon. Modifications to the inverted repeats were made early on in order to improve transposition efficiency [49]. Site-directed mutagenesis of the SB transposase produced alternative versions of the enzyme with different transposition efficiencies [50, 51]. The first SB transposase was SB10, and modified versions were numerically labelled (SB11 and so on).

A consequence of SB transposition is the creation of a 'footprint' mutation. The reason for this is that when SB inserts into genome it duplicates its target site (a TA dinucleotide) and excision generates 3-nucleotide overhangs. A potential advantage of this footprint mutation is that insertions of likely biological relevance can still be identified through genetic sequencing even if the transposon has subsequently mobilised into another site. On the other hand, a disadvantage of a footprint mutation is that some gene disruption may remain in the transposon insertion site, which may not be of relevance for the phenotype of interest. Another feature of the SB transposon is its tendency for so-called 'local hopping' wherein SB preferentially inserts into DNA near to the original site of the transposon, typically within 2 – 10 Mbp on the same chromosome. This bias in its insertion sites must be considered when analysing the data from a SB forward genetic screen, in practice meaning that investigators usually exclude all common insertion sites in the transposon donor locus.

The first sleeping beauty screens in mice for cancer used a constitutively active SB transposase and a transposon line (T2/Onc or T2/Onc2), in which the transposon was mobilising in all tissues [52]. This transposon can induce a gain-of-function or a loss-of-function of a gene, depending on where in the gene it inserts. T2/Onc contains a murine stem cell virus (MSCV) long terminal repeat (LTR) promoter with artificial exon with splice donor (SD). The LTR and SD can lead to fusion transcripts through splicing, and if T2/Onc has inserted upstream in a gene these fusion transcripts will be overexpressed because of the T2/Onc promoter. This is equivalent to activation of proto-oncogenes. The transposon also contains splice acceptors

and a bidirectional polyadenylation signal to cause termination of transcripts that arise when the transposon inserts into an intron of a gene. This is important for inactivating tumor suppressor genes and thus promoting tumorigenesis.

Several SB insertional mutagenesis screens have been conducted in mice, successfully contributing to driver gene discovery for many different cancers [53-63]. One constitutive SB screen in mice generated gliomas in the brain, although the incidence was low [64, 65]. To increase the incidence of these tumors, the authors crossed the mice with a *P19Arf* allele which is a known tumor suppressor gene in gliomas. Although this did indeed increase the incidence of brain tumors, the difference was small. A mixture of anaplastic astrocytomas (grade III) and glioblastomas (grade IV) were produced. Sequencing of the resulting 21 gliomas from this screen yielded 887 common integration sites (CIS), and identified *Csf1* as a recurrently hit gene. Immunohistochemistry in human tumors demonstrated overexpression of CSF1 in high-grade astrocytomas, providing some support for a role of this gene in supporting malignant glioma formation in humans as well as mice. Importantly however, the CIS genes did not include a number of well-established glioma genes such as *Egfr*, *Pdgfr* and *Tp53*, and there were only single insertions found in other important tumor-specific genes including *Pten* and *Akt*. This may be explained by certain insertion site preference biases of SB or that SB-induced gliomas represent only a subset of gliomas that is not driven by the major cancer drivers in most human gliomas. It is therefore crucial that transposon-driven gliomas are studied in other genetically-predisposed backgrounds in order to gain a more complete understanding of the cancer drivers, ideally with an alternative transposon system such as *piggyBac* as well. A systematic comparison of the insertion sites from their study and human glioma sequencing data was not performed.

To enable screens to be performed for tissue-specific cancers, a conditional SB transposase allele was also developed that is active in the presence of cre expression and has been used for screens of many cancer types in mice. One study which employed this conditional SB transposon system for investigating gliomas used a nestin-cre allele on a *Trp53*-mutant background to drive expression of the SB transposase in mouse neural stem cells, although these did not directly generate tumors *in vivo*. *In vitro* culturing of embryonic neural stem

cells derived from the subventricular zone of these mice demonstrated that these cells can be immortalised by mobilisation of SB: immortalisation occurred significantly more frequently in cell lines with both Trp*53*^{R172H} and mobilising SB than those with *Trp53*^{R172H} alone, and not at all in lines with neither *Trp53*^{R172H} nor mobilising SB. When these immortalised cells were subcutaneously transplanted into SCID mice, they generated tumors with a latency of two – four months. The authors identified 106 CIS genes in the immortalised cell lines and 114 in the tumors, of which 34 CIS genes were present in both cohorts. Comparing the CIS from the immortalised cells with those of the tumors in mice showed that a further round of transposon mobilisation for *in vivo* tumor establishment was needed in addition to the insertions present in immortalised cells alone [66]. The authors therefore were able to categorise SB insertions according to whether they drove cellular proliferation *in vitro* or *in vivo* tumor growth or both, as part of a two-step process of cancer evolution. Amongst the CIS in the immortalised cell lines were a few known glioma genes, including *Pten*, and similarly amongst the tumor CIS were genes such as *Pdgfrb* and *Nf1*. The study also identified cancer genes that were not previously linked with gliomas, such as *Met* and *Klf3* which were amongst their top-ranking tumor CIS. The CIS genes clustered into biological pathways that are thought to underlie gliomagenesis, in particular the Ras-MAPK and PI3K-Akt pathways, confirming that these major pathways that are mutated in human tumors can promote neural stem cell immortalisation *in vitro* and subcutaneous glioma formation. A systematic comparison with genetic and epigenetic alterations in human gliomas and glioma cell lines was not performed however, and neither was there an analysis for network interactions between the CIS genes.

More recently, work from the Bradley group has demonstrated the usefulness of a single-copy sleeping beauty transposon for cancer forward genetic screens in mice. The study employed a transposon linked with the *Pten* gene, such that mobilisation of this single copy transposon also led to heterozygous loss of *Pten* (an important cancer driver). The resulting developed tumors in multiple organs, particularly in the prostate. The advantage of this approach is that a single cell will simultaneously have transposition and *Pten* loss, helpful in establishing cooperativity between *Pten* and transposon insertion genes in tumorigenesis. Additionally, the study found that the common insertion sites from this screen gave the

strongest candidate driver genes from an otherwise longer list of genes produced from a multiple-copy transposon screen [67].

PiggyBac Transposon

The *piggyBac* (PB) transposon is naturally found in the cabbage-looper moth. *PiggyBac* has some key advantages over SB as a transposon: it has a larger cargo capacity than SB, and also can mobilise within the genome without leaving a footprint unlike SB. Given these advantages, efforts were made to adapt PB for mammalian systems [68], and specifically for forward genetic screens of cancer both *in vitro* and *in vivo*. There are several mouse PB transgenic lines, including both high-copy and low-copy PB number. The low-copy number PB transposon mice were demonstrated to be better for modelling solid cancers, since high copy mutagenesis has a tendency for embryonic lethality.

A constitutive PB screen in mice generated a variety of solid and haematological cancers [69]. In this work, there were a number of different transposon lines: it was demonstrated that lines with a transposon driven by the CAG promoter (ATP1) were more efficient at generating solid tumors, likely because of superior transposon mobilisation in the underlying tissues. Conversely, lines with the transposon under control of the MSCV promoter (ATP2) were better at producing leukaemias and lymphomas. With these models, the authors were able to identify known and novel cancer genes from the sequenced transposon common insertion sites. To enable screening in specific tissues, a conditional *piggyBac* transposase mouse line was engineered; a screen on pancreatic cancers employed this conditional *piggyBac* system, identifying new cancer genes such as *Foxp2* in the specific tissue of interest [70]. This approach for cancer screening is discussed in further detail in the relevant chapter of this thesis.

CRISPR / cas9 screening

Technologies for genetic manipulation of mammalian genomes based on engineered nucleases have evolved rapidly over the last few years. Zinc finger nucleases (ZFNs) were the first to be widely used in mammalian cells [71]. Their use established the basic paradigm for using double strand DNA breaks in the genome to stimulate error-prone or template targeted repair mechanisms to effect a genetic change. The use of ZFNs for this purpose quickly declined with the development of transcription-activator like effector nucleases (TALENs) primarily because the reliability of the nucleotide recognition encoded by the TAL repeats supported reliable synthesis of highly specific molecules [72]. The use of TALENs for mammalian genome engineering was also relatively short lived, as these were rapidly superseded by Clustered regularly interspaced short palindromic repeats-Cas9 (CRISPR-Cas9) because of the simplicity of deploying the cas9 nuclease to any nucleotide in highly complex genomes.

CRISPRs were identified in *E.coli* in 1987 and in other bacteria and archaea a decade later [73, 74]. The phage origin of these repeats and the identification of genes with putative nucleases associated with these repeats (CRISPR-associated) *cas*-genes led to the hypothesis and subsequent demonstration that the CRISPR-cas system had a role in microbial adaptive immunity [75]. This is achieved by directing the cas-nuclease to the incoming phage DNA by a guide RNA transcribed from the clustered repeats [76].

In contrast to the ZFN and TALEN systems in which specificity is achieved by complex protein-nucleic acid interactions, the cas-nuclease is directed to a genetic target by nucleic acid base pairing determined by a unique 20 nucleotide region of short guide RNA (sgRNA). Experimentally this sgRNA sequence can be adjusted to guide the nuclease to virtually any site in a complex genome [77]. The efficiency of the cas-nuclease coupled with the simplicity with which it can be directed has resulted in its rapid adoption. The CRISPR-cas9 system has been shown to be effective for manipulating genes in a variety of cell types from different organisms. When used as a nuclease, cleaved DNA is re-joined by an error-prone end-joining process resulting in small insertions and deletions (‘indels’) at the target site and concomitant

loss of the gene's function. Larger genetic alterations such as deletions and inversions can also be generated. In other applications the break generated by the nuclease will catalyse a process of homology directed repair if a suitable vector is also provided resulting in replacement of one sequence (for instance a defective copy) with a normal one provided by the vector, so called gene-editing, Fig 1.2. Studies have provided a cautionary note of potential off-target effects (unintended modifications at other sites in the genome) with this platform [78].

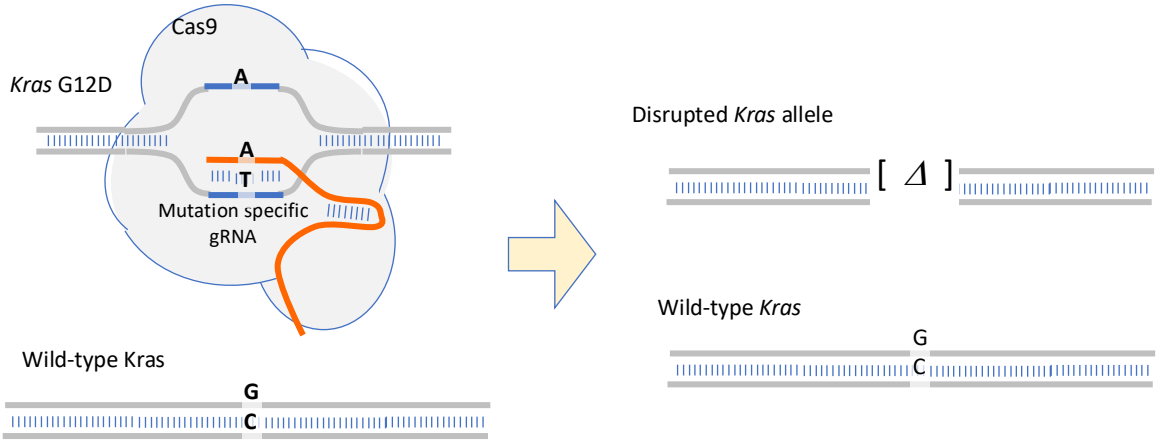


Figure 1.2. CRISPR/cas9 mediated introduction of DNA indels. Example is given of the Kras G12D oncogene which can be targeted with an sgRNA that leaved the wild-type allele untouched.

Over the last few years, a number of studies have demonstrated the usefulness of CRISPR-cas9 in both positive and negative selection genome-wide screens, including in human cancer cells [79] [80-83]. Such screens employ large lentiviral libraries with multiple sgRNAs per gene, and consequently require large starting populations of cells.

Glioma Mouse Models

The majority of glioma mouse models have employed cre / LoxP technology for specifically targeting cancer genes in certain neural tissues of interest. I will therefore describe this technology before discussing examples of glioma mouse models in more detail.

Cre / LoxP, Flp/FRT, RCAS Technology

Site-specific recombination allows for the generation of genetic alterations such as deletions, point mutations, duplications and inversions. The flippase / flippase recognition target (Flp/FRT) system was the first one to achieve site-specific recombination in multicellular organisms, and this was originally performed in Drosophila [84]; in this system, the flp recombinase mediates recombination between FRT sites in the genome. In the mouse however, the commonest method for recombination is the use of the cre/LoxP system in which the cre (cyclization recombinase) mediates recombination between two LoxP sites. The LoxP sites are 34-base pair consensus sequences, each with a central 8-bp core spacer sequence that determines the orientation of the LoxP site, and two inverted 13-bp flanking sequences that bind cre. The cre/LoxP system was first implemented in mice in the early 1990s, and since then has been widely used for generating conditional genetic alterations *in vivo* in a variety of specific tissues, including the brain. Indeed, many cre transgenic mouse lines have been created in the last few decades to allow study of organ or tissue-specific physiology and pathology [85].

An alternative system for introducing targeted mutations to cells of interest is the RCAS (replication-competent ASLV-long terminal repeat with a splice acceptor) vector system. These vectors derive from the Rous sarcoma virus, which belongs to the avian sarcoma – leukosis virus (ASLV) family, and they contain the src (oncogene) splice site and express an inserted gene (such as an oncogene) via a spliced message. This system is limited by the small size of the insert (2.5 kb) and the low number of cells that are typically infected and express the gene of interest [86].

Nestin is an intermediate filament protein expressed in neural stem cells and neural progenitors. Mice with the nestin-cre allele express cre from embryonic day 13, at which stage the embryonic neural progenitors are able to undergo differentiation into many cell types including astrocytes, neurons and oligodendrocytes. Therefore, in postnatal and adult mice containing nestin-cre, cre is expressed throughout most of the central nervous system, eye and also the kidneys – this was demonstrated by Dubois and colleagues who showed virtually complete cre-mediated recombination in these tissues by embryonic day 15.5 using LacZ based reporters [87, 88]. The allele is however insufficient for driving recombination in early embryonic ventricular zone neural progenitors and neural stem cells (before embryonic day 17.5), as determined using multiple cre-dependent reporters. [89]. An alternative cre line that is frequently used in glioma mouse models is hGFAP-cre, which is also expressed from pre-natal stages and in the majority of cell types in the brain and spinal cord [90]. For studies where the timing of recombination needs to be controlled, it is possible to do so through tamoxifen injections using nestin-creERT2 and GFAP-creERT2 mouse lines, which are inducible cre lines. These are useful for studying tumor origins from adult brain cells as opposed to embryonic cells, for example; the cre expression onset is controlled by specifying age of the mice at which tamoxifen is given. In order to induce recombination in more specific groups of brain cells, alternative cre lines can be used. For example, Olig2-cre allows site-specific recombination in oligodendrocyte precursor cells and oligodendrocytes, Syn-cre gives recombination specifically in neurons, and Glast-cre is a newer alternative line for recombination in neural stem cells in the SVZ. However, an important drawback of all of these cre lines is the specificity of the regions and cell types in which recombination occurs, in that there is typically recombination in other cells than those of interest.

Key Glioma Mouse Models

One of the earliest oncogenes to be discovered in gliomas is the epidermal growth factor receptor (*EGFR*) gene [91], which is mutated and/or amplified in 50 -60% of primary glioblastomas. EGFR is a cell-surface receptor that binds epidermal growth factor as its ligand

and then signals via intracellular cascades, including the Ras-MAPK and PI3K-Akt pathways. In primary glioblastomas, the variant III mutation of *EGFR* is particularly common, and involves deletion of exons 2 – 7 of the gene (the extracellular ligand binding domain) leading to constitutive signalling the resulting receptor. An early study aimed at determining whether excessive EGFR signalling can induce gliomas *in vivo* employed the RCAS vector system to introduce an *Egfr* activating mutation (the *EgfrvIII* deletion and another deletion that removes the intracellular regulatory kinase domain) in mice expressing the avian tumor virus receptor A (TVA) under brain cell specific promoters. The vector was introduced into the frontal lobes and hippocampus. After 15 weeks, none of the mice developed gliomas. In contrast, when an *Egfr* activating mutation was introduced in the presence of *Cdkn2a* loss, gliomas arose at a high frequency particularly on the nestin-TVA (Ntv) background. The authors concluded that *Egfr* activating mutations alone are insufficient to generate gliomas, but can cooperate with predisposing mutations such as those of *Cdkn2a* to produce these tumors [92]. Given the incidence of tumors was higher in Ntv compared with glial-specific GFAP-TVA (Gtv) mice, they suggested that the presence of these mutations in a neural stem cell lineage is a likely origin for gliomas.

Another early glioma mouse model that used the RCAS vector system was that by Holland et al in 2000 [29]. In this study, a $Kras^{G12D}$ mutated gene and a constitutively active *Akt* mutant were virally transferred into the brain of mice using RCAS vectors. Each of these genes was insufficient to induce gliomas when expressed alone; however, when they were expressed in combination with each other, lesions similar to human glioblastomas were produced. Although previously it was thought that neither *Kras* or *Akt* mutations are found in human GBMs, recent large-scale sequencing efforts of human tumors have demonstrated that *Kras* is a likely genetic driver of these cancers albeit at a low frequency [93]. Moreover, the authors demonstrated elevated Ras pathway activation in all GBMs they tested and increased Akt protein phosphorylation in the majority of GBMs, suggesting that upstream mutations are likely to lead to activation of these pathways.

Given that *TP53* is mutated in ~30% of GBMs and *PTEN* in around ~40%, Zheng et al hypothesised that these two mutations cooperate with each other in gliomagenesis[94]. They

crossed a hGFAP-cre mouse with *Trp53* mutant and *Pten* knockout mice, which led to grade III and grade IV gliomas at a median latency of approximately 7 months [94]. Gliomaspheres with stem cell like properties could be generated from these tumors, and the authors demonstrated that activation of myc was crucial in driving tumorigenesis in this model. Importantly, although *TP53* and *PTEN* mutations are commonly found in low grade gliomas as well, all of the tumors generated in these mice were either grade III or IV. Work from Luis Parada's laboratory supports these findings and also demonstrate cooperation of *Trp53* and *Pten* with *Nf1* in mice [95].

Zhu et al explored the cooperation between the *EGFRvIII* mutation and other genes in gliomagenesis by using transgenic mice [96]. They generated both an *EGFRvIII* transgenic mouse, in which the mutation was overexpressed at the *Col1a* locus, and also an *EGFRwt* transgenic mouse with the human gene sequence inserted and over-expressed at the *Col1a* locus. These mice were conditional and required injections of cre into the brain for the mutations to be expressed. Cre was injected into the basal ganglia (striatum) of adult mice. Neither of these mutations was sufficient to induce gliomas alone; but when expressed in combination with homozygous loss of *Pten* and *Ink4a*, both mutations were able to produce high-grade gliomas with a short latency. However, the *EGFRwt* allele produced tumors with a low incidence and long latency in comparison with a single *EGFRvIII* allele that dramatically enhanced the incidence and reduced the latency of tumor formation. Homozygous *EGFRvIII* was more efficient in producing tumors than heterozygous *EGFRvIII*, although the difference was rather small.

Glioma Cell of Origin

An important question in glioma biology is which cell type gives rise to the tumor. This is a well-studied yet still controversial topic, and is thus worth giving some consideration to here. Although it is unclear which is the key cell type of origin, it appears that the combination of genetic alterations affects whether one particular cell type can give rise to a glioma. Jacques et al introduced combinations of *Trp53* / *Pten* and *Rb* mutations in adult subventricular zone (SVZ) neural stem cells (NSCs) and in astrocytes of mice [97]. Only SVZ stem cells produced

tumors, whereas introducing these mutations into cultures astrocytes did not; moreover, *Trp53* and *Pten* mutations together induced gliomas, whereas deletion of *Rb* in addition to *Trp53* / *Pten* led to primitive neuroectodermal tumors (PNETs). Importantly, despite containing the same mutations as those induced in the SVZ, mature astrocytes were unable to form tumors.

A related study into the cellular origin of gliomas investigated the role of *Egfr* (activation) and *Cdkn2a* (loss) mutations in different brain cell types [98]. These mutations were introduced into cultured mouse astrocytes and neural stem cells, which were then transplanted into the brain (striatum) of SCID mice. If these mutations were introduced independently of each other, the cells were unable to induce gliomas. In combination however, they led to the formation of gliomas from both astrocytes and neural stem cells, suggesting that the combination of mutations rather than the cell type was more important in driving tumor formation. *Cdkn2a* loss led to dedifferentiation of the astrocytes, which allowed the cells to later be transformed if an activating *Egfr* mutation was introduced. The authors concluded that loss of *Cdkn2a* was a critical initial step in gliomagenesis that must precede *Egfr* activation if the latter is to trigger glioma formation.

To expand on these observations, Friedmann-Morvinski and colleagues used performed lentiviral vector injections to cause p53/Nf1 knockdown or H-ras expression with p53 knockdown in neurons, astrocytes and NSCs of mice. They found that all of these cell types generated malignant gliomas in mice with these genetic alterations, and concluded that most CNS cell types undergo dedifferentiation in response to defined oncogenic mutations to NSCs or progenitors, enabling tumor initiation and maintenance [99]. Although this demonstrated these differentiated cell types can give rise to GBMs *in vivo*, this does not necessarily establish which cell type is the most likely origin.

Another group investigated whether a particular cell type in the SVZ was particularly responsive to EGF; they demonstrated in mice that infusion of EGF into the lateral ventricles caused increased proliferation of C cells (transit amplifying progenitor cells that express nestin) in the SVZ, and these cells then invaded the brain parenchyma. Although no tumors

occurred in this model, the study demonstrated that exogenous EGF can increase proliferation of neural stem cells through the wild-type *Egfr* activation [100]. It is unclear from this study alone though what the effect of the *EGFRvIII* mutation in absence of exogenous EGF would be on these cells.

Another study elegantly used mosaic analysis with double markers (MADM) in mice with *p53*/*Nf1* inactivation in NSCs. Prior to GBM establishment, MADM-based lineage tracing identified aberrant growth only in oligodendrocyte precursor cells (OPCs), but not in NSCs or other NSC-lineages. Moreover, induction of *p53*/*Nf1* mutations directly in OPCs caused glioma formation, leading the authors to conclude that OPCs are the likely origin of glioma, even if the initiating mutations occur in NSCs [101], Fig 1.3.

Very recent work using sequencing data from human patients lends support to the subventricular zone being the origin of at least some GBMs [102] – the investigators performed deep sequencing of triple matched tissues from *IDH*-wild type GBM patients, including normal SVZ, tumor tissue and normal cerebral cortex. They found that normal SVZ in 56.3% of cases contained low level driver mutations (1% of tumor mutational burden) that were also present in the GBMs; introduction of driver mutations in astrocyte-like NSCs in the SVZ in mice led to migration of these cells and formation of GBMs at distant brain regions. This evidence supports these cells as being potential origins of GBMs.

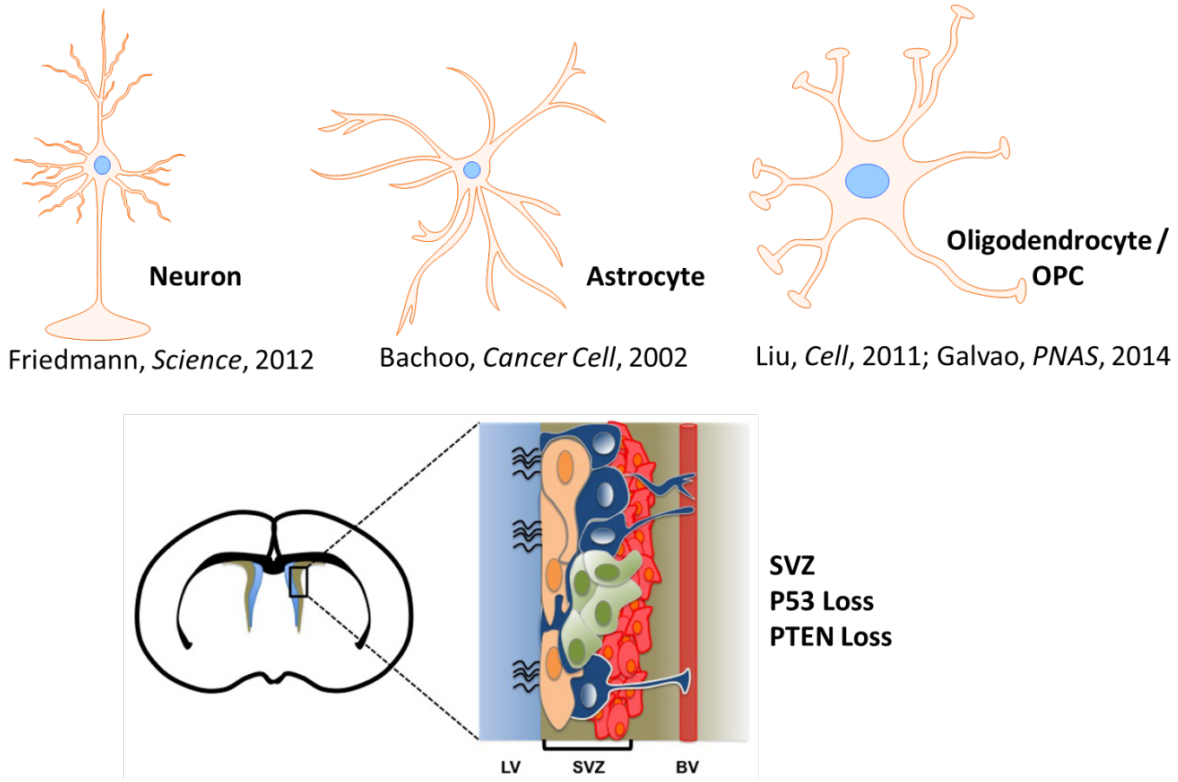


Figure 1.3. Potential sites of origin for gliomas as demonstrated in mouse models, as demonstrated in various studies suggesting these tumors may arise from neurons, astrocytes, oligodendrocyte precursor cells (OPCs) or subventricular zone (SVZ) neural stem cells [98, 99, 101, 103].

Spinal Glioma Mouse Model

Spinal gliomas cause significant morbidity such as limb paralysis, and the prognosis associated with these tumors is poor. There are very few animal models of spinal gliomas and their molecular pathology is poorly understood. Hitoshi et al used transgenic mice expressing *Pdgfβ* under the GFAP promoter using a tetracycline responsive element (TRE); they developed several mouse lines and selected a line with the highest expression of *Pdgfβ* in the spinal cord instead of the brain [104]. With this model, they demonstrated that mice developed spinal gliomas with a high incidence, and these tumors reflected a spectrum from oligodendrogliomas to astrocytomas. Loss of one copy of *Trp53* in addition to expression of

Pdgfβ led to acceleration in the time taken to develop spinal tumors; the incidence of brain gliomas is this model was relatively low (< 5%), likely due to the lower expression of *Pdgfβ* in the brain. This model provides an indication that gliomas can arise in the spinal cord with the same genetic aberration that can be used to generate brain gliomas. A strength of this model is the use of transgenic mice instead of RCAS vectors (a popular method amongst early glioma models), in which it is not possible to exclude a role for insertional mutagenesis in tumor formation and in which it is more difficult to target less accessible regions such as the spinal cord.

Summary

In summary, I have discussed the critical processes behind cancer development, introduced brain tumors and in particular gliomas, and how these tumors may be modelled and studied in mice. This will form the basis for understanding the experiments performed as part of this PhD thesis. The key aims of this Thesis, as will be described in the relevant Chapters, are to establish the role of *EGFR* in glioma initiation, and to map the cooperative mutational and functional genomic landscapes of gliomas in mice. Such knowledge will be important for deciphering complex human glioma genomes and potentially for developing new biomarkers and therapeutics.

Chapter Two: Materials and Methods

Mouse Breeding Strategies

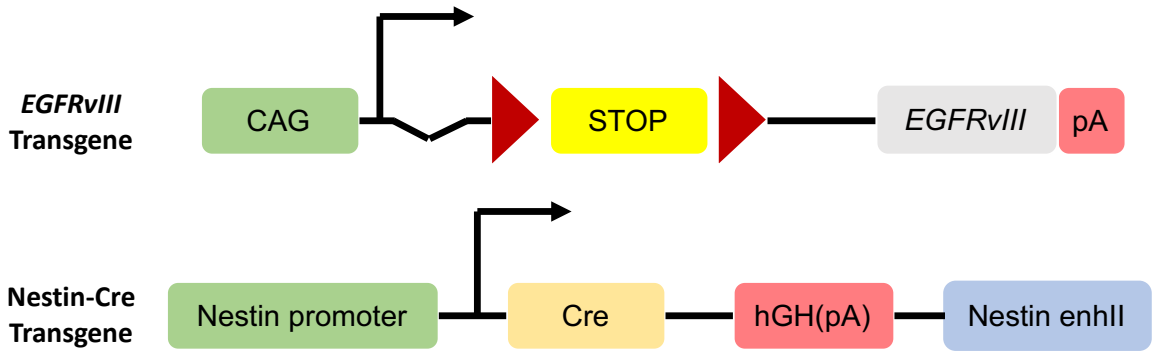
The breeding strategies used in my PhD typically involved multiple crosses between mice, in particular for the *piggyBac* transposon screens, with carefully considered breeding strategies. These strategies are described here.

EGFRvIII; *nes*-cre mice

EGFRvIII conditional mice (containing an integrated human *EGFRvIII* transgene in the mouse *Col1a1* locus, mouse chromosome 11) were acquired from the National Cancer Institute (NCI) and *nestin*-cre (*nes*-cre) mice from Jackson Laboratories, having been previously produced by other groups [87, 96], Fig 2.1. *EGFRvIII*-mice were crossed with each other in order to generate mice homozygous for this allele; similarly, nestin-cre mice were crossed with each other. *EGFRvIII*/*EGFRvIII* were then crossed with *nes*-cre/*nes*-cre mice to produce *EGFRvIII*/+; *nes*-cre/+ mice. These were placed on tumor watch from age 6 weeks, observing for signs of neurological disease caused by tumors of the central nervous system such as seizures, limb weakness, abnormal gait, incoordination, macrocephaly; and more general signs of illness such as piloerection, lethargy, and weight loss (see later section in Materials and Methods).

The strains of the original mice are as follows: *EGFRvIII* mice are FVB, *nes*-cre mice are C57BL/6J; the ATP1S2 and TSPB mice are C57BL/6J albino. Therefore, the final mouse cohorts were of a mixed background, with a predominance of C57BL/6J genetic background.

A



B

Figure 2.1. *EGFRvIII* was conditionally expressed in the central nervous system using *nes*-cre. A. Structures of *EGFRvIII [96]* and *nes*-cre [87] alleles. Notation: CAG = Cytomegalovirus (CMV) early enhancer, chicken β-actin promoter; pA = poly-adenylation signal; red triangle = *loxP* site; hGH(pA) = human growth hormone polyadenylation signal; nestin enhII = enhancer in second intron of rat nestin gene [87]. The EGFRvIII transgene is inserted into the mouse *Col1a1* locus on chromosome 11. Upon cre expression, the floxed stop cassette is excised, leading to expression of *EGFRvIII* driven by the CAG promoter from the Col1a1 locus. B. Expression of cre demonstrated in a conditional LacZ reporter mouse carrying the *nes*-cre allele at embryonic day 15, photograph provided by the Jackson

Laboratory (Mouse Genome Informatics). Further details of cells in which recombination occurs with nes-cre are described in [88, 89].

Generation of *EGFRvIII* mice with *piggyBac* transposition

In this study, we employed a conditional *piggyBac* transposon system for forward genetic screening for brain tumors. Specifically, we used the ATP1-S2 mouse line, which contains 20 copies of the *piggyBac* transposon driven by the CAG promoter (containing the cytomegalovirus early enhancer element; the promoter, first exon and first intron of the chicken beta-actin gene; and the splice acceptor of the rabbit beta-globin gene). This is considered a low-copy transposon line, in comparison with the high-copy lines that contain 80 copies of the transposon and these were found to be more likely to cause embryonic lethality due to excessive transposon mobilisation [69]. In order to enable conditional screening, we used a conditional transposase line that has LoxP sites either side of a neomycin-polyA cassette in front of the transposase sequence (when cre is expressed this neomycin-polyA sequence is removed and transposase is expressed). Thus, when mice containing the conditional transposase and ATP1S2 are crossed with *nes*-cre mice, the resulting offspring containing all three alleles have transposons being mobilised in the central nervous system.

For conducting transposon-based forward genetic screens for cancer, careful consideration must be given to the mouse breeding strategies so that transposon mobilisation only occurs at the intended time. Thus, the transposon and transposase should only be together in the final experimental cohorts rather than in any of the parental breeds. We crossed *EGFRvIII* mice with tissue-specific *piggyBac* transposase (conditional transposase; TSPB or Rosa26$^{LSL\text{-}PB}$) mice to yield offspring with *EGFRvIII*/+; TSPB/+. The offspring were crossed with each other to yield homozygotes for both alleles (*EGFRvIII*/*EGFRvIII*; TSPB/TSPB) and also *EGFRvIII*/*EGFRvIII*; TSPB/+ mice. Simultaneously, *nes*-cre mice were crossed with those carrying the ATP1S2 allele to yield *nes*-cre/+; ATP1S2/+ mice, which were then crossed with each other to give double homozygotes for these two alleles. To generate the main the

experimental cohort with both *EGFRvIII* expression and transposition, *EGFRvIII*/*EGFRvIII*; TSPB/TSPB mice were crossed with *nes*-cre/*nes*-cre; ATP1S2/ATP1S2 mice, giving mice heterozygous for these four alleles (*EGFRvIII*/+; TSPB/+; *nes*-cre/+; ATP1S2/+). Several breeding pairs were set up in this fashion in order to generate relatively large numbers of experimental mice in a timely manner. 120 animals with this genotype were produced in total. *EGFRvIII*/*EGFRvIII*; TSPB/+ were also crossed with mice doubly homozygous for *nes*-cre and ATP1S2, generating mice with *EGFRVIII*/+; TSPB+/+; *nes*-cre/+; ATP1S2/+ (these are controls with *EGFRvIII* expression but no transposition) and those with *EGFRvIII*/+; TSPB/+; *nes*-cre/+; ATP1S2/+. 60 animals with *EGFRvIII* expression without transposition were produced. In addition, 80 mice with transposition but no *EGFRvIII* allele were generated as a separate control cohort. A simplified outline for this breeding strategy is shown in Fig 2.2.

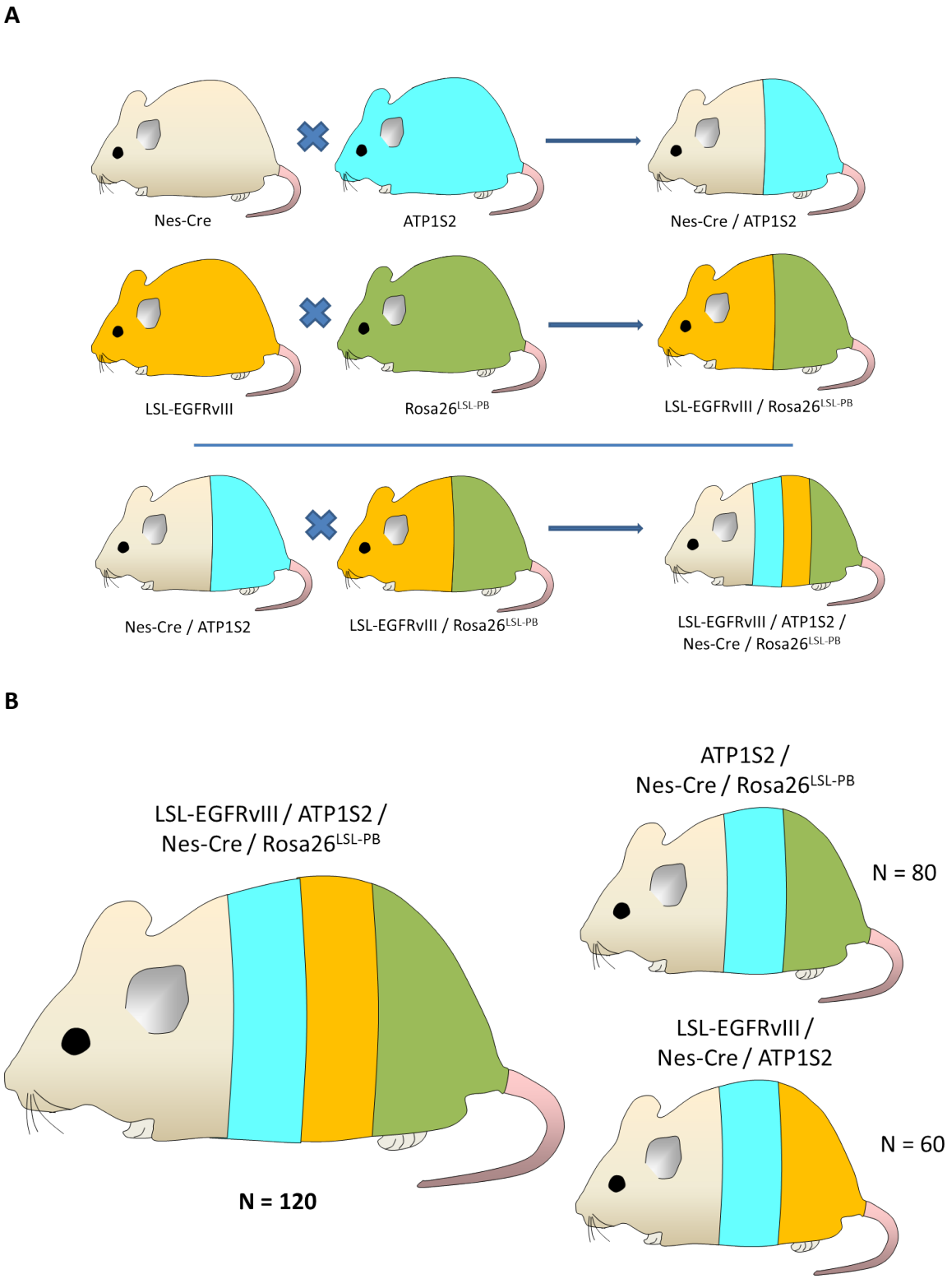


Figure 2.2. Breeding strategy for generating *EGFRvIII*-transposition mice. A. The mouse crosses required in this strategy. B. Final cohorts, with the number of mice with each genotype generated.

EGFRvIII; *Pten*^{loxP}; *nes*-cre mice

Mice homozygous for *EGFRvIII* were crossed with conditional mice homozygous for *Pten*^{loxP} (Fig 2.3) to generate *EGFRvIII*/+; *Pten*^{LoxP}/+ mice. These latter mice were then crossed with each other, and offspring with genotypes *EGFRvIII*/*EGFRvIII*; *Pten*^{LoxP}/^{LoxP} or *EGFRvIII*/+; *Pten*^{LoxP}/^{LoxP} were used for downstream crosses with homozygous *nes*-cre for production of experimental cohorts. Thus, *EGFRvIII*/+; *Pten*+/-; *nes*-cre/+ mice (n=12) were the main experimental cohort, and *EGFRvIII*+/+; *Pten*+/-; *nes*-cre/+ mice (n=10) were the control cohort lacking the *EGFRvIII* allele. These mice were put onto brain / spinal tumor watch from age 6 weeks, observing for the clinical signs as described previously.

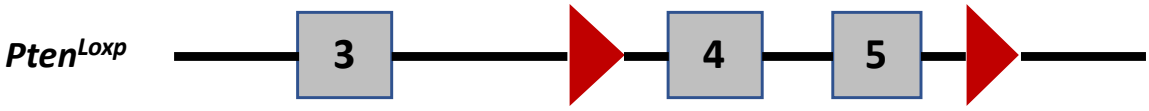


Figure 2.3. *Pten*^{LoxP} allele structure, with exons 4 and 5 flanked by LoxP sites (red arrows) and subsequently deleted with cre expression. Adapted from [42].

Generation of *Trp53*^{R172H} mice with *piggyBac* transposition

The two main cohorts of mice we generated for this study were:

Trp53^{R172H} / + ; *nes*-cre/+ ; ATP1S2/+ ; TSPB/+ - 120 mice. **Cohort A**

Trp53^{R172H} / + ; *nes*-cre/+ ; ATP1S2/+ - 60 mice. **Cohort B**

In order to generate these mice, we used similar breeding strategy principles as for the *EGFRvIII*-transposon mutagenesis screen. The *Trp53* allele used was *Trp53*^{R172H} (also denoted LSL-p53^{R172H}). *Trp53*^{R172H} / + mice were first bred with TSPB/+ mice, and the offspring were

crossed with each other to generate mice homozygous for both of these alleles. In parallel, *nes*-cre/+ mice were crossed with ATP1S2/+ mice, and offspring from this breeding containing both alleles were crossed with each other until mice homozygous for these two alleles were produced. As a final cross, doubly homozygous *Trp53*^{R172H} ; TSPB mice were bred with doubly homozygous *nes*-cre ; ATP1S2 mice to yield Cohort A mice. Cohort B mice were generated by crossing homozygous *Trp53*^{R172H} mice with doubly homozygous *nes*-cre ; ATP1S2 mice, Fig 2.4 and 2.5. A group of control mice with mobilising transposons in the CNS but no predisposing *Trp53*^{R172H} mutation were generated from the breeding pairs used for the *EGFRvIII*-transposon screen.

All appropriate mice were genotyped for *Trp53*^{R172H}, transposon, transposase and cre alleles to confirm the genetic identity of each individual mouse. Transposon mobilisation only occurred when the transposon, transposase and *nes*-cre alleles were all present in the same mouse, as demonstrated by polymerase chain reaction.

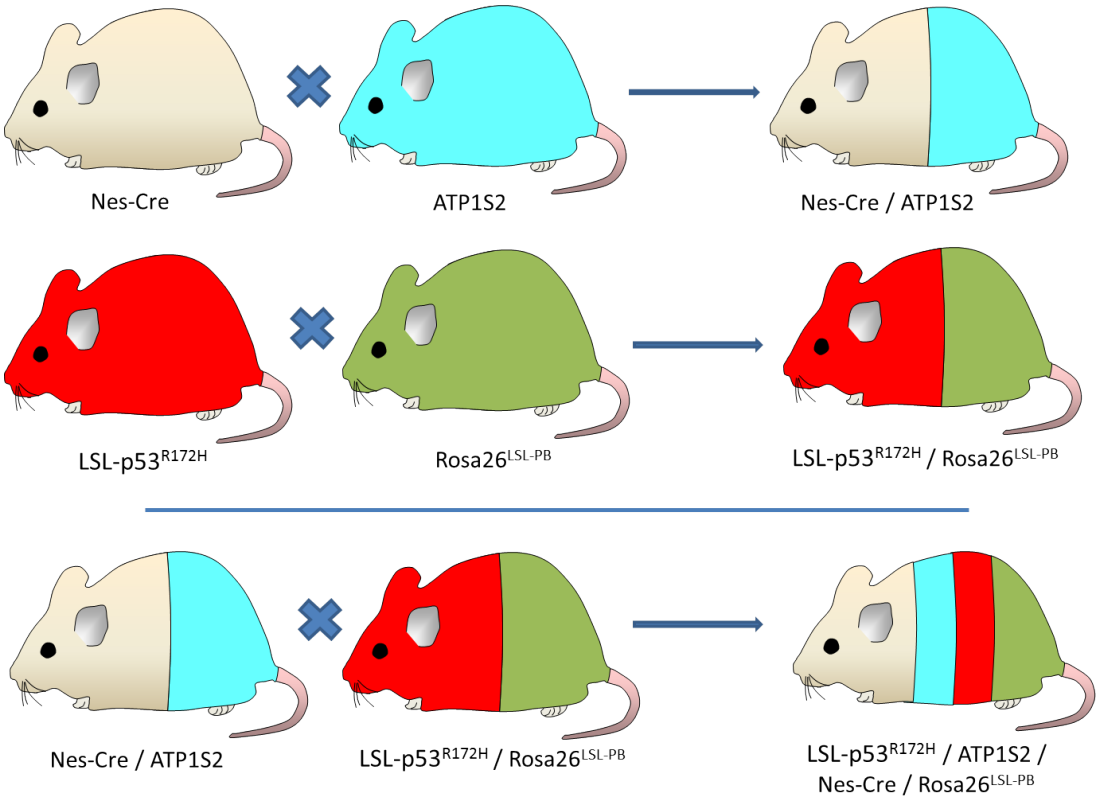


Figure 2.4. Breeding strategy for generating experimental *Trp53*^{R172H} – PB mice.

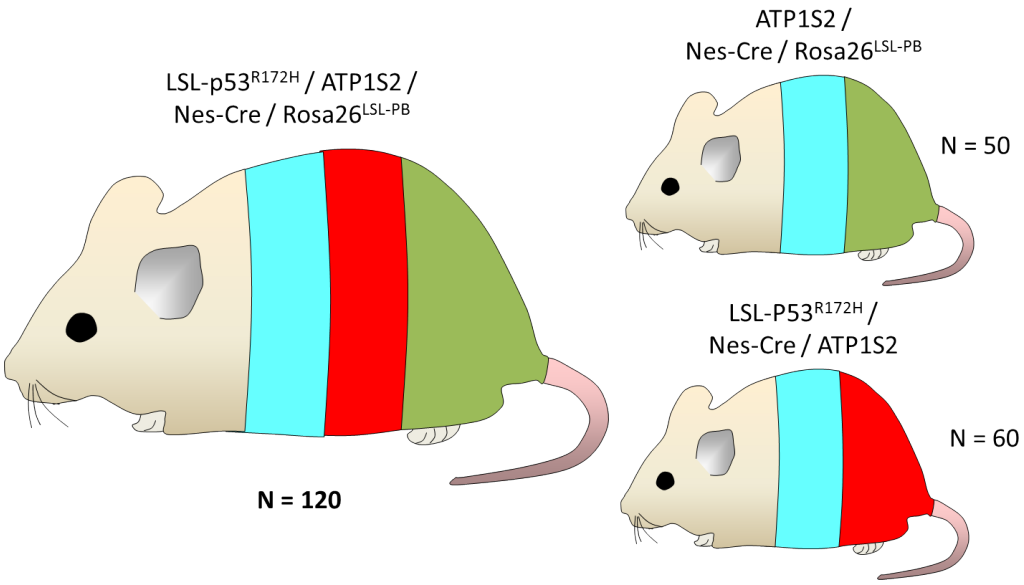


Figure 2.5. Cohorts of experimental mice (left) and control mice (right).

Generation of *Trp53^{R172H}; Pten$^{+/-}$* ; PB cohorts for screen

Given the long latency and low incidence of glioma formation in our *Trp53^{R172H}* / *piggyBac* mouse model, we decided to set up another smaller scale screen in mice carrying *Trp53^{R172H}* and *piggyBac* alleles in addition to a *Pten*-null allele, given that mutant-*Trp53* has been found to cooperate with *Pten* loss in gliomagenesis ([94].

In order to generate experimental mice with the genotype *Trp53^{R172H}* / +; *Pten^{LoxP}* / + ; *nes*-cre / + ; TSPB / + ; ATP1S2 / +, we followed a similar breeding strategy to that used for the *Trp53^{R172H}* - *piggyBac* screen: *nes*-cre / *nes*-cre ; ATP1S2 / ATP1S2 mice were produced and these were crossed with *Trp5^{R172H}* / + ; TSPB / TSPB ; *Pten^{LoxP}* / *Pten^{LoxP}* mice. Although there are reports that homozygous *Trp53* loss leads to acceleration of tumor formation in mice [94], we did not produce homozygous *Trp53^{R172H}* mice for the screen here because we found that these mice had a high incidence of clinical abnormalities outside of the CNS (lymphomas and sarcomas as described in [43]) even when expressed under control of *nes*-cre (these data are not presented in this Thesis given they are not the focus of this work). This was presumably

the result of having only one functional germline copy of *Trp53* in these mice. The mice produced for this screen were:

Trp53$^{LSL\text{-}R172H}$ / + ; *Pten*^{LoxP} / + ; *nes*-cre / + ; TSPB / + ; ATP1S2 / + - **n = 40.**

Pten^{LoxP} / + ; *nes*-cre / + ; TSPB / + ; ATP1S2 / + - **n = 20.**

Pten^{LoxP} / + ; *nes*-cre / + ; ATP1S2 / + - **n = 20.**

Mouse Genotyping Protocols

In this section, I describe the protocols I used for genotyping all the mice generated in for this Thesis.

Genotyping of *EGFRvIII* mice

Mice carrying the conditional *EGFRvIII* allele were crossed with those carrying *nes*-cre. Ear clips from mice were lysed according to the Kapa Mouse Genotyping Kit protocol for DNA extraction. The primers used for *EGFRvIII* genotyping are: forward, 5'-CCCCCTGAACCTGAAACATAA-3'; reverse, 5'-TAAATGCCACCGGCAGGATG-3'. The *EGFRvIII* amplicon size for this reaction is 670 bp. The reaction conditions are: 94°c for 3 minutes, 35 cycles at 94°c for 30 seconds, 62°c for 30 seconds, and 72°c for 1 minute, followed by an extension at 72°c for 3 minutes.

For distinguishing between heterozygous and homozygous *EGFRvIII* mice, a Real-Time PCR was performed using the TaqMan (ThermoFisher) assay with standard manufacturer's conditions and these primers: forward 5'- GCTATGAGATGGAGGAAGACG-3'; reverse 5'-TCACCAATACCTATTCCGTTACAC-3'; Probe 5'-FAM-AGGCCCTTCGCACTTCTTACACTT-TAM-3'. 'No template control' reaction mixtures were also made to detect any contamination of the reaction mix. For each sample, there was a corresponding reaction with Beta-actin primers, allowing for normalisation of *EGFRvIII* DNA and relative quantification. Values were then expressed relative to wild-type samples to quantify the *EGFRvIII* DNA copy number in mutant samples. Reactions were performed in triplicate for each sample and the mean copy number values were subsequently calculated.

Nes-cre genotyping

Mice carrying the *nes*-cre allele were imported from Jackson Laboratories, and re-derived in the animal house of the Wellcome Trust Sanger Institute. Genotyping for the *nes*-cre allele was also with the Kapa Mouse Genotyping kit and the following generic cre allele primers:

Primer	Sequence 5' --> 3'	Primer Type
oIMR1084	GCGGTCTGGCAGTAAAAACTATC	Transgene Forward
oIMR1085	GTGAAACAGCATTGCTGTCACTT	Transgene Reverse
oIMR7338	CTAGGCCACAGAATTGAAAGATCT	Internal Positive Control Forward
oIMR7339	GTAGGTGGAAATTCTAGCATCATCC	Internal Positive Control Reverse

DNA was extracted from ear clips. The PCR cycling parameters for genotyping the cre allele were: 94°c for 3 minutes, 35 cycles of 94°c for 30 seconds, 58°c for 1 minute, and 72°c for 1 minute, followed by a 2-minute extension at 72°c. The cre allele produced a 100-bp band, and the internal positive control band was 324-bp.

Genotyping of ATP1S2 and TSPB Alleles

Genotyping for the ATP1S2 allele was performed with PCR primers as follows:

ATP F: CTCGTTAATCGCCGAGCTAC

ATP R: GCCTTATCGCGATTTTACCA

This reaction yielded an 808 bp fragment in the presence of the ATP1S2 allele and no fragment in wild-type cases.

The following primers were employed to detect the TSPB allele:

BpA5F: GCTGGGGA TGCGGTGGGCTC

Rosa3R: GGCGGATCACAAGCAATAATAACCTGTAGTTT

This reaction yielded a 250 bp fragment in the presence of the TSPB allele and no fragment in wild-type cases.

Genotyping for *Pten*^{LoxP} allele

The *Pten*^{loxP} mouse allele we used was constructed by Trotman et al[42], with genotyping performed as described in their publication. The strain of this mouse is C56BL/6J. The DNA from ear clips of appropriate mice were genotyped for *Pten*^{loxP} allele with three primers in a PCR: primer 1 (5'-AAAAGTTCCCCTGCTGATGATTTGT-3'), primer 2 (5'-TGTTTTTGACCAATTAAAGTAGGCTGTG-3'), and primer 3 (5'- CCCCCAAGTCAATTGTTAGGTC TGT-3'). Universal PCR thermocycling parameters, as I have described previously, were implemented for these reactions. The wild-type *Pten* allele produces a 350bp band and the *Pten*^{loxP} allele yields a 450bp band in this reaction.

***Trp53*^{R172H} Allele Genotyping**

Mice carrying a conditional *Trp53*^{R172H} mutant allele (*Trp53*^{tm2Tyj}) were imported from The Jackson Laboratory. This mouse allele was originally produced in the laboratory of Professor Tyler Jacks, and is a dominant negative allele [43]. The strain of this *Trp53*^{R172H} allele is C57BL/6J.

DNA was extracted from mouse ear clips and genotyped using the KAPA Mouse Genotyping Kit (methods described previously). Primers used for genotyping the *Trp53*^{R172H} allele are:

Primer	Sequence 5' --> 3'	Primer Type	
25927	AGG TGT GGC TTC TGG CTT C	Wild type Forward	Reaction A
25928	GAA ACT TTT CAC AAG AAC CAG ATC A	Common	Reaction A
25929	CCA TGG CTT GAG TAA GTC TGC A	Mutant Forward	Reaction A

The following thermocycling conditions were applied to detect the *Trp53*^{R172H} allele:

Cycling

Step #	Temp °C	Time	Note
1	94	2 min	
2	94	20sec	
3	65	15sec	-0.5 C per cycle decrease
4	68	10sec	
5			repeat steps 2-4 for 10 cycles
6	94	15sec	
7	60	15sec	
8	72	10sec	
9			repeat steps 6-8 for 28 cycles
10	72	2 min	
11	10		Hold

The mutant allele produced a 174 bp fragment, and the wild-type allele gave a 370bp band.

Mouse Clinical Observation and Tissue Processing

In order to ensure clinical endpoints are consistent across our cohorts of mice, careful standards were put in place for clinical observation of mice and for culling when clinical endpoints were met. Here I describe these protocols and also the methods used for downstream processing of the collected tissues.

Tumor Watch

Mice heterozygous for *EGFRvIII* and *nes*-cre were placed onto 'tumor watch' from 4 weeks old. Specifically, these mice were monitored daily in particular for neurological signs, including limb weakness, ataxia, hydrocephalus / macrocephaly, head tilt and / or circling, lethargy, and weight loss. Mice were culled when the neurological signs were sufficient to impair basic functioning of the mouse such as feeding. The procedure for culling mice followed the Schedule 1 protocol for humane culling. All protocols involving mice were ethically approved locally by the animal facility at the Wellcome Trust Sanger Institute.

Dissection of eyes, brain and spinal cord

After culling the appropriate mice, eyes were dissected by holding the eyelids apart and using blunt forceps to lift the eyeball from its origin and to gently dissect it en bloc from the orbit, taking care not to apply pressure to the eyeball itself. Care was taken to dissect the eyeball with a portion of the optic nerve attached where possible, as this helps with orienting the eyeball in histological analysis.

The brain was dissected by using fine scissors to cut the cranium in the midline from posterior to anterior; cuts were made laterally to expose the brain completely. The brain was lifted en bloc from the skull base and removed. The spine was dissected by opening the thorax and abdomen with scissors; the thoracic and abdominal organs were removed with forceps. A transverse cut was made at the lumbar spine, and the spine was then dissected from the

posterior skin / subcutaneous tissue all the way up to the cervical spine. The brain, eyes and spine were placed directly into formalin (buffered) and left to fix for 24 – 48 hours. These specimens were stored at 4°c in the fridge, as previous studies have shown this temperature is better than room temperature for preserving nucleic acids over longer periods [105].

To facilitate RNA-sequencing of normal tissue, brain and spinal cord samples from control mice (lacking *EGFRvIII*) were dissected and stored in RNA-later (ThermoFisher). Age-matched mouse control samples (carrying the *nes*-cre allele but not *EGFRvIII*) were also stored in formalin and processed as described for histology to allow for a comparison with the *EGFRvIII*-mutant mouse samples; these included in particular, 7 brains, 10 eyes and 7 spines from wild-type mice.

Samples stored in RNA-later were kept at 4°c overnight then transferred to -20°c for longer term storage; RNA-later preserves RNA well by inactivating RNAses, avoiding the need for immediate freezing of samples.

Brain Tumor Dissection

Macroscopic photographs of mouse brains were correlated with pathological findings on H&E staining, allowing specific identification of the tumor regions. Tumor samples were then carefully dissected from the brain under a dissection microscope, aiming for at least 2 – 3 mm of tissue per sample (this is a sufficient amount to obtain an acceptable quantity of DNA for transposon-based sequencing). The tissue samples were placed in formalin, or if dissected immediately from fresh tissues then they were placed directly into RNA-later for downstream RNA extraction and / or part of the tumor was flash-frozen in liquid nitrogen for downstream DNA extraction (e.g. for whole-exome sequencing or transposon insertion mapping). Not all of the visible tumor material was removed from each mouse brain: some was left intact in order to process for histological diagnosis. In this way, we obtained material for transposon-based sequencing and for histopathology in the vast majority of cases.

Spinal Tumor Dissection

Mice that displayed signs of neurological disease such as paralysis were culled as described above. The spinal column was dissected and cut into two segments (cervical / thoracic spine, and lumbar spine). The cervical / thoracic spine was placed into formalin for histological processing. The lumbar spine was dissected under the microscope: the superficial soft tissue was cut and removed; the vertebral laminae were carefully removed with fine scissors to expose the spinal cord, and finally the vertebral spinous processes were gently removed to expose the entire spinal cord. In most tumor-watch mice, there was clearly abnormal tissue overlying the spinal cord and infiltrating the surrounding nerve roots, corresponding to the tumor tissue diagnosed on histology. Samples of this tumor tissue were removed, both at the subdural / subarachnoid components and the spinal root components; some tumor tissues were snap frozen in liquid nitrogen, others were placed in RNA-later. In cases where whole-exome sequencing of a brain or spinal tumor was planned, a spleen sample was snap frozen from the same mouse as to provide a normal DNA control.

Sectioning of Brains

In order to study precursor lesions of *EGFRvIII*-driven gliomas, we studied the histology of *EGFRvIII*; *nes*-cre mouse brains prior to clinically overt phenotypes. For all brain samples, including controls and those with tumors, we used the same sectioning technique in order to avoid biases in detection of lesions. Specifically, four coronal slices were made for each brain, from anterior to posterior including three sections with cerebral cortex and one with cerebellum. The subventricular zone was present in at least one of these slices. The pathologist examining these sections was blinded to genotype, which again helps reduce a bias in detection of lesions in particular genotypes. Haematoxylin and eosin (H&E) staining was performed on all samples that contained areas that appeared to be proliferative or tumor-like.

The majority of tumor samples were photographed using a high-resolution camera (Panasonic DMC-T27), both macroscopically and through a dissecting microscope. These images provided a record for later correlation between tumor samples and their underlying histopathology.

Tissue Extraction and Storage

Mice with the appropriate genotypes were placed on tumor watch in order to observe for clinical signs of neurological disease, including lethargy, weight loss, seizures, weakness, macrocephaly, and abnormal gait. Those mice with such clinical signs were culled with a schedule 1 protocol using CO_2 (cervical dislocation was avoided to avoid damage to brain and spinal cord tissue). Brains and spinal columns of mice were dissected from whole bodies as described previously. These samples were stored in formalin at 4^{o}c, allowing long-term preservation of tissue and nucleic acids. All brains and spinal cords from tumor watch mice were processed for histological analysis by paraffin-embedding and sectioning to determine if there were tumors or tumor precursor lesions present, Fig 2.6.

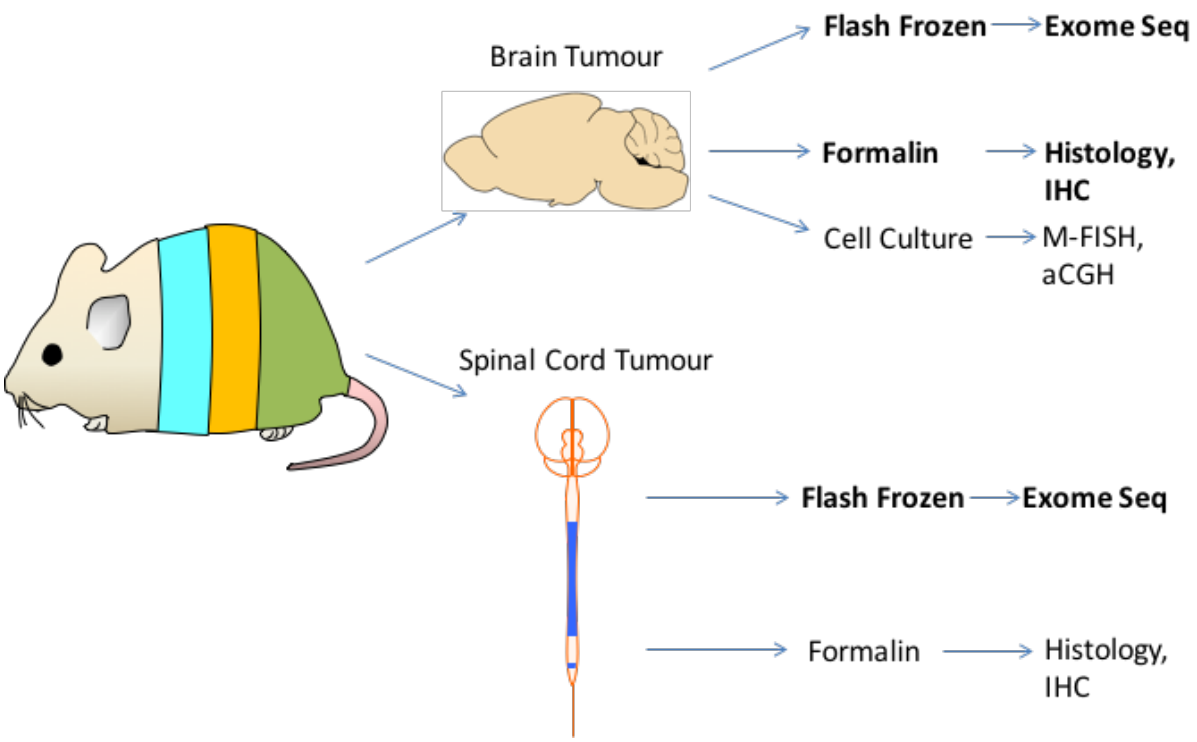


Figure 2.6. Processing strategy for tumors induced in mice. Brain and spinal tumors were collected into different storage reagents dependent on downstream sequencing and experimental requirements.

Histology

Brain and spinal tissues were fixed in 4% paraformaldehyde and then embedded in paraffin. 4μm sections were stained with haematoxylin and eosin (H&E) for morphological analysis. A consultant neuropathologist (Professor Sebastian Brandner, Department of Neuropathology, National Hospital for Neurology and Neurosurgery, UK), who has extensive experience in the pathology of central nervous system tumors from humans and mice and who was blinded to genotype, reviewed all histological sections for pathological diagnosis. In addition to reviewing samples with mutant alleles (*EGFRvIII*, transposition, $Trp53^{R172H}$, $Pten^{+/-}$), our neuropathologist reviewed 7 brains and 7 spinal cords from adult (age range 12 -43 weeks) mice containing only *nes*-cre, to serve as controls.

Neuropathological diagnosis and grading of gliomas was established using the following grading system by our neuropathologist (Professor Sebastian Brandner): grade I: tumors of low-to moderate cellularity, overall bland cytological appearance, bland nuclear morphology and only rare, or no mitotic figures. Grade II: tumors with moderate or high cellularity, occasional mitotic figures, and absence of microvascular proliferation and necrosis. Grade III: tumors with high cellularity, clear presence of mitotic figures, including brisk mitotic activity, hyperchromatic nuclei, but with no microvascular proliferations and no necrosis. Grade IV: highly cellular tumors, with densely packed nuclei, often a high nucleus to cytoplasm ratio, frequent mitotic figures, and with either microvascular (vascular endothelial) proliferations, or necrosis, or both. Microneoplasias were defined as gliomas with features of glioma grade I or II but were smaller in size (100 – 300μm length). Primitive neuroectodermal tumors were diagnosed for tumors displaying hyperchromatic cells and mitosis or necrosis, and some Homer-Wright rosettes with central solid cores of neurofibrillary material. All histopathological images used in this Thesis were discussed with and agreed by Professor Brandner to provide the pathological interpretation described herein, and all histological diagnoses were also provided by Professor Brandner.

Immunohistochemistry

Immunohistochemistry staining was performed using the Ventana Discovery XT instrument, using the Ventana DAB Map detection Kit (760-124), an automated system for high-throughput immunohistochemical (IHC) staining. The protocol for IHC conducted by this instrument is as follows: deparaffinisation in 'EZ prep' solution at 75°c for 8 minutes, then cell conditioning using Cell Conditioning (CC1) solution at 95°c for 44 minutes. This is followed by blocking with inhibitor ChloroMap (CM) at 37°c for 4 minutes then incubation with the primary antibody for 60 minutes. After this, one drop of either anti-rabbit or anti-mouse horseradish peroxidase as appropriate is added with incubation for 16 minutes. One drop of DAB (diaminobenzene) CM and one drop H_2O_2 are applied with incubation for 8 minutes, followed by one drop of copper CM with 5 minutes of incubation. Slides were haematoxylin

counterstained (8 minutes incubation then post counterstaining with Bluing reagent for 8 minutes). These solutions were obtained from Roche, Ventana Medical. The antibodies used, with 100μl volume of each, are shown in Table 2.1.

Antibody	Dilution	Source	Pre-treatment	Primary Antibody incubation	Swine anti Rabbit Dako E0353	Rabbit anti Mouse Dako E0354
Olig2	1:100	Millipore ab9610	CC1 45min	4h	32min	
Sox2	1:500	Abcam ab97959	CC1 60min	1h	32min	
Nestin	1:500	Abcam ab22035	CC1 45min	1h		32min
Ki67	1:100	Cell Signalling 12202S	CC1 30min	1h	32min	
GFAP	1:1000	Dako Z0334	Protease 1 12min	32min	32min	
PDGFRa	Pre-diluted	Abcam ab15501	CC1 30min	12h	32min	
EGFR (31G7)	1:100	Life Technologies (ref 280005)	Protease 1 12min	1h		32min
EGFRvIII	1:100	Sigma MABS1915	Protease 1 12min	1h		32min

Table 2.1. Antibodies used for immunohistochemistry.

In addition to staining tumor samples, 5 normal brain controls and 5 normal spinal cord controls (from *nes*-cre mice without *EGFRvIII*) were subjected to the same IHC stains. These IHC stains were kindly performed in collaboration with the Department of Neuropathology at National Hospital for Neurology and Neurosurgery, UK (IQPath, Ms Angela Richard-Londt).

Establishing Primary Cultures

Mouse brain tumors were carefully dissected under the microscope; instruments were cleaned with ethanol prior to each use to reduce the chances of tissue contamination. A small portion of the brain tumor was placed in cold saline on ice. This sample was then processed as soon as possible for primary culture establishment: it was cut into small pieces with a scalpel, and incubated in Accutase (STEMCELL Technologies) for 15 minutes at 37°c to dissociate the cells under a sterile hood. The cell suspension was centrifuged at 300g for 3 minutes, and Accutase removed; the cells were washed with PBS three times before being adding to culture medium and plated in a 6-well plate. The culture medium was composed of: DMEM/F12 medium (50%), neurobasal medium (50%), hEGF (25ng/ml), bFGF (25ng/ml), N2 (1x), B2 (1x), BME (1x), PSL (1x). Penicillin/streptomycin and amphotericin B were used in the first passage only to reduce the risk of bacterial and fungal infection, given that tumor tissues were extracted from mice under non-sterile conditions. The cultures were incubated at 37°c, and split every two – three days as required.

To preserve the cell lines, samples were cryopreserved: cells were washed with PBS, split into single cells, and added to 1ml of a combination of 90% neural media and 10% DMSO and transferred to a cryovial. Cryovials were placed in a freeze-container and put in the -80°c freezer. These were transferred to liquid nitrogen the following day for long-term storage. In addition, multiple cell pellets were flash frozen at -80°c for downstream RNA and / or DNA extraction.

Genetic and Transcriptomic Characterization of Tumors

DNA Extraction

DNA was extracted from mouse tumor and spleen tissue according to the MagMax DNA Multi-sample Kit (ThermoFisher) instructions. Briefly, 2-3mm of tissue was incubated for 24 hours in proteinase K solution at 55°c in a shaking incubator. For formalin-fixed tissue, the sample was washed with phosphate-buffered saline (PBS) prior to proteinase K digestion and after digestion it was then incubated for one hour at 95°c, which helps reverse some of the formalin cross-links in the DNA. 100% isopropanol was then added to the lysate, which was then vortexed. Magnetic beads were added to the mixture and then this was placed on a magnetic bead stand. The sample was washed twice with ethanol / isopropanol based wash buffers, and RNAse was added to lyse RNA in the sample. After two further wash steps, DNA was eluted in 200μl of elution buffer, with sample tubes placed on the magnetic stand to separate the eluate from the beads. The eluate was transferred to a fresh tube and DNA quantified using the NanoDrop (ThermoFisher; this is a spectrophotometer for absorbance-based quantification of nucleic acids). All samples were stored at -20°c for long term preservation.

FISH

For multiplex-fluorescence in situ hybridization (M-FISH), chromosome-specific DNA library for each mouse chromosome was generated from 5,000 copies of flow-sorted chromosomes, provided by Flow Cytometry Core Facility of the Wellcome Trust Sanger Institute, using GenomePlexWhole Genome Amplification (WGA2) kit (Sigma-Aldrich). Mouse 21-color painting probe was made following the pooling strategy [106]. Five chromosome-pools were labelled with ATTO 425-, ATTO 488-, CY3-, CY5-, and Texas Red-dUTPs (Jena Bioscience), respectively. We performed this by the use of WGA-3 re-amplification kit (Sigma-Aldrich). Next, the labelled products were pooled and sonicated to obtain a size range of 200–1,000 bp, required for use in chromosome painting. Sonicated DNA (sufficient for 10 hybridizations) was precipitated with ethanol together with mouse Cot-1 DNA (Invitrogen)

and this was re-suspended in hybridization buffer. Metaphase preparations were dropped onto pre-cleaned microscopic slides; these were fixed in acetone and dehydration through an ethanol series. Metaphase spreads on slides were then denatured by immersion in an alkaline denaturation solution and dehydration. The M-FISH probe was denatured before application onto the denatured slides. Hybridization was performed in a 37 °C incubator for two nights. Post-hybridization washes included a 5-minute stringent wash in 0.5 × SSC at 75°C, and then a 5-minute rinse in 2 × SSC containing 0.05% Tween20 (VWR) and a two-minute rinse in 1 × PBS, both at room temperature.

Slides were mounted and the images were visualised on a Zeiss Axio-Imager D1 fluorescent microscope built with narrow band-pass filters for DAPI, DEAC, FITC, CY3, TEXAS RED, and CY5 fluorescence and an ORCA-EA CCD camera (Hamamatsu). The SmartCapture software (Digital Scientific UK) was used to capture M-FISH digital images, and these were processed with the SmartType Karyotyper software (Digital Scientific, UK). At least 10 metaphases for each sample were fully karyotyped.

Counting FISH Chromosomal Aberrations

We quantified the cytogenetic anomalies found on FISH as follows: single translocations, copy number gains or losses were counted as one anomaly for each chromosome; for polyploidy in all chromosomes, this was counted as one anomaly for each cell in which this was seen for a particular culture.

Whole-exome sequencing

DNA was extracted from mouse tumors and the matching spleen (as a control to enable later filtering of germline single nucleotide polymorphisms, SNPs) from the same mice. For whole-exome sequencing (WES), extracted DNA was first quantified (using Accuclear UltraHS dsDNA Standards Assay reagent kit and BMG FLUOStar Omega fluorescence reader), followed by normalising each sample to 4.17ng/μl in 120μl in preparation for library creation (performed

by Wellcome Trust Sanger Institute Sequencing Facility). DNA was sheared into fragments of 150bp (on the Covaris LC220 and Agilent Bravo automated workstation) followed by library creation and amplification using unique indexed tags and adaptors (Agilent's SureSelectXT Automated Library Prep & Capture Kits and MJ Tetrad). The amplified libraries were then purified (using Agencourt AMPure XP and Beckman Coulter Biomek NX96 automation) and eluted in nuclease-free water, followed by a second round of quantification. The libraries were then diluted to an appropriate concentration for introduction into the exome-capture stage. Exome pulldown (hybridization) was performed using *Mouse-All-Exon* oligo-baits (Agilent) for 23 hours at 65ºC. Uniquely indexed samples were baited and captured into pools. The pulldown was then purified and eluted using streptavidin-coated Dynal beads ready to be amplified (on the MJ Tetrad). The amplified product was further purified, and subsequently quantified using the Agilent Bioanalyzer and finally, subjected to sequencing on the HiSeq Illumina 2500 platform.

Somatic variant calling and CNV analysis

Sequencing reads were mapped to the *Mus musculus* genome (GRCm38/mm10) using BWA-MEM (version 0.7.16a; Burrows-Wheeler Aligner – a software for mapping low-divergent sequences against a large reference genome)[107], with default parameters. Duplicate reads were marked by biobambam2 (tools based on collation of read alignments in BAM files by read name), and base quality scores were recalibrated with GATK (Genome Analysis Toolkit, version 3.7 – tools focused on identifying variants and genotyping from high-throughput sequencing data)[108]. Sequencing coverage ranged from 50 – 80 x for each sample, as confirmed by sequencing read counts. Somatic variant calling of tumor and its matched normal BAM files were performed using Mutect2 (version 3.8). Mutect2 is a publicly available tool for calling single-nucleotide variants/SNVs and insertions/deletions/INDELs, via local assembly of haplotypes; the tool applies a Bayesian classifier for detecting somatic mutations even with low allelic fractions (including below 0.1) [109]. Mutations were annotated to a database of GRCm38.86 by SnpEff-4.3i[110]. Significantly mutated genes (SMGs) were

identified by the Mutational Significance in Cancer framework (MuSiC, Version 0.4)[111] with default parameters; genes were called SMGs if: mutated in two or more tumors; corrected Likelihood ratio test p-value < 0.01 and FDR < 0.2, and Convolution test p value < 0.01 . The MuSiC framework is a method for identifying SMGs as genes that display a significantly higher mutation rate than the background mutation rate (BMR), taking into account multiple mutational mechanisms such as splice site mutations, coding indels and single nucleotide variants (SNVs).

To detect somatic copy-number alterations, the pileup files of tumor and its matched normal BAM files were generated by samtools mpileup (version1.5 – tools for the manipulation of alignments in BAM format)[112], followed by copy number analysis using varScan2 (version 2.4.2 – 'Variant detection in massively parallel sequencing data') [113] with default parameters. Copy number variations (CNVs) were segmented using circular binary segmentation algorithm[114], which was implemented in DNAcopy (version 1.52). GISTIC2 (version 2.0.23)[115] with the following parameters: "qvt = 0.05, confidence level = 0.99, and maxseg = 20000" was performed to find focal CNVs using the Mus musculus (mm10) refSeq gene annotations. This bioinformatic analysis was performed in collaboration with Dr JK Kim (Daegu Gyeongbuk Institute of Science and Technology, South Korea).

RNA Extraction

RNA was extracted from stored tissue samples using the MN Nucleospin RNA kit according to manufacturer's instructions. Briefly, 5mg pieces of tissue samples were homogenised, then lysed using lysis buffer and DDT. The lysates were filtered through the Nucleospin filters, and 70% ethanol was added. The lysates were passed through Nucleospin RNA Columns, the membranes of which were then desalted with desalting buffer. DNase reaction mixture was prepared and added to the column membranes to digest DNA. Following this, three wash steps and drying of the membranes were performed, and RNA from each column was eluted in 60μl of RNAse-free water.

RNA-sequencing and bioinformatics analysis

RNA-seq libraries were constructed using the Illumina Tru-Seq Stranded RNA protocol with oligo dT pulldown and sequenced on Illumina HiSeq2500 by 75-bp paired-end sequencing. The RNA-seq data for samples were generated as 75 bp paired-end Illumina reads and aligned using STAR[116] to the human genome (GRCh37). The total number of reads that align to the exons of each gene as defined by Ensembl (version 75 – a genome browser for vertebrate genomes)[117] were obtained using STAR (a software for aligning RNA-seq reads)[116]. The obtained gene counts were used to obtain expression fold changes (FC) and False Discovery Rates (FDRs) for genes between any two conditions using DESeq2 (a tool for differential expression analysis) [118]. The genes were considered differentially expressed if their -2.0 > log-FC > 2.0 and the Benjamini-Hochberg adjusted p-value ≤ 0.01 (p-value from Wald's test).. The Gene Set Enrichment Analysis (GSEA) against each of the MsigDB (Molecular Signatures Database) [119] gene datasets were performed using the GSEA tool[120]. Brain tumor RNA-seq data were compared with normal brain samples (cerebral cortex from *nes*-cre only mice without *EGFRvIII*, n=6), and spinal tumor RNA-seq data were compared with normal spinal cords (from *nes*-cre mice as previous, n=6). Analysis of RNA-seq data was done in collaboration with Dr MS Vijayabaskar (Wellcome Trust Sanger Institute).

In order to detect specifically the presence of human *EGFRvIII* transcripts in RNA-seq data from mouse tumors (therefore also indicating that recombination of the conditional *EGFRvIII* allele has occurred), the human *EGFR* sequence (obtained from Ensembl) with exons 2 to 7 removed was introduced into the mouse reference genome as a separate gene prior to RNA-seq alignment. The total number of reads aligned to the *EGFRvIII* gene was then counted as described above. This process was applied both to brain and spinal tumors as well as to control wild-type brain and spine samples (which do not contain the *EGFRvIII* allele).

PiggyBac Fusion Transcript Detection

Transposon insertion sites from RNA-sequencing were obtained using IM-Fusion (Insertional Mutagenesis-Fusion – a tool that employs fusion-aware RNA-seq alignment to identify insertions as a result of splicing between endogenous genes and a transposon) [121]. In any given tumor sample, genes with at least one read traversing the transposon-gene junction or by a fragment (read pairs) spanning across the junction were identified. Based on the orientation of the inserted transposon and the feature (splice donor, or splice acceptor) of the integrated transposon, the gene transcript was either declared as activated or truncated. As controls, we analysed 10 *EGFRvIII* ; *nes*-cre; ATP1S2 tumors (lacking TSPB) – there were no read counts supporting fusion transcripts in these tumors, implying fusion transcripts with transposons occur specifically in the presence of transposition only as expected.

Statistical Analysis

Software calculations were performed using Microsoft Excel, GraphPad Prism version 7 or R version 3.2.0 (The R Project for Statistical Computing, http://www.r-project.org/). The p-values, specific test and data representation for each analysis is described in the main text or figure legends. Data were verified to meet the assumptions of the statistical tests used. Stars to represent significance levels are shown in certain figures, with the following meaning: * = p value less than 0.05; ** p value less than 0.01; *** p value less than 0.001; **** p values less than 0.0001.

Transposon Mobilisation

We tested for mobilisation of the ATP1-S2 transposons using a 'jumping' PCR. DNA was extracted from the brain or tumor of mice containing transposase and transposon alleles expressed under *nes*-cre control. The first samples were taken from brains of mice aged 4 -6 weeks in order to confirm that transposon mobilisation starts early (indeed, cre is expressed from embryonic day 13 under the rat nestin promoter). Tissue samples were taken from

various sites of the brain to confirm widespread transposon mobilisation: basal ganglia, cerebral cortex, brainstem, and cerebellum. As control samples, we used brain tissue specimens from mice carrying the ATP1-S2 and *nes*-cre alleles but not the TSPB allele (and therefore there should not be any transposition in these samples). Three separate PCRs were employed to test for jumping (or mobilisation), and two PCRs were implemented for non-mobilisation of the transposons. The jumping PCRs employ reaction primers flanking either side of the ATP1-S2 concatemer; whereas the non-jumping PCRs employ one primer flanking the concatemer and another nested within the concatemer sequence. These were described in the paper by Rad et al [69]. The primers used for these reactions and the associated expected fragment sizes are as follows:

PCR	Forward Primer	Reverse Primer	Band size (bp)
Jumping PCR1	GGCCTCTTCGCTATTACGC	TCAAACGAAGATTCTATGACGTG	253
Jumping PCR2	GGGCCTCTTCGCTATTACG	GGTCGAGTAAAGCGCAAATC	220
Jumping PCR3	GTGCTGCAAGGCGATTAAGT	GGTCGAGTAAAGCGCAAATC	182
Non-Jumping PCR1	GGGCCTCTTCGCTATTACT	CCGATAAAACACATGCGTCA	274
Non-Jumping PCR2	AACAAGCTCGTCATCGCTTT	GGTCGAGTAAAGCGCAAATC	423

Jumping PCR1 gave faint or no bands, and therefore we employed jumping PCR2 and PCR3 for most experiments as these yielded clear bands.

Reaction conditions for these PCRs were the 'universal' conditions, as described previously. For non-jumping control samples, brain tissues were extracted from mice carrying *nes*-cre and ATP1S2 alleles but not the transposase. Kidney and spleen samples from the same mice were

used to as control samples with negligible cre expression to demonstrate that the transposon mobilisation occurred only where cre was expressed.

Splinkerette PCR and Sequencing for PB Integration Sites

Tradis (transposon-directed insertion site sequencing) library preparation was performed as described in [122]. Briefly, DNA extracted from tumor tissue was quantified using the Qubit (a fluorometer for nucleic acid and protein quantification). 2µg of DNA from each tumor was diluted in 1x – low TE buffer to a total volume of 120µl. DNA from the samples was plated in a Covaris plate and sheared on the Covaris instrument with the following parameters:

Duty cycle	**20%**
Intensity	5
Cycles per burst	200
Time	80 s
Temperature	4–7 °C

The resulting sheared DNA samples were then quality-control (QC) assessed using the Agilent High Sensitivity DNA chip to check for a mean fragment size between 200 and 300bp (with re-shearing to be done if the fragment size were considerably larger). Following AmpureXP bead clean-up of the DNA samples, the DNA was end-repaired using the NEBnext DNA Sample Prep Reagent Kit according to manufacturer's instructions, which briefly involved incubating the DNA for 30 minutes at 20°c with T4 DNA polymerase and Klenow DNA polymerase. The end-repaired DNA was then 'A-tailed' through incubation for 1 hour at 37°c with Klenow fragment exo- and dATP. After a further bead clean-up of the DNA, the samples were subjected to adaptor ligation: DNA was incubated with annealed Splinkerette v1.2 adaptor and Quick ligase at 20°c for 1 hour, and a further bead clean-up was done. In order to check for success of

adaptor ligation, the DNA was checked with the Agilent High Sensitivity chip once more to ensure an ~100bp shift to the larger end of the scale in the electropherogram was observed. The library was then split into plates, one for the 3' and one for 5' end of the DNA. A PCR for amplification of the adaptor-ligated library was performed using the DNA samples with the KapaHiFi HotStart kit and a separate primer for each DNA end (3' and 5'), with the following cycling parameters: 95°c for 2 minutes, then 18 cycles of 95°c for 20 seconds, 63°c for 20 seconds and 72°c for 40 seconds, and finally 72°c for 5 minutes. After bead clean-up of the resulting DNA, a further PCR was performed using a separate primer for each library (one for 3' and one for 5') and an index, barcode-containing primer for each individual sample (allowing for multiplexing of the samples for sequencing). The thermocycling parameters for this second PCR were: 95°c for 2 min, 12 cycles of 95°c for 20 seconds, 60°c for 20 seconds and 72°c for 40 seconds, and finally 72°c for 5 minutes. The DNA was bead purified once more, and checked on the Agilent DNA High Sensitivity chip to reveal a multi-spiked profile as compared to smooth curves seen previously.

In order to avoid individual samples being heavily overrepresented in the sequencing pool, the barcoded samples in the libraries were quantified and then combined into an equimolar pool. Briefly, a (standard quantification curve) qPCR was performed using the KAPA SYBR Fast qPCR kit and the diluted DNA samples, with the following thermocycling parameters: 95°c for 5 min, and 32 cycles of 95°c for 15 seconds and 60°c for 45 seconds. The primer sequences for this qPCR reaction are: qPCR2.1 100 µM, 5'-A*ATGATACGGCGACCACCGAGAT*C-3'; qPCR2.2 100 µM, 5'-C*AAGCAGAAGACGGCATACGAGA*T-3'. The data from these qPCR reactions allowed the samples from each library to be pooled equimolarly for multiplex-barcoded sequencing.

Each library pool (one for each transposon end) was sequenced on a separate Illumina MiSeq run, giving 75bp paired-end reads. The libraries were multiplexed for up to 55-samples in each pool in this study, requiring 4 MiSeq runs in total, in order to give high coverage sequencing. Given that previous studies report that there is often a large percentage of non-mobilised transposons in tumors, our sequencing runs had 20% PhiX (a small control genome to allow

quick alignment and estimation of error rates in sequencing) spiked in. The sequencing primers used were:

PB_L_pr_seq, 5'-C*ACCGAGATCTACACCACGCATGATTATCTTTAACGTACGTCACAATATGATTATCTTT*C-3';

PB_R_pr_seq MiSeq 5'-C*ACCGAGATCTACACATGCGTCAATTTTACGCAGACTATCTTT*C-3';

SB_L_pr_seq MiSeq 5'-G*TGAGTTTAAATGTATTTGGCTAAGGTGTATGTAAACTTCC*G-3';

SB_R_pr_seq MiSeq 5'-A*AAAACGAGTTTTAATGACTCCAACTTAAGTGTATGTAAACTTCC*G-3';

Spl_rev_seq MiSeq 5'-T*AATACGACTCACTATAGGTGACAGCGAGCGC*T-3';

Spl_tag_seq MiSeq 5'-A*GCGCTCGCTGTCACCTATAGTGAGTCGTATT*A-3'.

Insertion Mapping

We used the Gaussian Kernel Convolution (GKC) approach of de Ridder et al[123] for identifying *piggyBac* (PB) common insertion sites (CIS), as described previously [69, 122]. CIS are genomic regions of several tens of kilobases in length where transposons insert significantly more frequently than by chance considering the background rate of insertions and number of TTAA canonical insertion motifs. The GKC framework underpins our analysis for identifying *piggyBac* CIS in this study[123]. This framework essentially places a Gaussian kernel function at each insertion in the genome identified by next-generation sequencing. The kernel functions at each position are then summed in order to yield an estimate of the number of insertions; therefore insertions located close to one another will give a taller peak in the estimate of number of insertions, reflecting the fact that neighbouring insertions may produce identical effects on nearby genes. When a peak exceeds a given threshold (α-level), it is determined to be significant. The kernel width is seen as a scale parameter, which can be

altered to produce CISs of different widths. The Bonferroni (multiple-testing) correction is applied to the data in order to reduce the number of false positive CISs.

The sequencing reads were filtered for Splinkerette primer sequences contained within the PB inverted terminal repeats (ITRs). Transposon insertion sites (IS) were established by mapping the sequencing reads to the mouse genome (assembly version GRCm38) using the SMALT aligner (http://smalt.sourceforge.net). For each tumor sample, sequencing reads mapping to the same location in the genome counted as a single IS. The top 300 IS, by read count, of each sample were pooled in a non-redundant set and subjected to a GKC analysis with 'window sizes' (kernel widths), ranging from 10kb to 100kb in 10 kb steps. Similar numbers of CISs were found for each window size, and most CISs were detected across multiple windows. Significant CISs were taken to be those with a Bonferroni-corrected p-value < 0.001 for multiple window sizes. Significant CISs were associated with genes as annotated in Ensembl release 90[124]. Mouse genes labelled as 'predicted' in the Ensembl annotation were not considered in the analyses. Cancer genes were obtained from COSMIC v82 (Catalogue of Somatic Mutations in Cancer) [125]. Analysis of insertion sites was performed in collaboration with Dr Hannes Ponstingl (Wellcome Trust Sanger Institute).

Accession Codes

All the sequencing data generated in this Thesis are available from the European Nucleotide Archive (ENA), accession code ERP024282.

Human Sequencing Data Comparative Analysis

In order to compare the genetic data gained from our mouse work with that from human patients, I used large patient databases of gliomas using online tools as described here.

Reviewing Patient Data on Known Drivers

To develop a clearer background of the spectrum of *EGFR*, *TP53* and *PTEN* mutations and copy number changes that occur in human gliomas, I used The Cancer Genome Atlas (TCGA) patient datasets of low-grade glioma and glioblastoma DNA-sequencing and RNA-sequencing. These datasets are publicly accessible online through the website www.cbioportal.org [126], which displays the frequency copy number changes and mutational profiles for genes of interest, and also shows mRNA expression data thereby enabling us to observe the putative consequences of altered copy number (such as amplification) on gene expression. These datasets only included brain tumors; there are currently no such large published genomic datasets from human spinal gliomas for comparative genomics analysis that we are aware of. Plots were automatically generated on www.cbioportal.org and were formatted as appropriate for presentation on Adobe Illustrator.

Comparative analysis of mouse WES data with human glioma sequencing

To cross-validate our findings from mouse glioma whole-exome sequencing (WES) data and determine if genes which are most frequently and recurrently altered in these mouse tumors are also altered in gliomas from patients, I analysed TCGA low-grade glioma dataset using the website www.cbioportal.org as previously described. To ensure that the full range of genetic alterations in human gliomas were captured, including heterozygous loss and single-copy gain, I used the following terms when inputting my gene list into Cbioportal: 'HETLOSS', 'HOMDEL', 'MUT', 'GAIN' and 'AMP', which respectively code for heterozygous deletion, homozygous deletion, mutation, single-copy gain and amplification. To determine if pairs of genes had significantly co-occurring or mutually exclusive alterations in these patients'

tumors, Fisher's exact test was implemented on gene pairs with a Bonferroni-adjusted significance level of $p < 0.05$. For gene pairs (or larger groups) that had significantly co-occurring alterations in these patients' tumors, the locations of these genes were individually verified using the Ensembl human sequence dataset in order to determine if these genes were located in neighbouring regions such as the same chromosome arm in humans.

Gene Interactions and Pathway Analyses

To analyse for potential interactions for genes based on our RNA-seq data, we used the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database, which is publicly available online (https://string-db.org/). STRING contains information on confirmed and putative protein-protein interactions from multiple sources, including experimentally-derived interactions identified through literature curation, and computationally predicted interactions. A Benjamini-Hochberg adjusted p-value < 0.05 (hypergeometric test) was taken as statistically significant for a collection of genes to have more interactions than predicted by chance compared to a random set of genes of similar size.

Gene Ontology (GO) analysis was performed to determine which biological processes are over-represented in the differentially expressed genes in mouse gliomas. This analysis was conducted using the online tool, DAVID Bioinformatics Resource 6.8 (https://david.ncifcrf.gov/).

To analyse for functional interactions between proteins represented by CIS genes, I also used the STRING tool, and enriched pathways were similarly demonstrated by GO analysis.

CIS Genes Comparative Genomics Analysis using TCGA Datasets

To determine whether genes found to be mutated or transposon CIS in our mice are also genetically / epigenetically altered in patients, we reviewed copy number, methylation and

mutational data for these genes in TCGA datasets using Cbioportal as described above for the other genes.

Chapter Three: Evolution of *EGFRvIII*-induced Gliomas in Mice

Abstract

EGFR is one of the most frequently mutated and amplified genes in gliomas, which represent the commonest type of intrinsic brain tumor. It is unclear at what stage of gliomagenesis these alterations are acquired and the effect they have on the genetic evolution of cancer. Gliomas also occur in the spinal cord, but the genetics of these tumors and how they compare with their brain counterparts are poorly understood. Here, we set out to determine how a common activating *EGFR* mutation (*EGFRvIII*) influences glioma genetic evolution by leveraging mouse genomics. We expressed *EGFRvIII* in the mouse central nervous system, and found that this is sufficient to initiate glioma formation both in the brain and spinal cord with long latency. Whole-exome sequencing of resultant tumors revealed the secondary molecular alterations spontaneously acquired after *EGFRvIII*-tumor intiation, including amplification of *EGFRvIII*, deletions of *Cdkn2a* and *Nlrp1b*, and mutations of *Trp53*, *Tead2*, *Sub1* and *Nt5c2*. Transcriptomic profiling through RNA-sequencing of these tumors revealed enrichment for gene sets in multiple pathways, including Wnt, MAPK, p53, JAK-STAT and stem cell related pathways. Comparative analysis of these data with human glioma sequencing data demonstrates recurrent deletions in *TEAD2* and *NT5C2*, as well as methylation of *Sub1* and *Nlrp1* implicating these genes as putative contributors to human gliomagenesis too. This chapter presents data showing *EGFRvIII* can initiate gliomagenesis *in vivo* and the subsequent genetic alterations somatically acquired in tumorigenesis.

Introduction

EGFR as a glioma driver

EGFR is mutated in up to 60% of *IDH1*-wild type GBMs, and *EGFRvIII* is one of the commonest types of *EGFR* mutation in these tumors. As one would expect, previous analysis of TCGA data showed that *EGFR* mutation and / or amplification is associated with a significant increase in EGFR expression suggesting these genetic changes have functional consequences on the tumor [27]. In fact, this observation that increased *EGFR* expression is associated with its amplification in GBM was first described in 1987 [127]. Previous studies have suggested that *EGFRvIII* (variant III) mutations only cause gliomas in mice in the presence of other predisposing mutations in genes such as *Pten* and *Cdkn2a* [92, 96, 128]. These studies typically expressed the mutation(s) in specific locations, such as the basal ganglia or cerebral cortex, reflecting the location of cre injection. Studies have since implicated a role for the subventricular zone (SVZ) in the earliest phase of glioma formation, although this may be context-dependent on the genetic background; indeed the majority of studies that demonstrate SVZ as a glioma site of origin have used *Trp53* as a predisposing mutation. Another issue is that the observation times of these earlier studies were generally short (eg 12 weeks in the case of [29]), meaning tumors that arise after long latency were not detected.

In the TCGA cohort, GBMs have a somatic mutation rate in *EGFR* of 32.4% [27]. A review of the locations of the point mutations within the *EGFR* gene demonstrates that they are clustered either within the extracellular receptor domains or in the intracellular tyrosine kinase domains, Fig 3.1. This is consistent with the idea that (at least some of) these mutations are activating mutations, that either switch on the tyrosine kinase domain or the ligand-binding domain.

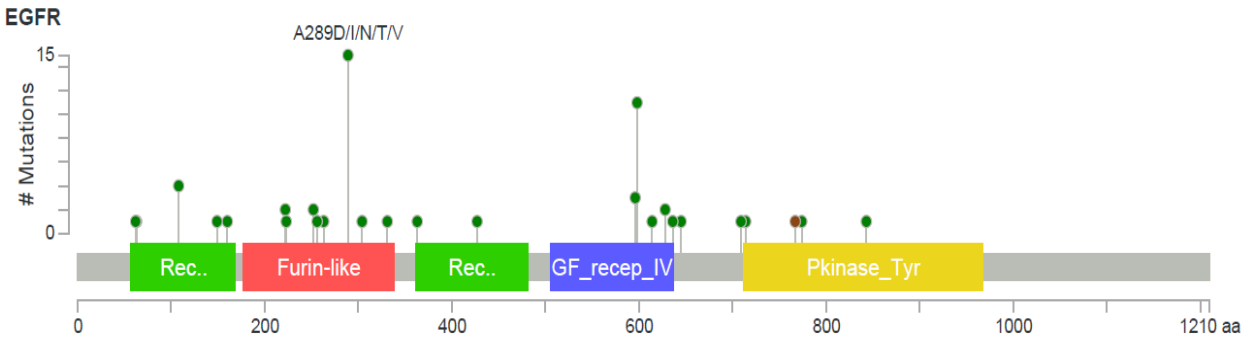


Figure 3.1. Genetic alterations in *EGFR* in human GBMs, data from TCGA. Bottom panel details the location in the gene where mutations are located; the height of the pin represents the number of tumors with a mutation in that locus. Data extracted using the publicly available software, CBioportal (see Materials and Methods). Green block = receptor ligand domain; red block = furin-like cysteine rich region; blue block = growth factor receptor domain IV; yellow block = protein tyrosine kinase.

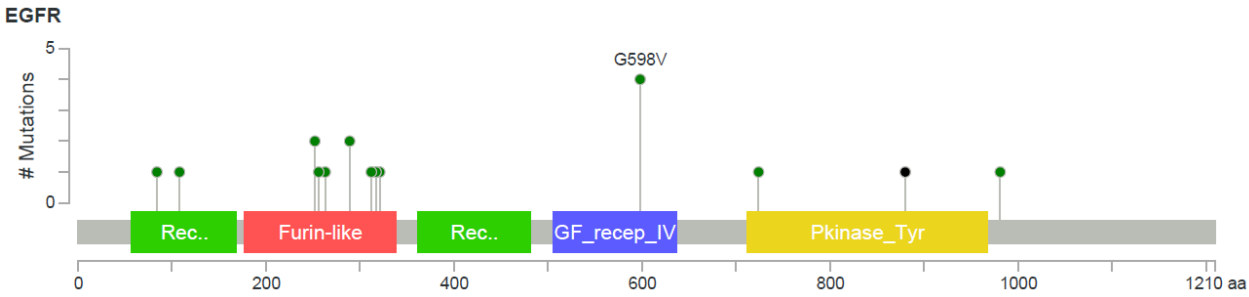


Figure 3.2. Analysis of genetic alterations of *EGFR* in human low grade gliomas from the TCGA cohort. Bottom panel displays the location of mutations within *EGFR*, the majority of which occur in the furin-like domain although there is a prominent mutation (G598V) in the growth factor receptor domain.

Analysis of TCGA data from 530 human low grade gliomas identifies a difference in *EGFR* mutational rates compared with GBMs: the *EGFR* somatic mutation rate in LGGs was 5.3% in

this cohort [93]. Similar to GBMs, the vast majority of the mutations occurred either in the extracellular ligand-binding domains or the tyrosine kinase domain, Fig 3.2. However, these data do not include extrachromosomal *EGFR* mutations and mutations which are known to be present (therefore increasing the prevalence of *EGFR* driver mutations) in both LGGs and GBMs [129].

Mechanistically, *EGFRvIII* is a constitutively active form of *EGFR* and signals predominantly via the PI3K-Akt-mTOR pathway, whose ultimate effects include increased cellular proliferation [130, 131]; thus, this can explain why this mutation would predispose to tumor formation on its own, particularly if the activation of the PI3K-Akt pathway were very strong. However, a counter argument is that constitutive activation of EGFR may trigger senescence, thereby avoiding cancer formation[92]. Intriguingly, previous work has demonstrated that *EGFRvIII* may not activate the Ras-MAPK pathway as strongly as PI3K-Akt signalling, unlike its wildtype counterpart which can activate both PI3K-Akt and Ras-MAPK pathways strongly in the presence of its ligand epidermal growth factor (EGF). Glioma cells can express both wild-type *EGFR* and *EGFRvIII*, in which case EGFR phosphorylates EGFRvIII and this leads to enhanced STAT3 signalling which may increase tumorigenicity even further [132]. EGFRvIII can activate the c-MET receptor tyrosine kinase that can stimulate cell proliferation too, such that inhibiting both EGFR and c-MET together may have a strong anti-proliferative effect [133]. It has been reported that the oncogenic mutant EGFRvIII protein can be transferred between glioma cells via extracellular vesicles, thus enabling oncogenic pathway activation in neighbouring cells [134].

Human GBMs, although they frequently carry the *EGFRvIII* mutation, display substantial intra- and inter-tumor heterogeneity, with many cells not expressing the mutant protein although the tumor as a whole carries the mutation[135]. This makes it challenging to determine whether *EGFRvIII* is an initiating tumorigenic event or whether it emerges late in gliomagenesis and then establishes itself as a dominant genetic driver. However, studies have

demonstrated that mutant *EGFR* provides a selective growth advantage for glioma cells *in vivo*, specifically in maintaining glioma growth following tumor initiation [136]. Whether *EGFRvIII* can initiate glioma formation and the subsequent genetic alterations driving tumor formation are still unknown.

EGFR targeted therapies for gliomas

There are a number of targeted therapies against the epidermal growth factor receptor (EGFR, either already in clinical use or in development. Thus, it is worth considering key issues surrounding these approaches here (and these are also discussed further in the Discussion Chapter). EGFR inhibitors have been demonstrated to be beneficial in certain cancers carrying an *EGFR* mutation, such as lung and colon cancers where there is a clear survival benefit associated with this treatment in subsets of patients [137]. The original EGFR inhibitors like erlotinib and gefitinib are classed as reversible, in that mutations in the gene can easily lead to tumor resistance [138-141]; as a result, newer irreversible inhibitors, such as afatinib, have been developed that bind the receptor more strongly. Early clinical trials using EGFR inhibitors in patients with gliomas demonstrated evidence of tumor regression with this treatment, particularly in tumors also containing *PTEN* co-expression [142]. However, larger trials have not demonstrated an improvement in survival with this treatment [143]. Potential reasons for this are that gliomas carrying *EGFR* mutations are no longer dependent on these for their growth and instead are dependent on other acquired mutations, and / or that these drugs do not completely block EGFR signalling [144]. Indeed, a recent study in mice demonstrated that genetic ablation of *Egfr* gave a stronger tumor inhibitory effect than EGFR inhibitors, suggesting complete suppression of the receptor is important for slowing tumor growth [128]. Given the potential for resistance to this form of therapy, studies are investigating the role of additional therapies to complement EGFR inhibitors.

Immunotherapy is a promising area within cancer research, which has revolutionized treatment of malignant melanoma for example. A 13-amino acid peptide vaccine, rindopepimut, based on the EGFRvIII protein has been generated and is being studied in

randomised controlled trials after showing benefit in mouse models and in small glioblastoma patient cohorts [145]. This method relies upon the body's immune system to trigger a response to the foreign EGFRvIII peptide, which contains a novel surface glycine residue that is not normally present on the wild-type EGFR and is therefore an immunogenic epitope. The precise component of the immune system (whether antibodies or T-cells) that is responsible for removal of the tumor cells is still unknown however. In some pre-clinical and clinical studies, rindopepimut is also being injected subcutaneously with GM-CSF to enhance the immunogenic response. Antibodies against EGFR have been described in experimental models of glioma, with some promising results [146].

The *EGFRvIII* mutation has been implicated in resistance to radiotherapy in GBM, potentially through conferring an increased rate of double strand break repair compared with normal brain which mitigates the effect of radiotherapy on tumor cell killing [147, 148]. Resistance to EGFR tyrosine kinase inhibitors (TKIs) may occur through multiple mechanisms. A possible such mechanism is maintenance of mutant *EGFR* on extrachromosomal DNA, and following treatment with *EGFR* TKIs there is elimination of mutant *EGFR* from extrachromosomal DNA to give treatment resistance; after withdrawal of treatment, mutant *EGFR* may reappear on extrachromosomal DNA to drive tumor re-expansion [149]. Amplified copies of *EGFR* may also be contained on double-minute chromosomes [150]. Alternatively, *EGFRvIII* may transcriptionally suppress alternative receptors that activate similar pathways, such as PDGFRβ; treatment with EGFR TKIs may then increase transcription of *PDGFR*β that can then drive tumor growth [151]. Other EGFR-related members of the ERBB protein family may similarly become activated in glioma stem cells if EGFR is inhibited ([152]. *EGFRvIII* itself may drive resistance to erlotinib when used to treat tumors with an *EGFR* amplification: in this situation, *EGFRvIII* becomes upregulated resulting in an increase in PI3K signalling via increased expression of PI3Kp110ð (a regulatory subunit of PI3K) [153]. There are important effects of *EGFRvIII* on transcriptional programs as well, for example acting via transcription factors SOX9 and FOXG1[154]. *EGFR* mutants are also likely to enhance tumor invasion into normal brain, as it has been demonstrated to upregulate enzymatic effectors of invasion such as metalloproteases and serine proteases [155].

Although these novel therapeutics are generating much interest in EGFR as a clinical target, this molecule has thus far proven to be an unsuccessful target in the treatment of glioma patients. Potential reasons for this are an incomplete understanding of the biology of this gene, and particularly of its interactions with other molecules in a tumor such as cooperativity between EGFRvIII and other proteins.

Aims of Study

Whole-genome sequencing studies of human brain gliomas have shed much light on the genetic and epigenetic landscapes of these tumor types[28, 30, 93, 156]′[157]. In addition to mutations, driver genes may be altered through transcriptional, methylation or large copy number changes, and these are more difficult to identify as cancer drivers. Another complicating issue is that mutations in individual tumors can occur in different combinations, which can affect prognosis and response to therapy[1, 158], but this makes it more difficult to confidently identify which genes are truly collaborating with one another. Given the extensive intra-tumor heterogeneity of end-stage tumors, the timing of key mutations acquired during the natural history of gliomas and how these driver mutations influence tumor genetic evolution cannot easily be inferred from human genomic studies.

Activating mutations in the epidermal growth factor receptor (*EGFR*) occur in up to 60% of *IDH1*-wild-type GBMs[28] of which *EGFRvIII* is the most common (an in-frame deletion of exon 2 to 7 in the extracellular domain leading to constitutive receptor activation[96, 133]). Frequent mutations and amplifications of *EGFR* (including extrachromosomal ones) have recently been detected in *IDH1*-wild-type, histologically low-grade gliomas (LGGs)[129, 159], highlighting a need for integrated molecular diagnosis. In addition to brain tumors, studies on small cohorts of patients have identified *EGFR* amplification and expression in spinal gliomas, particularly in leptomeningeal-disseminated paediatric LGGs [160, 161], suggesting increased EGFR signalling may promote tumorigenesis in a subset of spinal gliomas. However, the timing of *EGFR* mutations in gliomagenesis, their role in spinal gliomas and their cooperative genetic lesions remain largely unknown. In particular, the genetic drivers of spinal gliomas and how they compare with their brain counterparts are obscure[162]. Understanding the functional genomic landscapes of gliomas is therefore of the utmost importance and will help us decipher human patient glioma genomes.

Here, we aimed to address these challenges by investigating the genetics of gliomas from mice expressing a constitutively-active *EGFR* mutation (*EGFRvIII*) in the central nervous system under control of the nestin-promoter. We show that *EGFRvIII* is sufficient to initiate gliomagenesis from the normal mouse brain and spinal cord with long latency in this model. By combining whole-exome sequencing, transcriptomics, and genome-wide *piggyBac* transposon mutagenesis (discussed further in Chapter Four), we identified recurrent mutations in known and novel putative glioma genes and characterized the functional genomic landscapes of *EGFR*-mutant brain and spinal gliomas in mice.

Results

Here we aimed to study the role of *EGFRvIII* in gliomagenesis. For this, we generated double heterozygous mice carrying a conditional human *EGFRvIII* transgene (integrated in the Col1a1 locus, chromosome 11) [96] and expressing cre under the control of the Nestin promoter[87] (*nes*-cre), which activates *EGFRvIII* expression primarily in the central nervous system. It has been previously shown that mice with the nestin-cre allele express cre from embryonic day 13 throughout most of the central nervous system, eye and also the kidneys – this was demonstrated by Dubois and colleagues who showed almost complete cre-mediated recombination in these tissues by embryonic day 15.5 using LacZ based reporters [88], except in early embryonic ventricular zone neural progenitors and neural stem cells in which recombination is complete by embryonic day 17.5) [89]. The outline of the experiment for our study is shown in Fig 3.3.

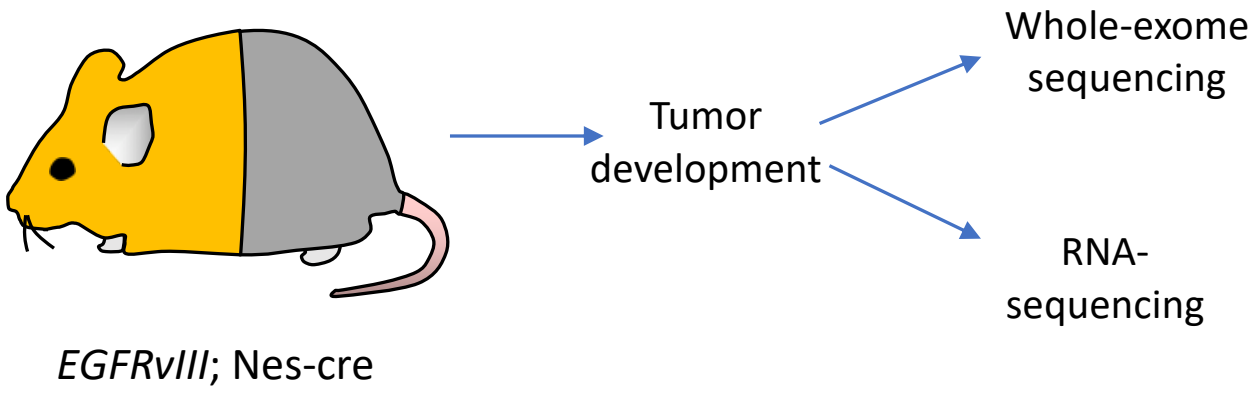


Fig 3.3. Outline of the experiment: *EGFRvIII* was conditionally expressed throughout the central nervous system using *nes-cre*, and resulting tumors were subjected to whole-exome sequencing and RNA-sequencing.

Eye lesions in *EGFRvIII*; *nes*-cre mice

EGFRvIII; *nes*-cre mice started developing lesions within the eye that were clinically apparent from around 7 weeks of age. These lesions typically eventually affect both eyes with 100% penetrance. The complex eye phenotype presented with a number of features, including one or more of: cataract, hyphema, secondary glaucoma and proptosis, Fig 3.4. Histology revealed abnormal neovascularisation in the retina associated with chronic micro-haemorrhages and occasionally larger bleeds (n=24 eyes). In all cases, there was either partial or complete degeneration of the lens. There was also a proliferation within the ciliary body in many cases; although this could be mistaken for an ocular melanoma, the cytology of the proliferation does not match that of a melanoma. It is likely therefore that the ciliary body proliferation is a secondary reaction to chronic haemorrhages within the eye, rather than presenting a tumor or tumor precursor. No such lesions were observed histologically in age-matched control eyes from wild-type mice (with EGFRvIII but lacking nestin-cre; n=10 eyes). However, these lesions were not the focus of this project so these data are not presented here.

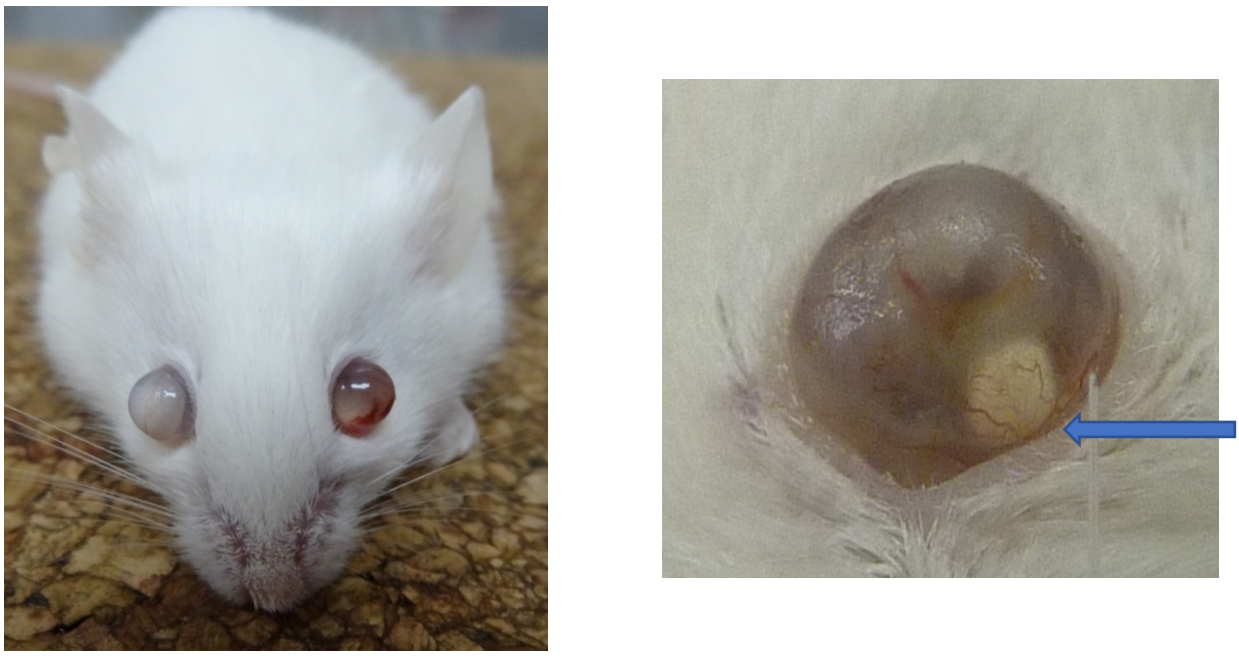


Figure 3.4. Eye lesions in *EGFRvIII*; *nes*-cre mice. Left panel shows proptosis of eyes and small haemorrhages in left eye. Right panel shows apoptotic lens, confirmed with histology (histology not shown).

Clinical Phenotypes of Mice

EGFRvIII; *nes*-cre mice started developing signs of neurological disease from 14 weeks onwards. These include progressive macrocephaly (enlargement of the head, Fig 3.5), lateral leaning behaviours or a head tilt, circling, weakness of limbs, and seizures. At later stages, the mice displayed a combination of these signs, and the vast majority also display the eye phenotype with a degenerate lens. Mice required culling when the phenotype restricted their basic functioning, such as inability to mobilise and therefore to feed. Records were kept of the age at which these mice were culled and of their clinical phenotype. Control mice with *nes*-cre but not *EGFRvIII* did not display any signs of neurological disease after one year of observation.

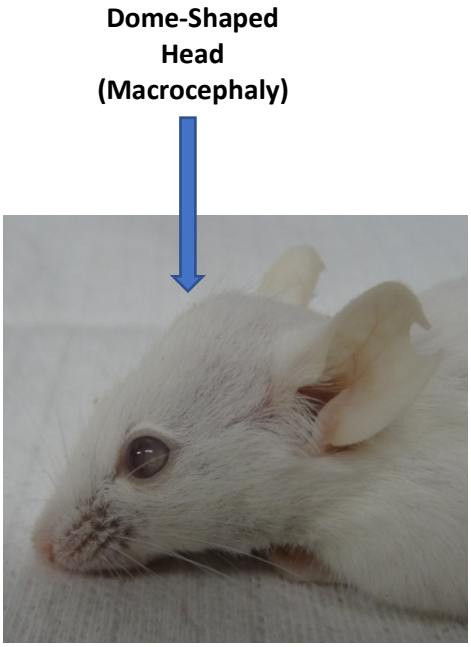


Figure 3.5. Clinical phenotypes of mice expressing the *EGFRvIII* allele in nestin-expressing tissues. The phenotypes are primarily neurological; left panel shows a typical example of macrocephaly due to hydrocephalus.

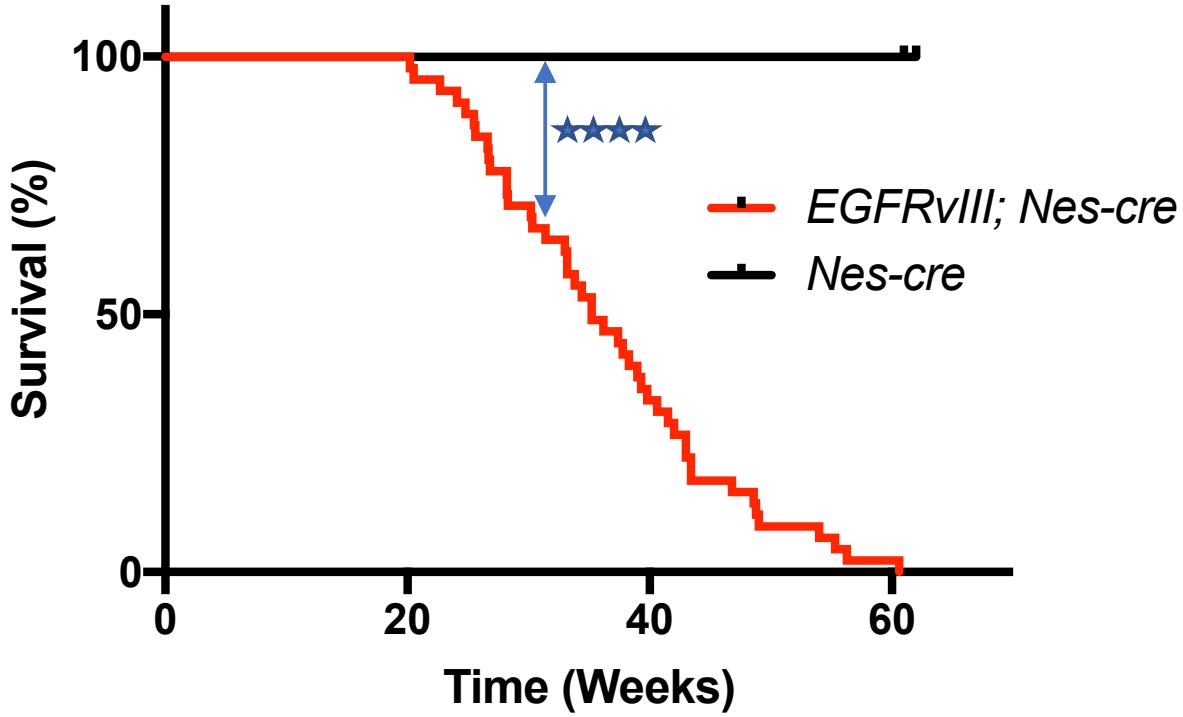


Fig 3.6. Kaplan-Meier plot of *EGFRvIII ; nes-cre* mice and control (*nes-cre*) mice (**** denotes $p < 0.0001$, log-rank test, n=31 and n=10 mice respectively). Control mice with *nes*-cre but lacking *EGFRvIII* did not display signs of neurological disease after one year of observation.

EGFRvIII **initiates gliomagenesis**

By 60 weeks of age, 100% of mice had succumbed to brain and/or spinal tumors (n=31), Fig 3.6. Spinal tumors will be described in the next sub-section.

Pathological examination of brains prior to clinically overt disease (mice aged 12-28-weeks) revealed small glioma precursor lesions with proliferative activity (as indicated by immunohistochemical staining with Ki67) – these lesions are also described as 'microneoplasias' (13/13 mice) and have also been reported in mice from different genetic contexts before[97, 163]. The size of each microneoplasia was between 100 and 200µm. Multiple microneoplasias were detected bilaterally protruding into the lateral ventricles, third ventricle and from the brain surface, and they had subpopulations of proliferating cells as detected by Ki67 immunostaining, Fig 3.7 and Fig 3.9. In contrast, no such lesions were seen

in 5 control mice carrying only the *nes*-cre allele (age 10-30 weeks; p = 0.0001, two-sided Fisher's exact test), Fig 3.8. The control and experimental brain samples were examined in exactly the same way by our Consultant Neuropathologist, with four standardised cut sections from each brain, as described in the Materials and Methods. Importantly, Professor Sebastian Brandner (Consultant Neuropathologist) provided histological diagnoses for all samples in this study. Using immunohistochemical staining, we demonstrated these microneoplasias expressed protein markers of neural stem cells and transit-amplifying cells, specifically Sox2, Nestin, PDGFRa, GFAP and Olig2, Fig 3.10 and 3.11.

Next, we examined mice that were culled following development of clinical signs of underlying disease. *EGFRvIII*; *nes*-cre mice displayed neurological signs due to one or multiple gliomas within the lateral ventricles and / or brain surface with histological evidence of subarachnoid involvement (26/31 mice had brain gliomas; mean survival 36.2 weeks), Fig 3.12. In order to confirm *EGFRvIII* recombination had occurred specifically in tumor cells from microneoplasias and gliomas, we performed immunostaining for human EGFR and EGFRvIII in these mouse brains which demonstrated strong EGFR and EGFRvIII expression specifically in gliomas and their precursors but not in normal brain (5/5 tumors positive for EGFR immunostaining; 4/4 tumors positive for EGFRvIII immunostaining), Fig 3.13 and 3.14. Histopathological analysis by a Consultant Neuropathologist revealed these tumors had histological features and expressed protein markers comparable to those of human gliomas, Fig 3.15. The cells of these tumors have relatively monomorphic round tumor cell nuclei of a glial nature. The tumors displayed small lakes of myxoid matrix, similar to those observed in human astrocytomas or oligodendrogliomas, and range in size from 200μm upwards. Although the majority were histologically LGGs, a small proportion also displayed necrosis and microvascular proliferation (endothelial hyperplasia) that is characteristic of human GBMs, Fig 3.16. The grading system we used to define glioma grades is described in the Materials and Methods. Tables 3.1 and3.2 show all mice in this study and their associated pathologies.

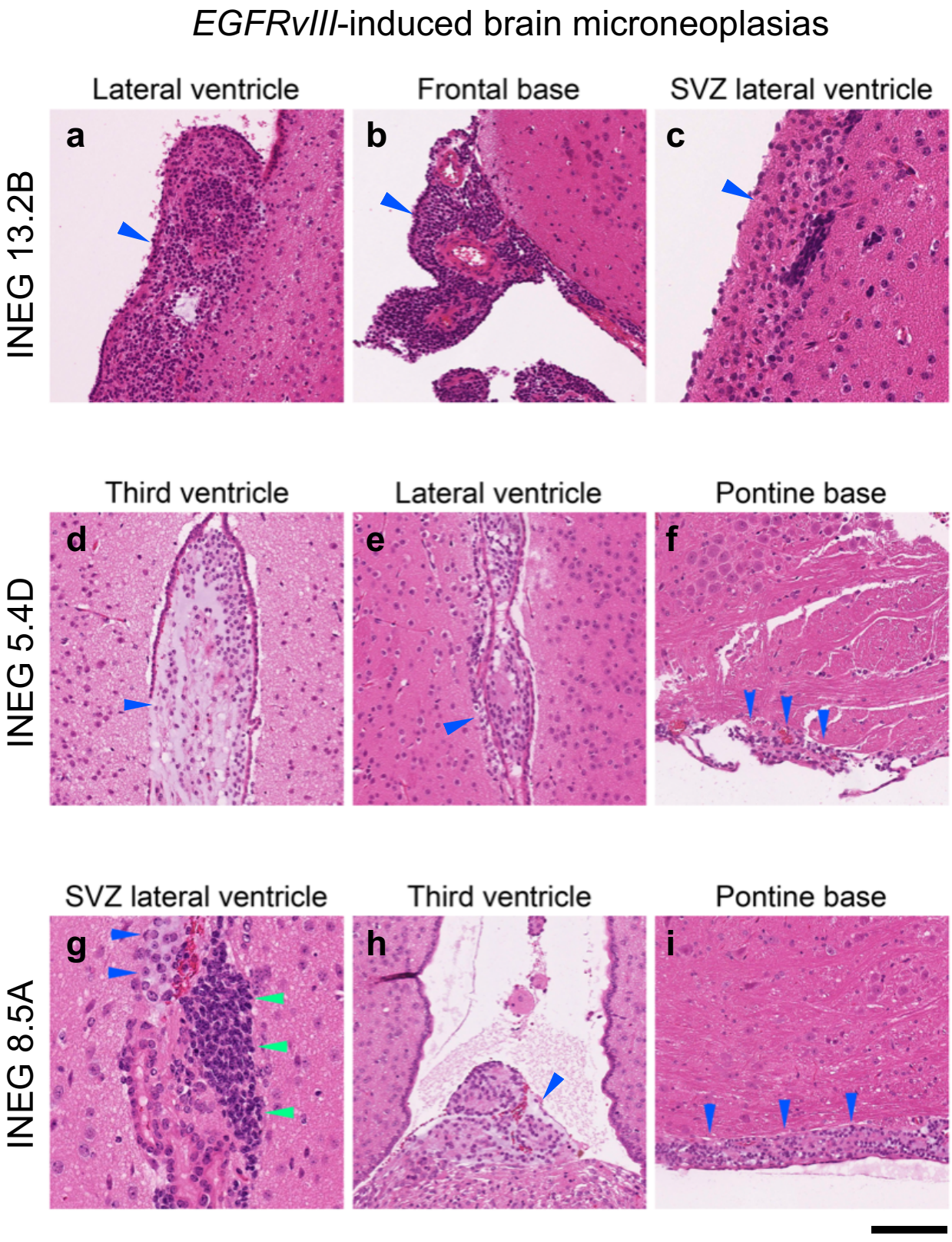
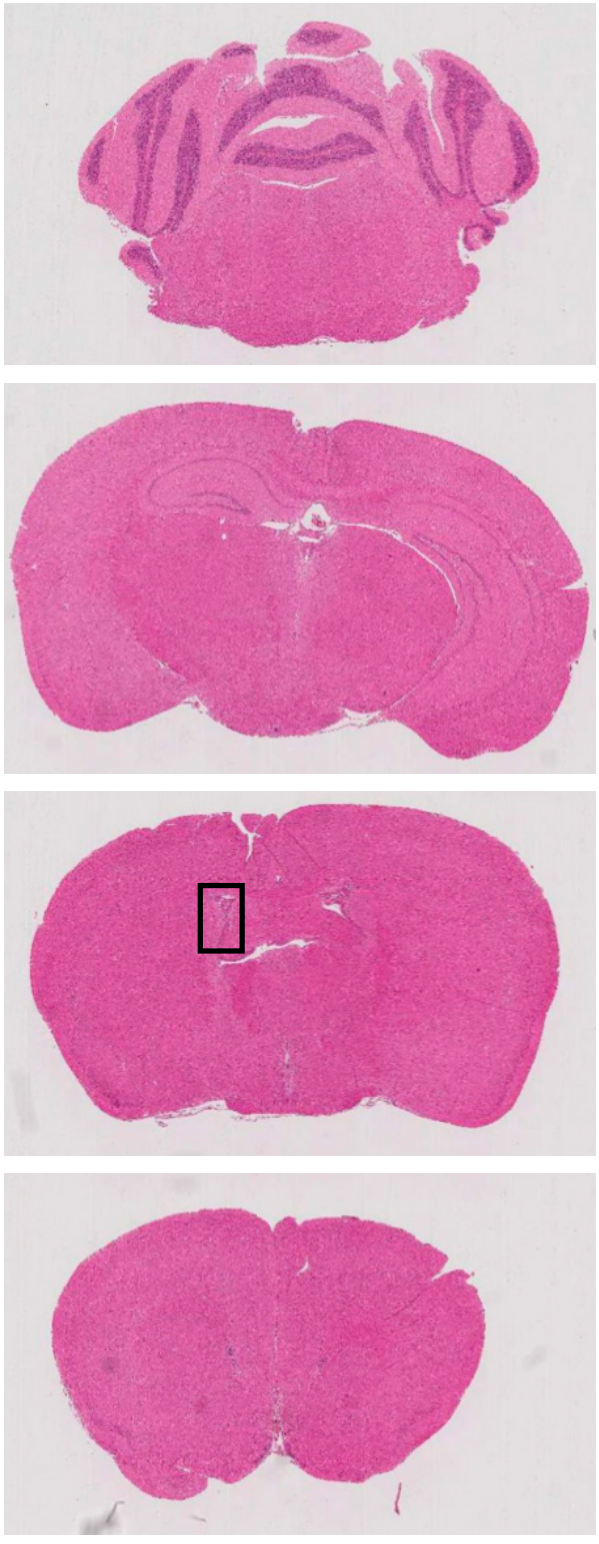

EGFRvIII-induced brain microneoplasias
INEG 13.2B
Lateral ventricle
Frontal base
SVZ lateral ventricle
a
b
c
INEG 5.4D
Third ventricle
Lateral ventricle
Pontine base
d
e
f
INEG 8.5A
SVZ lateral ventricle
Third ventricle
Pontine base
g
h
i

Fig 3.7. Microneoplasias in EGFRvIII mouse brains. Examples of the formation of small tumors in the ventricular system and subarachnoid space. A, B, C, tumor growth in the lateral ventricle, the base of the frontal brain and the subventricular zone (SVZ) adjacent to the lateral ventricle. D, E, F, formation of the hypercellular myxoid intrinsic tumor in the third ventricle (D) the lateral ventricle (E) and the base of the pons (F, arrows). G, H, I, hypercellular cluster (dark nuclei of expanded SVZ stem/progenitor cells (green arrows) and adjacent a small glial neoplasm (blue arrows). H, small glioma protruding from the floor of the 3rd ventricle and I, subarachnoid spread of a glial neoplasm on the base of the pons, in a "sugarcoat" fashion (arrows). Lettering on sides of panels reflect mouse IDs from which these tumor originated. All histology in this Chapter and this Thesis was reviewed by Consultant Neuropathologist, Professor Sebastian Brandner (University College London, Department of Neuropathology).

Posterior



Anterior

Lateral Ventricle

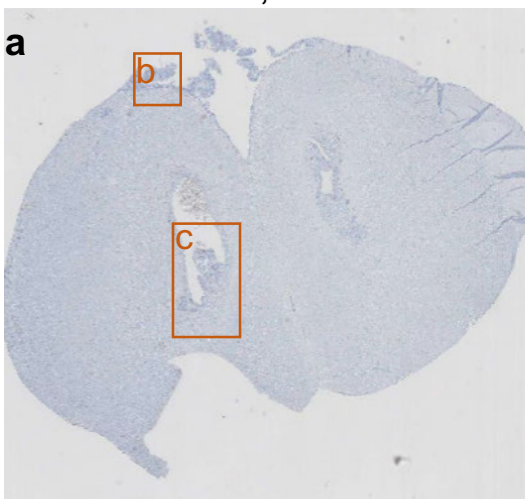
Fig 3.8. H&E stain of a typical control (nes-cre) mouse brain, showing no evidence of a microneoplasia or glioma in any location from 4 independent coronal sections. 5/5 such mice were examined with the same result. Scale bar represents 1mm for left panels and 100μm for the right panel.

Mouse ID	Age (weeks)	Pathology of brain
13.2B	12.9	hydrocephalus, hyperplasia of SVZ with multple cell clusters. Microneoplasia - Budding SVZ growth, small tumor SVZ derived
13.2f	22.1	hydrocephalus, small ventricular glioma
8.2e	24.1	bilateral intraventricular glial neoplasm/ glioma (small - microneoplasia), mild hydrocephalus, skull base microneoplasia
8.3g	21.8	Microneoplasia
31.1h	17.8	Focal clusters in SVZ, microneoplasia
5.5d	20.2	small Glioma / microneoplasia on brain surface, base of brain
5.4d	28.2	small Glioma/ microneoplasia in lateral ventricle and 3rd ventricle
25.2e	18.9	Subventricular cell cluster, small tumors on base of brain and SVZ (microneoplasias).
30.2e	16.2	Microneoplasia on base of brain and cerebellar flocculus
36.1j	21.1	Brain surface microneoplasia
8.3B	28	hydrocephalus, tumor cells (microneoplasia) in corner of lateral ventricle, 3rd ventricle and brain base
20.1e	27.6	Microneoplasia in lateral ventricle and 3rd ventricle, base of the brain.
20.3b	20.2	Microneoplasia in lateral ventricle and 3rd ventricle, base of the brain.

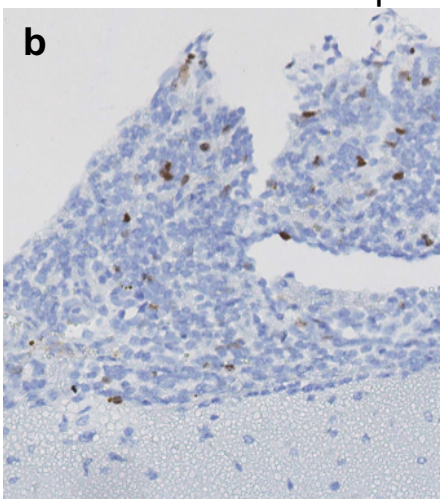
Table 3.1. Clinical and pathological details of *EGFRvIII* ; *nes*-cre mice with brain microneoplasias.

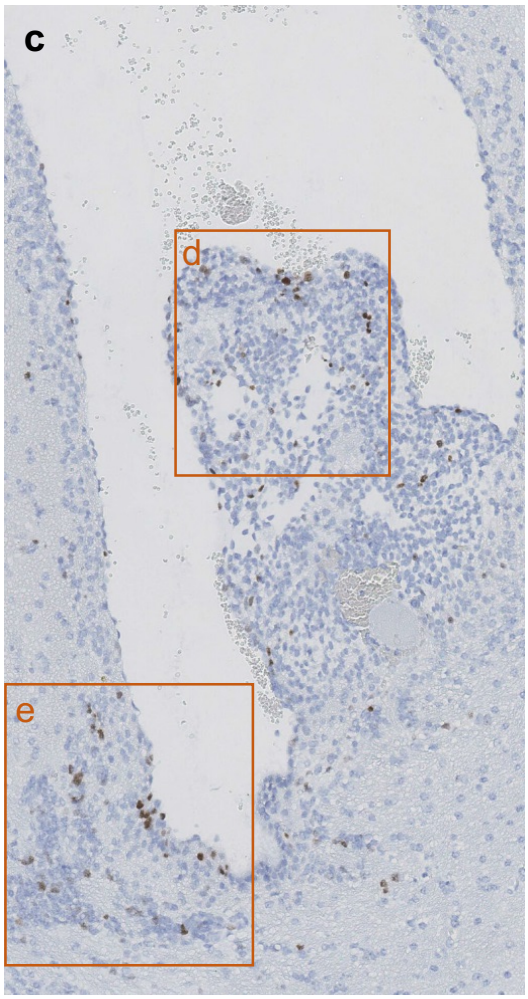
Ki67 Proliferation index of microneoplasias in a brain of an *EGFRvIII* ; Nes-cre mouse

Overview, INEG 13.2B

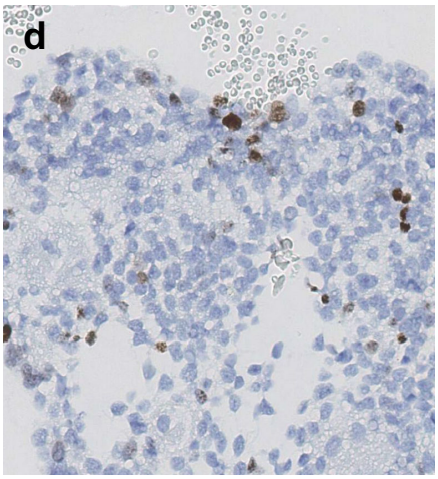


Brain surface microneoplasia

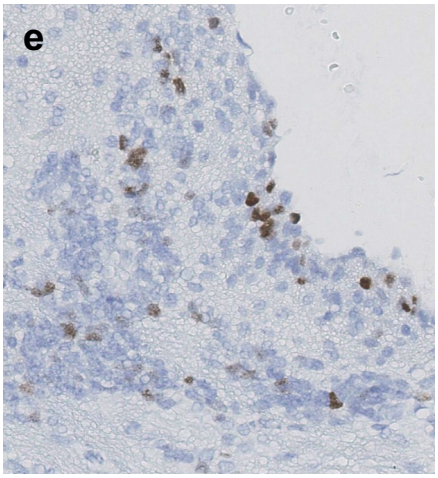




SVZ microneoplasia



SVZ cellular expansion



100

Fig 3.9. Ki67 immunohistochemical staining of early gliomas from *EGFRvIII*-conditional mouse. Positivity for the stain is observed exclusively in a minority of cells in this early glioma and in a nearby SVZ cellular expansion. Ki67, a proliferative marker, stained a small proportion of these cells suggesting these lesions are not characterised by brisk mitotic activity. Scale bar represents 1mm for a, 50 μm for b, d, e, and 200μm for c . 3/3 mice had early tumors showing similar positivity for Ki67.

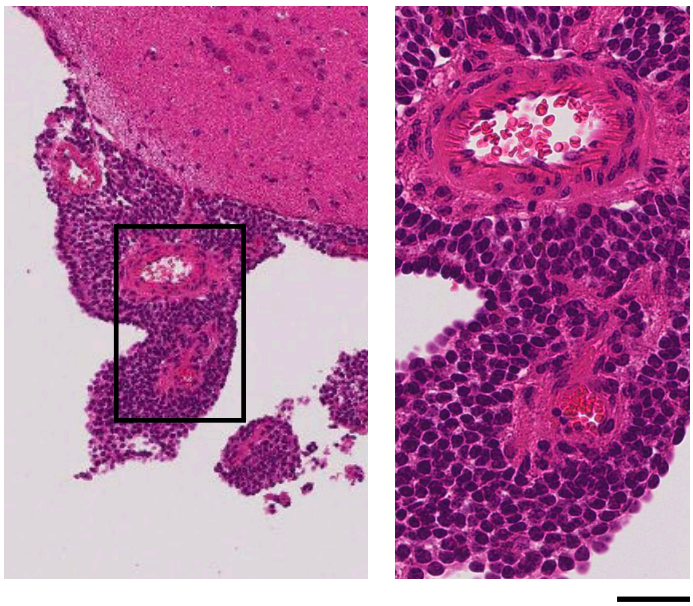


Fig 3.10. Low and high power views of a small glioma (microneoplasia) protruding from the cortical surface of the brain. Scale bar corresponds to 90μm for left panel, and 25μm for right panel.

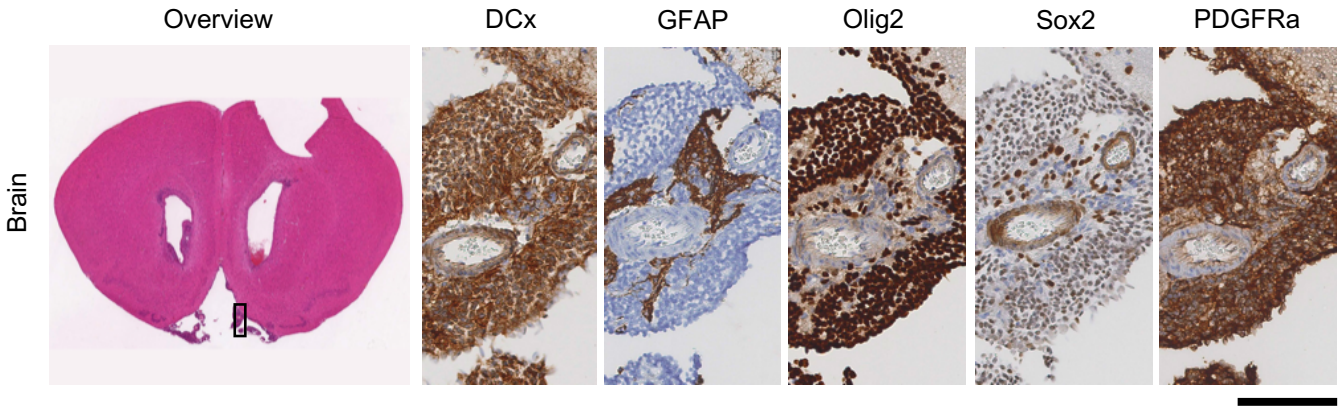


Fig 3.11. Histopathology of brain microneoplasia (left to right): low power view of H&E stain of a brain with a typical microneoplasia (same as in Fig 3.11), and high power view of immunostains of this neoplasm showing positivity for neural lineage markers double-cortin (DCx), GFAP (reflecting reactive astrocytes between tumor cells), Olig2, Sox2 and PDGFRa (n=5/5). Scale bar corresponds to 1mm for left H&E panel, 70μm for immunostain panels.

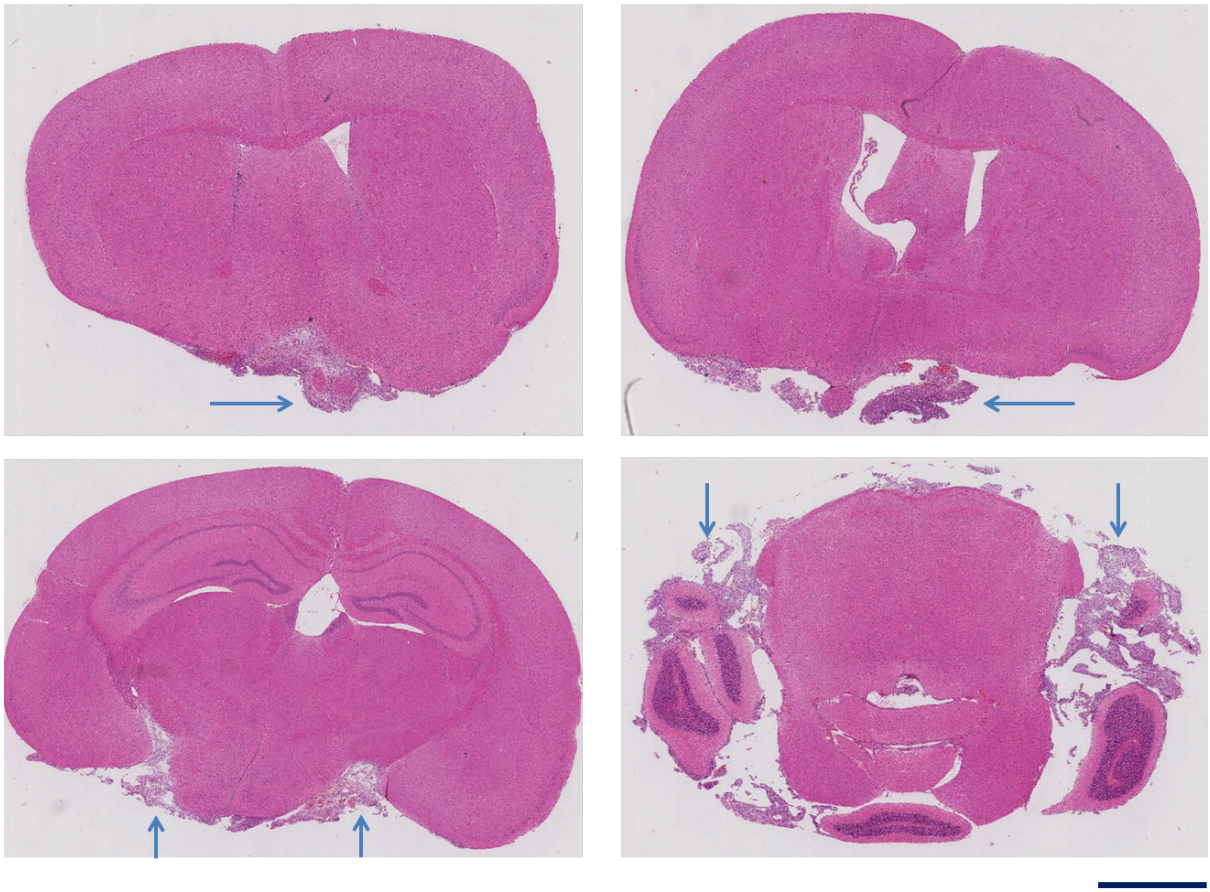


Fig 3.12. Serial sections showing subarachnoid spread of gliomas in *EGFRvIII* mice. Serial slices of a whole brain from one mouse showing dissemination of a glioma on the brain surface (highlighted by arrows); these are H&E stained sections. Scale bar represents 1.2mm.

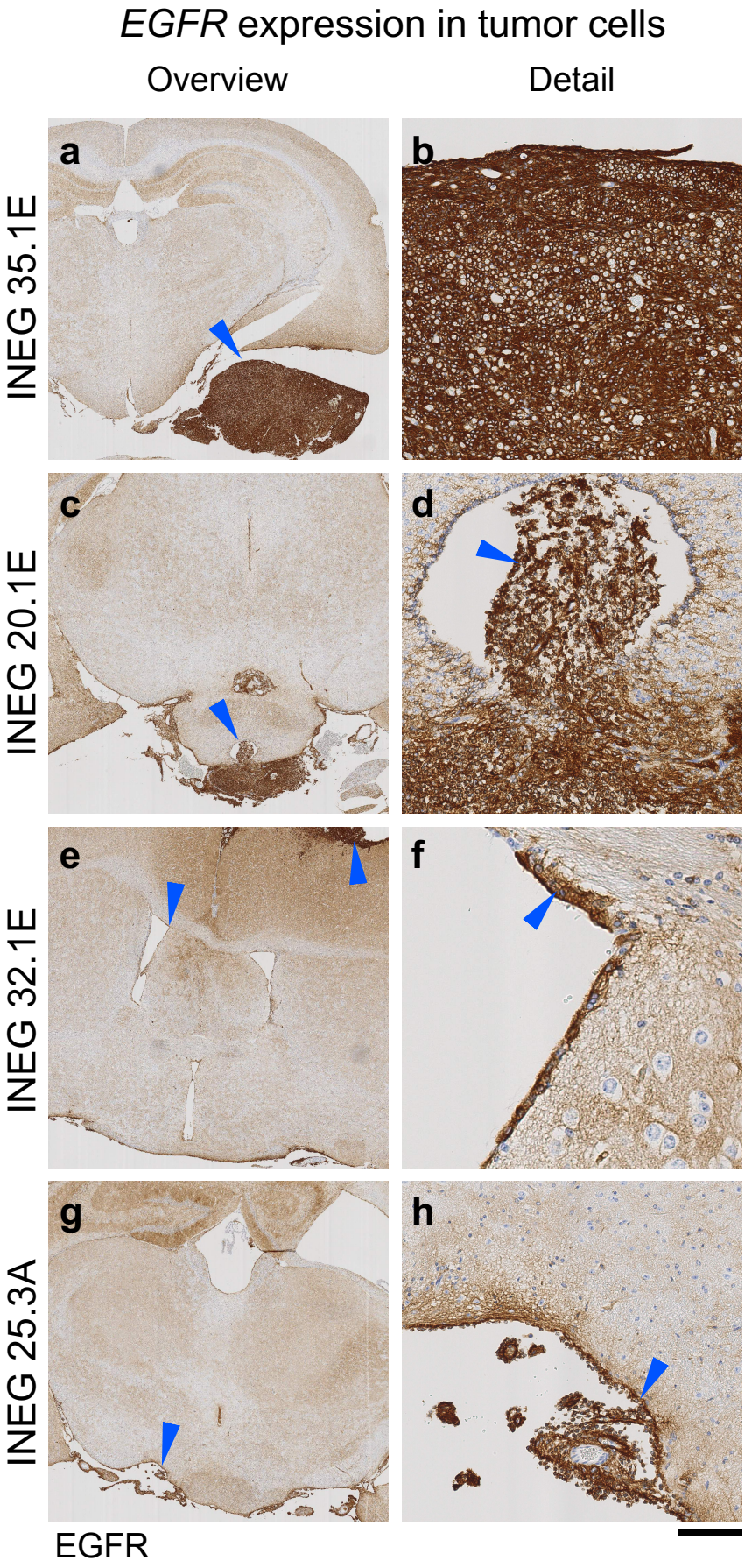

EGFR expression in tumor cells
Overview
Detail
INEG 35.1E
INEG 20.1E
INEG 32.1E
INEG 25.3A
a
b
c
d
e
f
g
h
EGFR

Fig 3.13. Expression of human EGFR, as detected by immunostaining, is limited to tumor cells in *EGFRvIII*; *nes*-cre mouse brains. 5/5 tumors from these mice stained positive for EGFR. Overview (left) and detail (right) of tumors and microneoplasias of different sizes and locations. A, B, medium-sized circumscribed, extra-parenchymal growing tumor attached to the temporal lobe. B, detail showing strong and diffuse EGFR expression specifically in the tumor. C, D, small circumscribed tumor growing on the floor of the third ventricle and expanding towards the optic tract. E, F, Likely transformed cells, with possibly incipient formation of microneoplasia in the left lateral ventricle. G, H, scattered small neoplastic lesion on the floor of the midbrain. Scale bar corresponds to 0.7mm for A, 200μm for B, 0.4mm for C, 200μm for D, 0.7mm for E, 100μm for F, 0.5mm for G, 200μm for H.

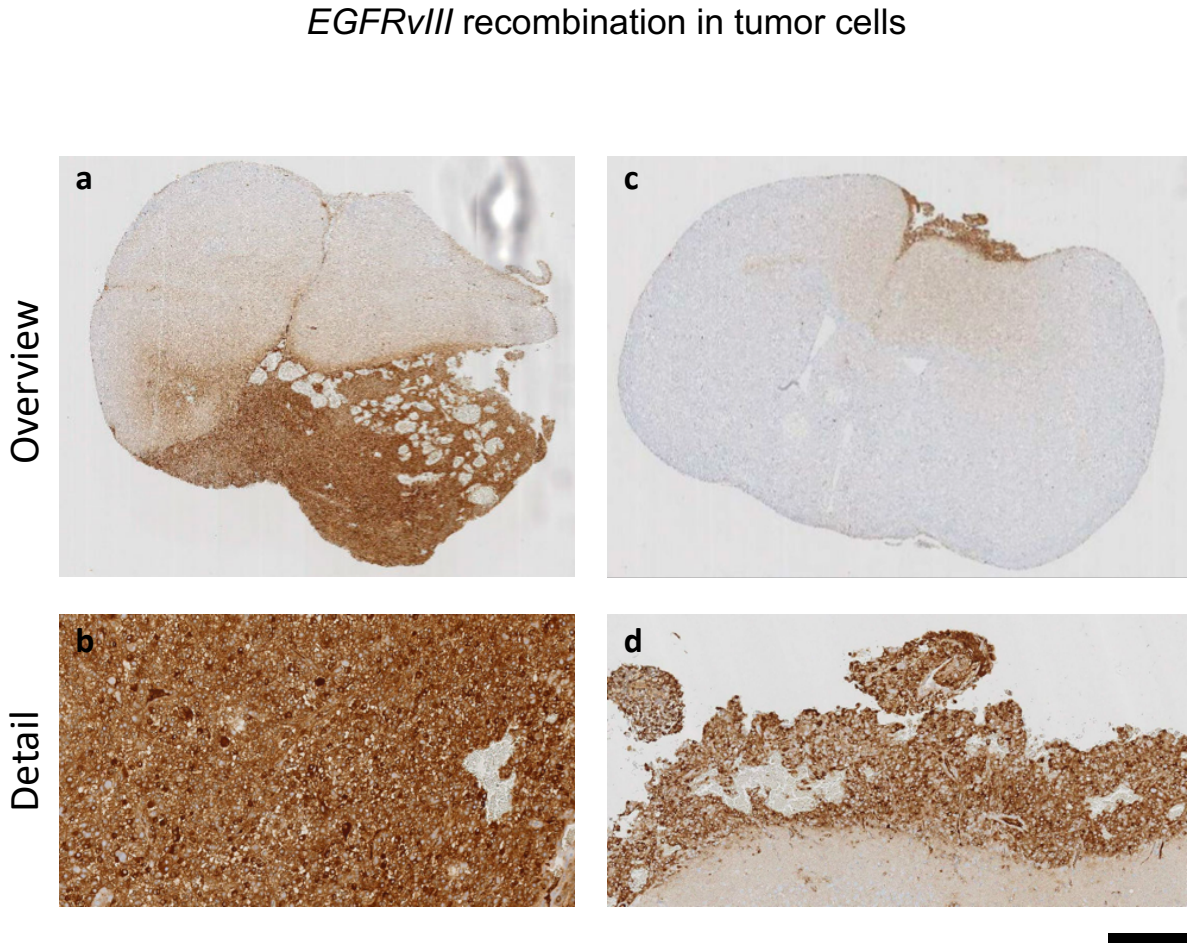


Fig 3.14. Expression of human EGFRvIII is limited to tumor cells. A, B, overview and detail images demonstrating EGFRvIII immunostaining is positive across glioma cells but not normal mouse brain in *EGFRvIII*; *nes*-cre mice (n=4). C, D, overview and detail images demonstrating *EGFRvIII* is expressed in smaller glioma nests (precursors to larger tumors) in these mice. Scale bar corresponds to 1mm for A and C, and 100μm for B and D.

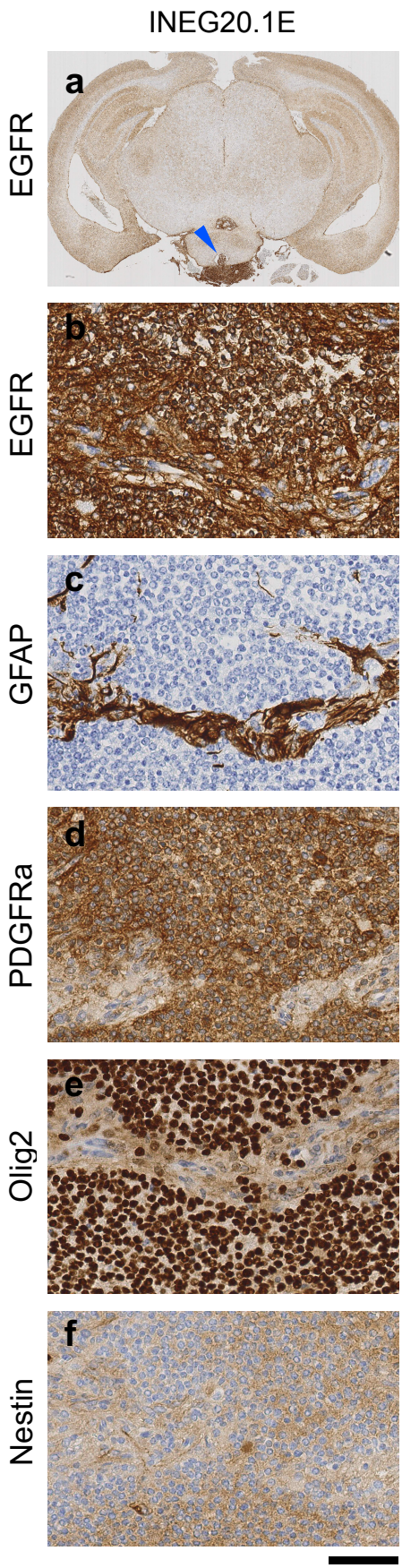

INEG20.1E
a
EGFR
b
EGFR
c
GFAP
d
PDGFRa
e
Olig2
f
Nestin

Fig 3.15. Histopathology of a typical small glioma from an *EGFRvIII*; *nes*-cre mouse. a, overview of the coronal brain section with a circumscribed extracerebral intrinsic tumor, highlighted with an immunostaining for EGFR. b, detail of the tumor, stained for EGFR. c, immunostain for GFAP shows negative tumor cells enclosing a strand of reactive glial tissue. All tumor cells strongly express PDGFRa (d), Olig2 (e), and Nestin (f), which are typical markers expressed by human gliomas. Scale bar corresponds to 50μm for b, c, d, e, f; 1.3mm for a.

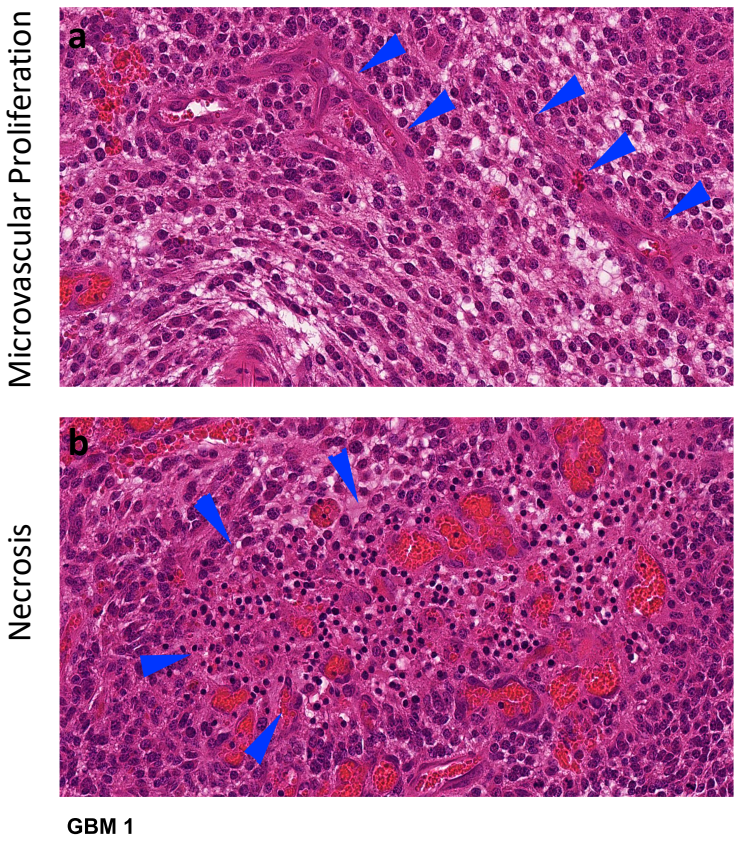

Microvascular Proliferation
a
Necrosis
b
GBM 1
Microvascular Proliferation
c
Necrosis
d
GBM 2

Fig 3.16. Histological features of GBM in *EGFRvIII*-mice. Typical examples showing defining histological features of glioblastoma in two mouse tumors – A, B show microvascular proliferation and necrosis in one GBM, and C, D show these features in another. Scale bar corresponds to 50μm.

Spinal Cord Gliomas

We next examined the spinal cords of *EGFRvIII*; *nes*-cre mice. Aside from brain tumors, *EGFRvIII*; *nes-cre* mice also developed multiple and widespread spinal tumors with 100% penetrance (31/31 mice), which account for the apparent neurological deficits including focal limb weakness. The tumors were located on the spinal cord surface, with evidence of local invasion into the surrounding nerve roots, soft tissue and cranial nerve ganglia. In contrast, most tumors did not show infiltration into the cord: only two *EGFRvIII*; *nes*-cre spinal tumors showed histopathological evidence of spinal cord parenchymal invasion by tumor cells, reminiscent of intramedullary spread of spinal astrocytomas in humans. The spinal tumors were present throughout the leptomeningeal space (in cervical, thoracic and lumbar spine) indicating leptomeningeal-spread, which is a poor prognostic indicator in human patients [164].

In 5/31 mice without established brain tumors (but with microneoplasia), there were still spinal tumors present at all levels of the spines examined (cervical, thoracic and lumbar). Histological examination of the spinal tumors classified them as grade II glioma in all cases (Fig 3.17), even in the presence of grade IV intracranial gliomas, suggesting these are primary spinal gliomas, most likely arising independently. These spinal tumors expressed classical glioma markers, such as GFAP, Sox2, Olig2 and PDGFRa, Fig 3.18.

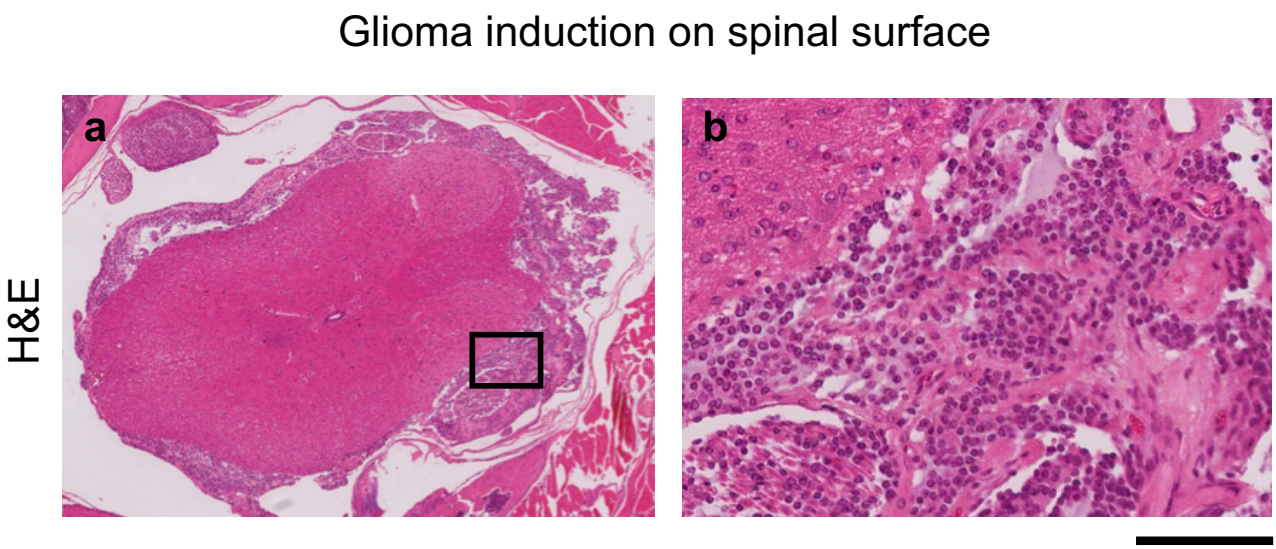


Figure 3.17. Spinal cord glioma initiated by *EGFRvIII*. Haematoxylin and eosin stains. High powered view is presented on the right. Scale bar represents 1mm for a and 100μm for b.

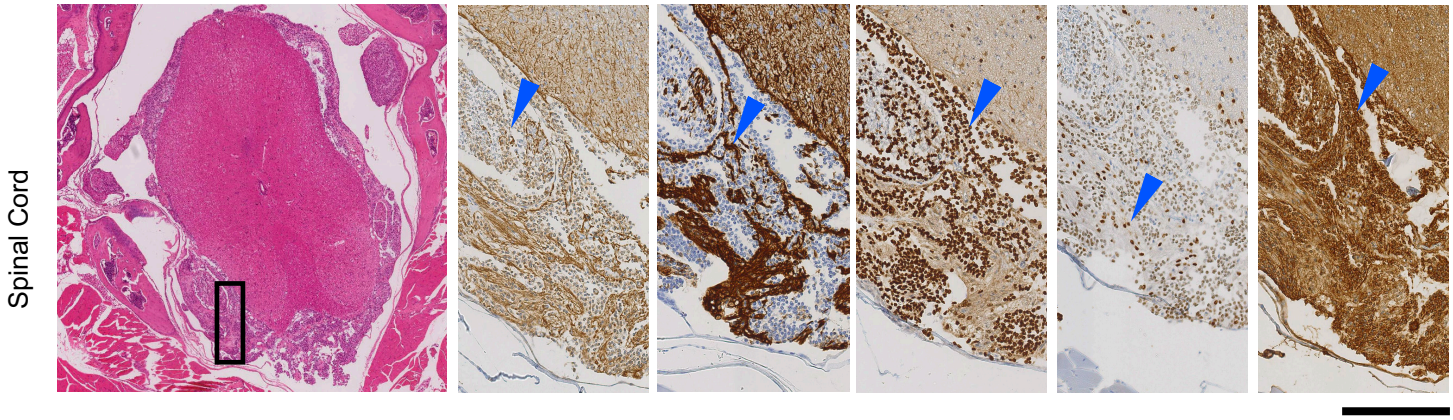


Fig 3.18. Histopathology of a typical spinal glioma in an *EGFRvIII*; *nes*-cre mouse (left to right): low power view of spinal cord with an encasing glioma, and high power views of immunostains of this tumor for neural lineage markers – tumor cells are negative for DCx, reactive astrocytes are positive for GFAP, and tumor cells are positive for Olig2, Sox2 and PDGFRa. Scale bar corresponds to 1mm for left H&E panel, 70μm for upper immunostain panels; 0.7mm for lower H&E panel, 140μm for lower immunostain panels.

Mouse ID	Age (weeks)	Brain Pathology	Spine Pathology
50.1h	34	tumor on lateral ventricle and subarachnoid spread, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
35.2e	23.8	mild hydrocephalus, subarachnoid tumor, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
36.1D	40	SVZ thickening and tumor spread, subarachnoid widespread, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
51.1f	22.7	brain base tumor, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
36.2c	24.8	SVZ thickening and tumor spread, subarachnoid widespread, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
8.2g	41.5	intraventricular and subarachnoid tumor, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
36.2b	37.8	SVZ thickening, widespread thick tumor growth, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
8.5d	33.2	Tumor cells basal and possibly also ventricular, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
30.1b	33	normal SVZ, subarachnoid tumor spread, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
35.1a	33.8	widespread subarachnoidal growth, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
50.1f	26.6	intraventricular tumor growth grade 2/3 glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
35.2a	28.2	extraventricular subarachnoid tumor growth grade 2/3 glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
51.1j	26.7	large extra cerebral tumor grade 2/3 glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
30.1d	34.4	glioblastoma, large, lateral grade 4	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
30.2d	31.4	tumor spread grade 2/3 glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
5.3g	43.4	svz thickening, extra cerebral subdural spread grade 2/3 glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II

50.1a	28.3	grade 4, MVP (microvascular proliferation), necrosis, no infiltration	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
5.4c	42	grade 2, ventricular tumor, bilateral at the base of the brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
Mouse ID	Age (weeks)	Brain Pathology	Spine Pathology
5.5e	39.3	grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
33.1a	37.8	grade 2, Ventricular growth, base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
51.1d	30.2	grade 2, Widespread ventricular and subarachnoidal spread wth base accentuate	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
36.1g	38.3	grade 2, Ventricular growth, base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
5.3d	46.8	grade 2, Ventricular growth, base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
33.1b	50.6	grade 2, Ventricular growth, base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
31.1j	43	grade 3, large intracerebral glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
30.1h	42.2	microneoplasia, small tumor nest at the base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
51.1a	35.2	microneoplasia, small tumor islands at the base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
51.1c	35.2	microneoplasia, intraventricular tumor, small tumor islands at the base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
5.4H	48.8	microneoplasia, small ventricular tumor nests and base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
5.4a	49	microneoplasia, small intraventricular tumor, small tumor islands at the base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
35.2b	39	grade II, ventricular tumor and islands at the base of brain.	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II

Table 3.2. Clinical and pathological details of all EGFRvIII-only mice. Although brain tumors displayed heterogeneity, spinal tumors were pathologically very homogenous.

Primary Cultures

Human glioma cells tend to grow as gliomaspheres (spheres of tumors cells) in neural stem cell media [165]. *EGFRvIII*-expressing mouse tumor cells placed into neural stem cell culture media led to the rapid production (within 1 – 2 days) of gliomaspheres: clusters of tumor cells in suspension, akin to the growth of normal neural stem cells in these culture conditions, Fig 3.19. The cells continued to proliferate (measured up to 8 passages) and the cultures required splitting approximately twice weekly. Gliomasphere cultures were produced for 8 / 8 mouse tumors.

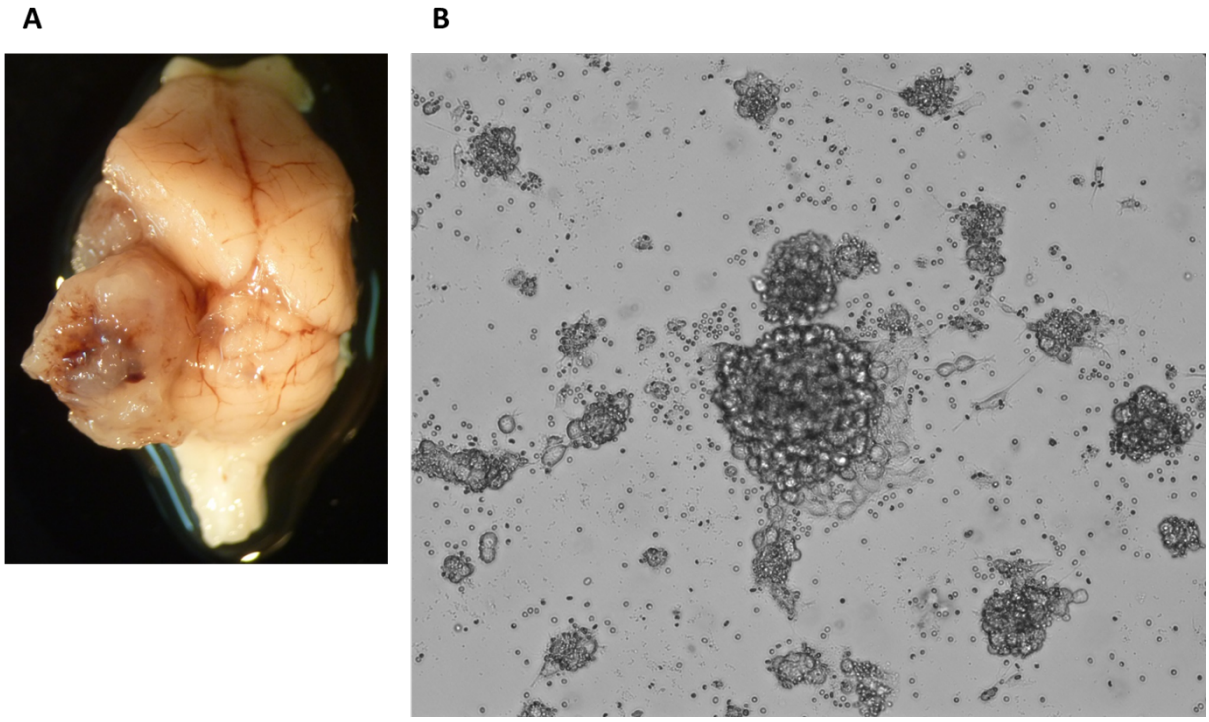


Fig 3.19. Establishing primary cultures from *EGFRvIII*-expressing mouse gliomas. A shows a large glioma on brain surface, which was cultured in neural stem cell media to yield gliomaspheres as shown in B.

Cytogenetic Analysis

Although there was some heterogeneity between the samples, in general these tumors were diploid but with extensive polyploid components. M-FISH karyotyping was performed for 3 *EGFRvIII*-only tumors. Cytogenetic analysis is shown for two representative tumor samples, Fig 3.20 and 3.21.

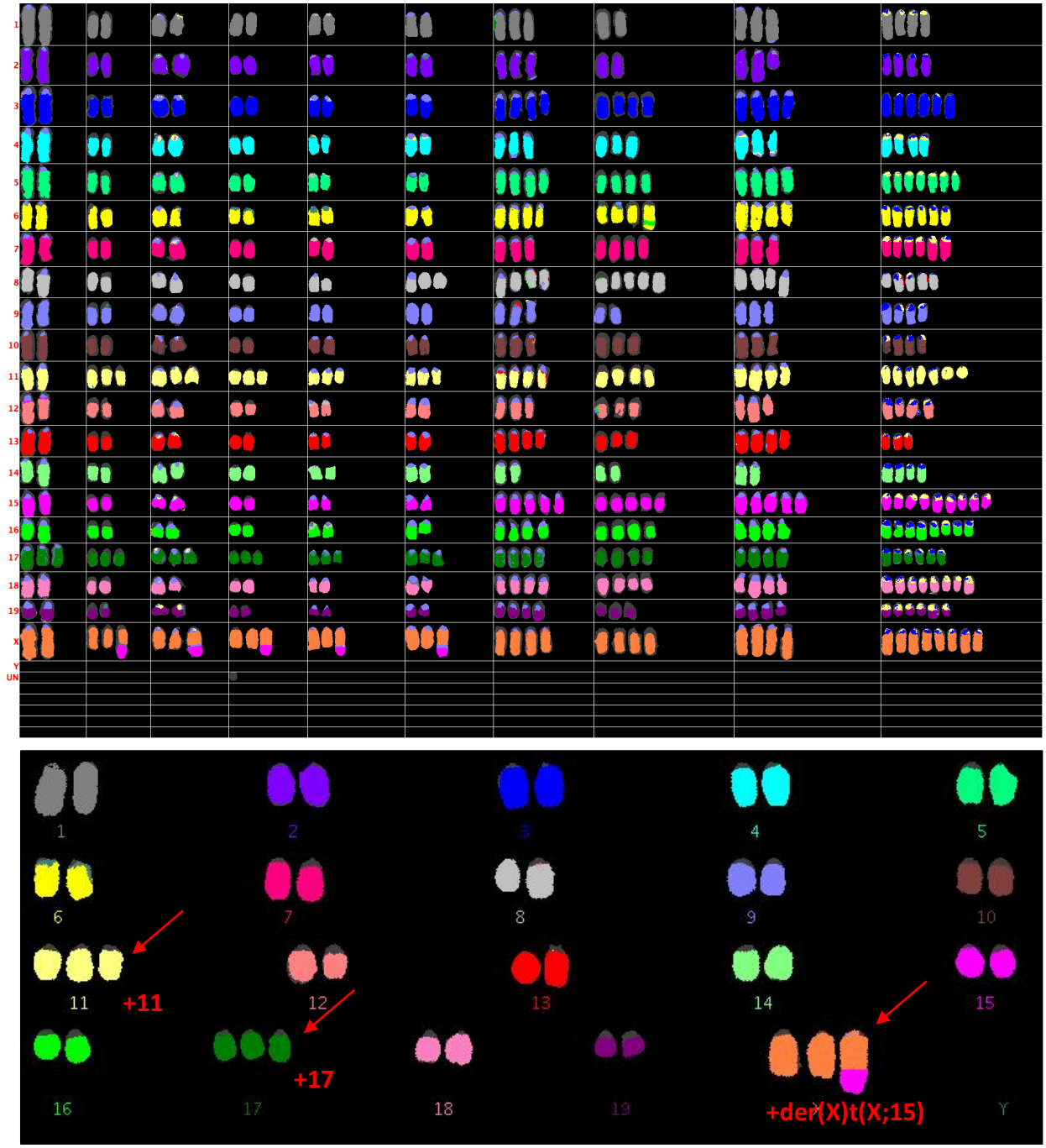


Figure 3.20. Cytogenetic analysis of an *EGFRvIII*-expressing mouse glioma. Top panel shows 10 sets of mitotic chromosomes in metaphase spreads from a single tumor; a metaphase is shown in a column and each row is a chromosome. Note the high degree of polyploidy. Bottom panel shows the dominant subclone (40% of cells had this chromosome profile), with an extra copy of chromosome 11 and 17, as well as an amplification of X and chromosome 15 (with a translocation between them)

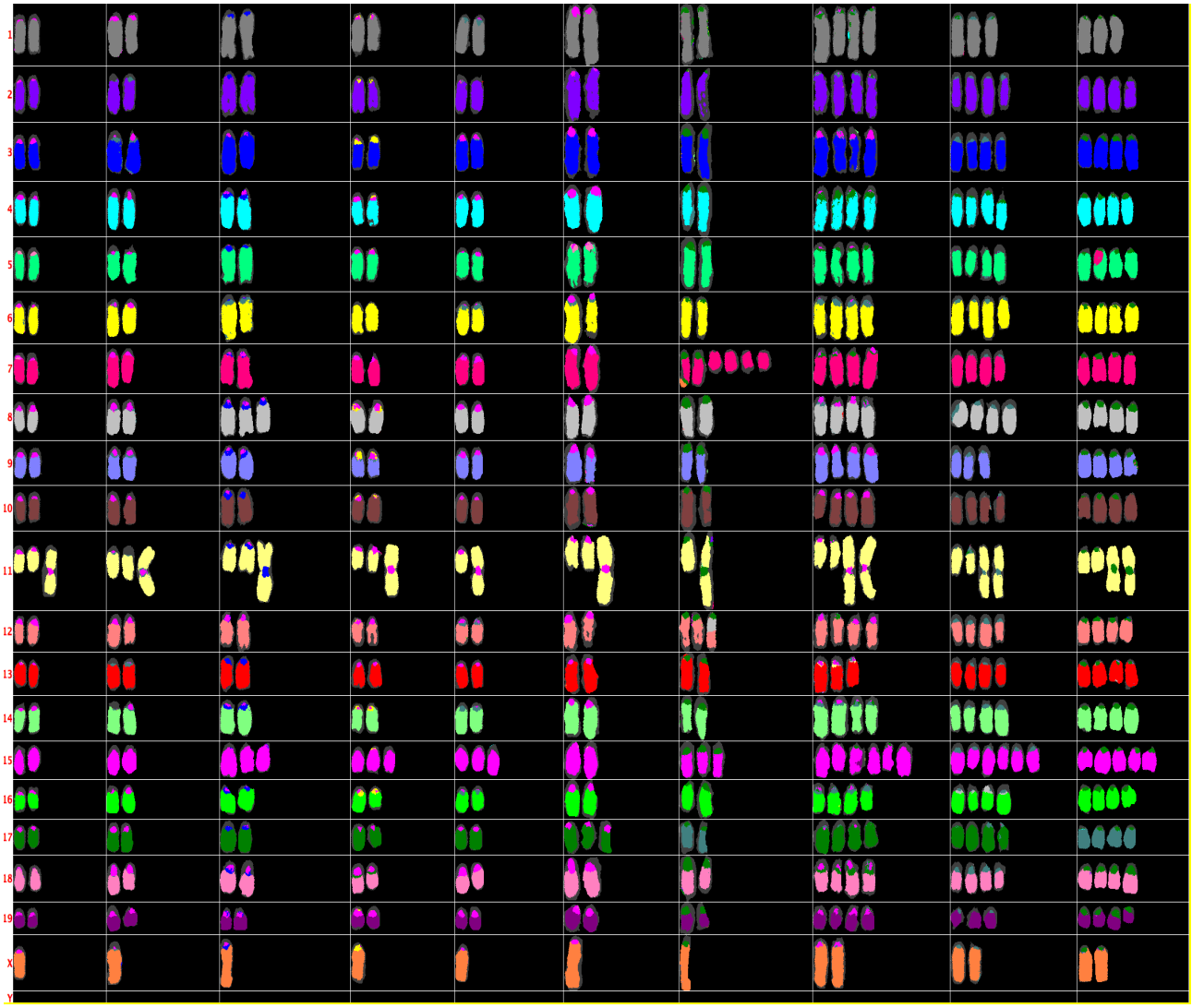


Fig 3.21. Cytogenetic profile for a second *EGFRvIII* mouse glioma, highlighting chromosome 11 amplification (in this case with a Robertsonian translocation) in all cells. Note also in this case that 4 out of 7 diploids have trisomy for chromosome 15.

RNA-Sequencing

We sequenced the RNA from *EGFRvIII*-expressing mouse gliomas (primary tumors) in order to determine the pathways activated in these tumors to provide insight into the relevant tumorigenic mechanisms. Whole-transcriptomic profiling of 11 brain tumors revealed a distinctive expression pattern for *EGFRvIII*-gliomas. There were 2000 genes that were significantly upregulated in *EGFRvIII*-expressing brain gliomas compared with control (wild-type mouse brains, n=6) with a log-fold change of more than 2 and a Benjamini-Hochberg adjusted (for multiple testing) $p < 0.01$, Wald's test (Supplementary Table 2). Gene ontology analysis for pathway enrichment of the 300 most upregulated genes in these brain tumors showed there was significant enrichment for multiple pathways, particularly those related to the cell cycle and mitosis, cell differentiation, central nervous system development and neurogenesis (FDR < 0.001 in all cases), Table 3.3. Downregulated genes showed enrichment for pathways such as neuron differentiation and migration (FDR < 0.001; downregulated genes are shown in Supplementary Table 3).

The most "upregulated" gene was the *EGFRvIII* transgene but as this human transgene is not present in normal mouse tissue, fold-change is not meaningful. The endogenous mouse *Egfr* gene was also upregulated (mean $\log_2$-fold-change = 3.71) in both brain and spinal tumors, suggesting both mutant-*EGFR* and wild-type *Egfr* expression are advantageous to tumor growth (Fig 3.22; human EGFRvIII and mouse Egfr could be differentiated based on sequence differences between the species, see Materials and Methods). In brain and spinal tumors, we confirmed that the majority of the top mutated genes are also expressed, including *Sub1*, *Trp53*, *Tead2, Nlrp1b, Nt5c2*, *Prex2*, *Uimc1* and *Itga6*.

Hox (homeobox) genes have been implicated in escape from apoptosis, epithelial-mesenchymal transition, and angiogenesis in other cancers[166]. Nineteen of the 30 most strongly upregulated genes in the brain tumors compared with wild-type brains were homeobox (*Hox*) genes (Benjamini-Hochberg adjusted $p < 1 \times 10^{-12}$, Wald's test, Fig 3.23).

Gene set enrichment analysis (GSEA) is a type of analysis used to delineate sets of genes that are over-represented in a large set of genes and that therefore may be linked to a disease phenotype, in this case gliomas. GSEA of differentially expressed genes in *EGFRvIII*- brain gliomas showed significantly enriched gene sets (FDR < 0.01, Kolmogorov-Smirnov test) including the p53 pathway, Wnt and Jak-Stat pathways, Rb pathway, and stem cell-related pathways, implicating these oncogenic pathways in gliomagenesis by cooperating with EGFR signalling, Fig 3.24.

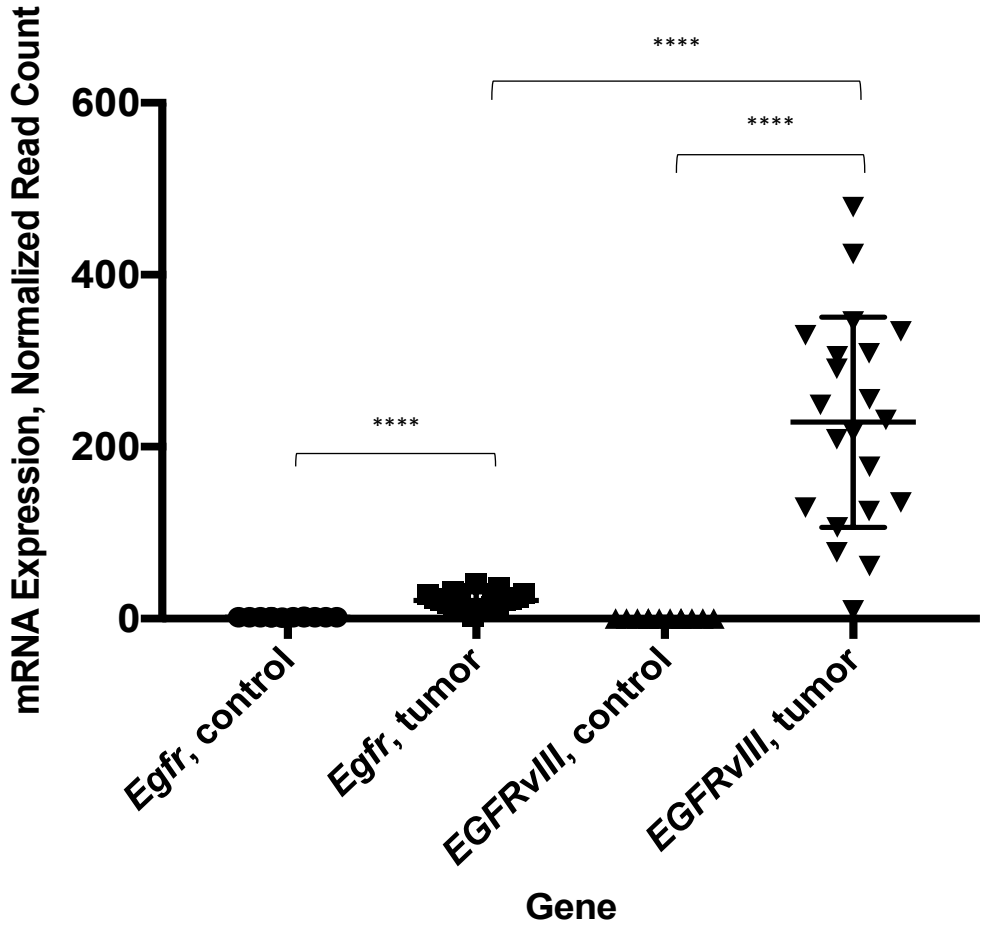


Figure 3.22. Plot showing stronger upregulation of EGFRvIII mRNA expression (from RNA-sequencing) compared with wild-type Egfr in tumors, highlighting the former is the more prominent driver (**** denotes p < 0.0001, paired t-test; n=11 brain tumors, n=10 spinal

tumors, relative to wild-type brain, n=6, and spinal cord, n=6). Mean expression and standard deviation values (error bars) are plotted.

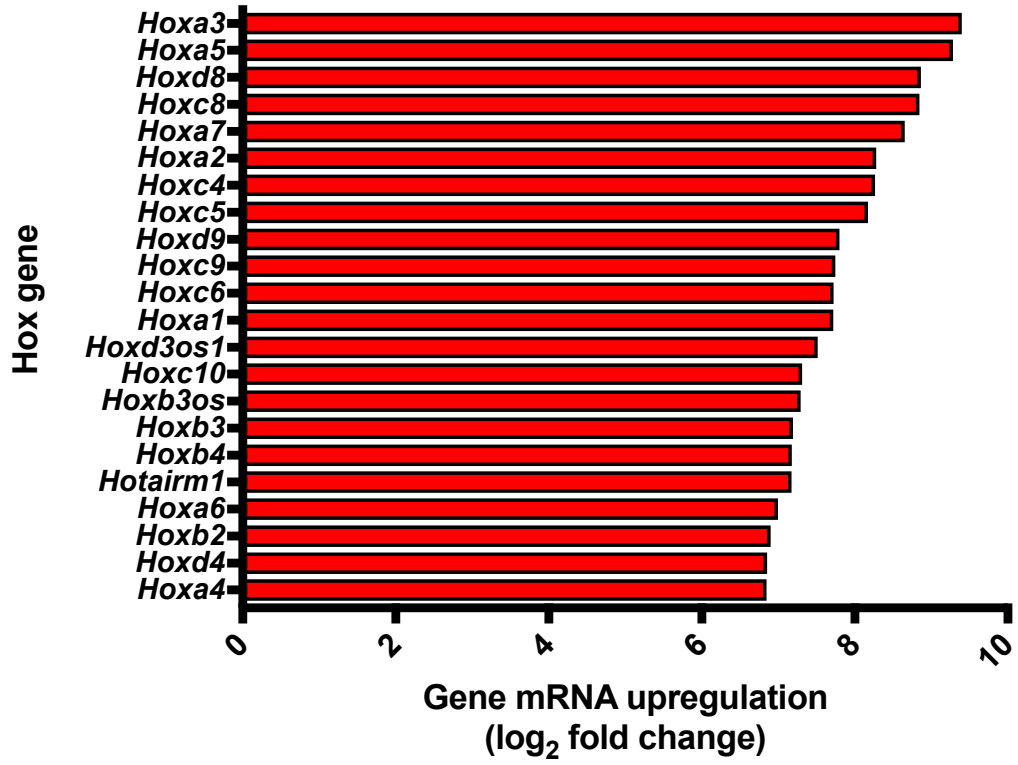


Fig 3.23. *Hox* gene upregulation in EGFRvIII gliomas arising in the mouse brain. Genes are ranked according to $\log_2$-fold change compared to wild-type mouse brain, Benjamini-Hochberg adjusted $p < 1 \times 10^{-12}$ for each gene (Wald's test).

	Biological Process (GO)		
Pathway ID	**Pathway Description**	**Count in Gene Set**	**False Discovery Rate**
GO:0048598	embryonic morphogenesis	43	2.00E-20
GO:0048513	organ development	79	2.06E-15
GO:0051301	cell division	30	5.96E-13
GO:0000278	mitotic cell cycle	31	5.32E-12
GO:0030154	cell differentiation	75	1.30E-10
GO:0007067	mitotic nuclear division	23	2.84E-10
GO:0080090	regulation of primary metabolic process	95	6.02E-10
GO:0000280	nuclear division	25	1.90E-09
GO:0048523	negative regulation of cellular process	80	1.93E-09
GO:0022402	cell cycle process	33	2.47E-09
GO:0031323	regulation of cellular metabolic process	95	3.51E-09
GO:0007049	cell cycle	38	3.66E-09
GO:0006351	transcription, DNA-templated	55	3.96E-09
GO:0048519	negative regulation of biological process	83	4.37E-09
GO:0019219	regulation of nucleobase-containing compound metabolic process	71	1.24E-08
GO:0050794	regulation of cellular process	129	1.25E-08
GO:0051171	regulation of nitrogen compound metabolic process	74	1.54E-08
GO:0060255	regulation of macromolecule metabolic process	91	1.56E-08
GO:0031326	regulation of cellular biosynthetic process	73	1.94E-08
GO:1901362	organic cyclic compound biosynthetic process	60	2.38E-08
GO:0051252	regulation of RNA metabolic process	65	9.47E-08
GO:0007059	chromosome segregation	16	1.21E-07
GO:2001141	regulation of RNA biosynthetic process	63	1.80E-07
GO:0045595	regulation of cell differentiation	42	1.82E-07
GO:0035270	endocrine system development	13	2.66E-07
GO:0009059	macromolecule biosynthetic process	63	3.06E-07
GO:0006355	regulation of transcription, DNA-templated	62	3.81E-07
GO:0048522	positive regulation of cellular process	78	1.97E-06
GO:0022008	Neurogenesis	37	3.32E-06
GO:0021546	rhombomere development	5	3.53E-06
GO:0090304	nucleic acid metabolic process	64	3.89E-06
GO:0007399	nervous system development	45	4.46E-06
GO:0031324	negative regulation of cellular metabolic process	48	7.19E-06
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	26	7.84E-06
GO:0006357	regulation of transcription from RNA polymerase II promoter	41	1.14E-05
GO:0048518	positive regulation of biological process	83	1.37E-05
GO:0010605	negative regulation of macromolecule metabolic process	47	1.58E-05
GO:0048699	generation of neurons	34	1.58E-05
GO:0007417	central nervous system development	27	1.82E-05
GO:1902679	negative regulation of RNA biosynthetic process	32	1.91E-05

Table 3.3. Gene ontology (DAVID) analysis for enriched pathways in transcriptomic profile of *EGFRvIII*-induced brain tumors. Fisher's exact test with FDR multiple testing correction is used for significance testing. This list was generated using publicly available DAVID analysis software, https://david.ncifcrf.gov/.

Comparison of the RNA-sequencing profile of 11 *EGFRvIII*-induced brain gliomas with 10 spinal gliomas revealed that there were significantly differentially expressed genes between these two groups of tumors. Differentially expressed genes were ranked according to the log-fold change from one tumor type to another. With a cut-off of log-fold change of 2 and a p-value less than 0.05 (Benjamini-Hochberg corrected for multiple testing, Wald's test), there were 68 upregulated genes in the brain tumor cohort compared with spinal tumors and 228 significantly downregulated genes. Analysis of the upregulated genes in brain tumors using STRING demonstrated that there was enrichment for interactions between the proteins represented by the genes, suggesting that these genes are connected as part of a network (p-value = 2.6×10^{-7}, Hypergeometric test), Fig 3.25. Gene ontology (GO) analysis of this gene set by DAVID (see Methods) showed significant enrichment for pathways involved in brain and head development, and specifically in forebrain development, Table 3.4. GO analysis of genes upregulated in spinal compared with brain tumors demonstrated significant enrichment for GO pathways involved in sensory and motor processes, such as detection of pain and thermal stimuli, response to external stimuli and regulation of muscle contraction, Table 3.5 and Fig 3.26. Collectively, these pathways are representative of the major functions of the spinal cord and spinal nerves, reflecting the location (and likely independent origin) of spinal tumors.

RNA-sequencing of spinal tumors showed gene set enrichment for the same pathways in tumors compared with wild-type spinal cord (FDR q-value < 0.01 in all cases, Kolmogorov-Smirnov test), including for example p53, Wnt and MAPK pathways, Fig 3.28. Supplementary Table 4 shows upregulated genes in spinal tumors and Supplementary Table 5 shows downregulated genes in spinal tumors.

To determine if the transcriptional profiles of these mouse tumors have similarity to human

cancers, we used GSEA comparing with all known datasets in MSigDB. Critically, the transcriptional profile of both brain and spinal *EGFR*-mutant gliomas significantly overlapped with the human Verhaak mesenchymal glioblastoma profile (FDR q-value < 0.01, Kolmogorov-Smirnov test) ([167]), Fig 3.27 and Fig 3.28. These data imply the transcriptome of these mouse gliomas resembles that of this particular subtype of human glioblastoma.

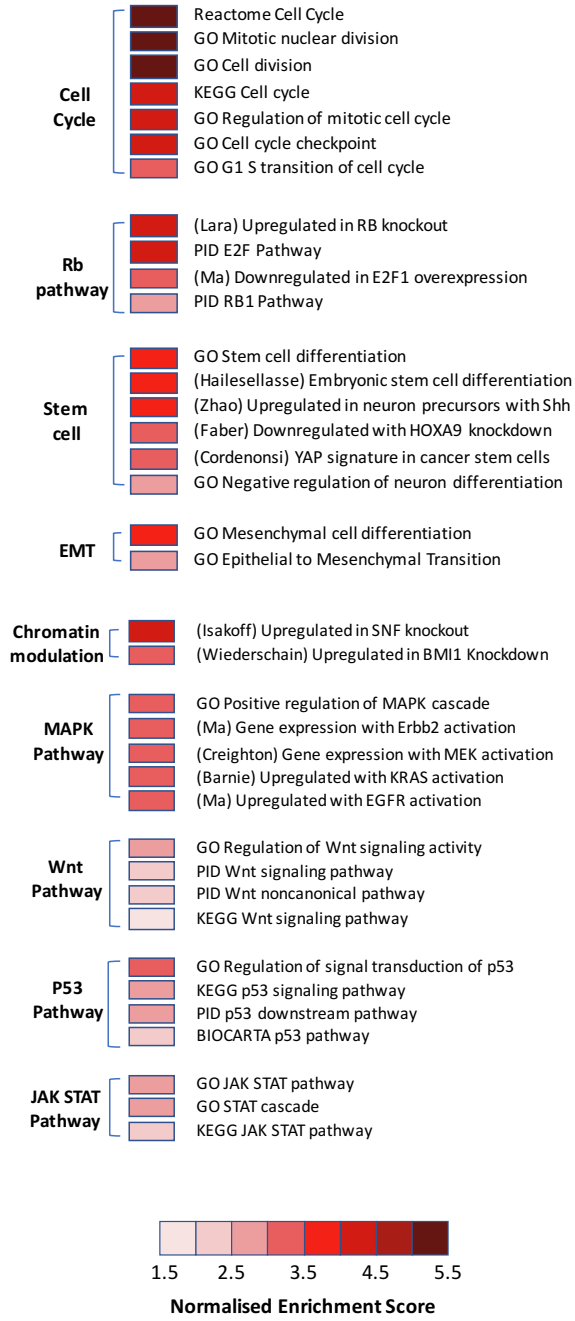


Fig 3.24. Gene set enrichment analysis reveals collaborative pathways in *EGFR*-mutant brain tumors identifies oncogenic pathways, stem-cell related genes and epithelial to mesenchymal transition (EMT) related genes. Each line identifies a transcriptomic profile with an FDR q-value < 0.01 (Kolmogorov-Smirnov test). Although not displayed here, spinal tumors enrich for the same pathways implying conserved molecular mechanisms. Gene set enrichment analysis produces an enrichment

score for a given gene set, which is then 'normalised' to account for the gene set size (as this can influence the enrichment score), producing a normalised enrichment score (NES), [119].

Biological Process (GO)			
Pathway ID	Pathway Description	Count in Gene Set	False Discovery Rate
GO:0060322	head development	13	4.95E-05
GO:0007417	central nervous system development	13	0.000186
GO:0007399	nervous system development	18	0.000206
GO:0035108	limb morphogenesis	7	0.00022
GO:0007420	brain development	11	0.000293
GO:0060173	limb development	7	0.000293
GO:0007389	pattern specification process	9	0.000873
GO:0030326	embryonic limb morphogenesis	6	0.000873
GO:0003002	Regionalization	8	0.0012
GO:0006366	transcription from RNA polymerase II promoter	9	0.0012
GO:0009653	anatomical structure morphogenesis	17	0.00128
GO:0030900	forebrain development	8	0.00139
GO:0060021	palate development	5	0.00141
GO:0007267	cell-cell signaling	9	0.00193
GO:0021761	limbic system development	5	0.00211
GO:0048731	system development	22	0.00211
GO:0009887	organ morphogenesis	11	0.00286
GO:0021877	forebrain neuron fate commitment	3	0.00286

Table 3.4. The gene ontology pathways most enriched in *EGFRvIII* brain gliomas compared with spinal gliomas. Analysis was performed using DAVID, with all genes upregulated in brain compared with spine tumors with $\log_2$-fold change > 2.0 and Benjamini-Hochberg adjusted p-value < 0.01 (Wald's test). Fisher's exact test with FDR multiple testing correction is used for significance testing. This list was generated using publicly available DAVID analysis software, https://david.ncifcrf.gov/.

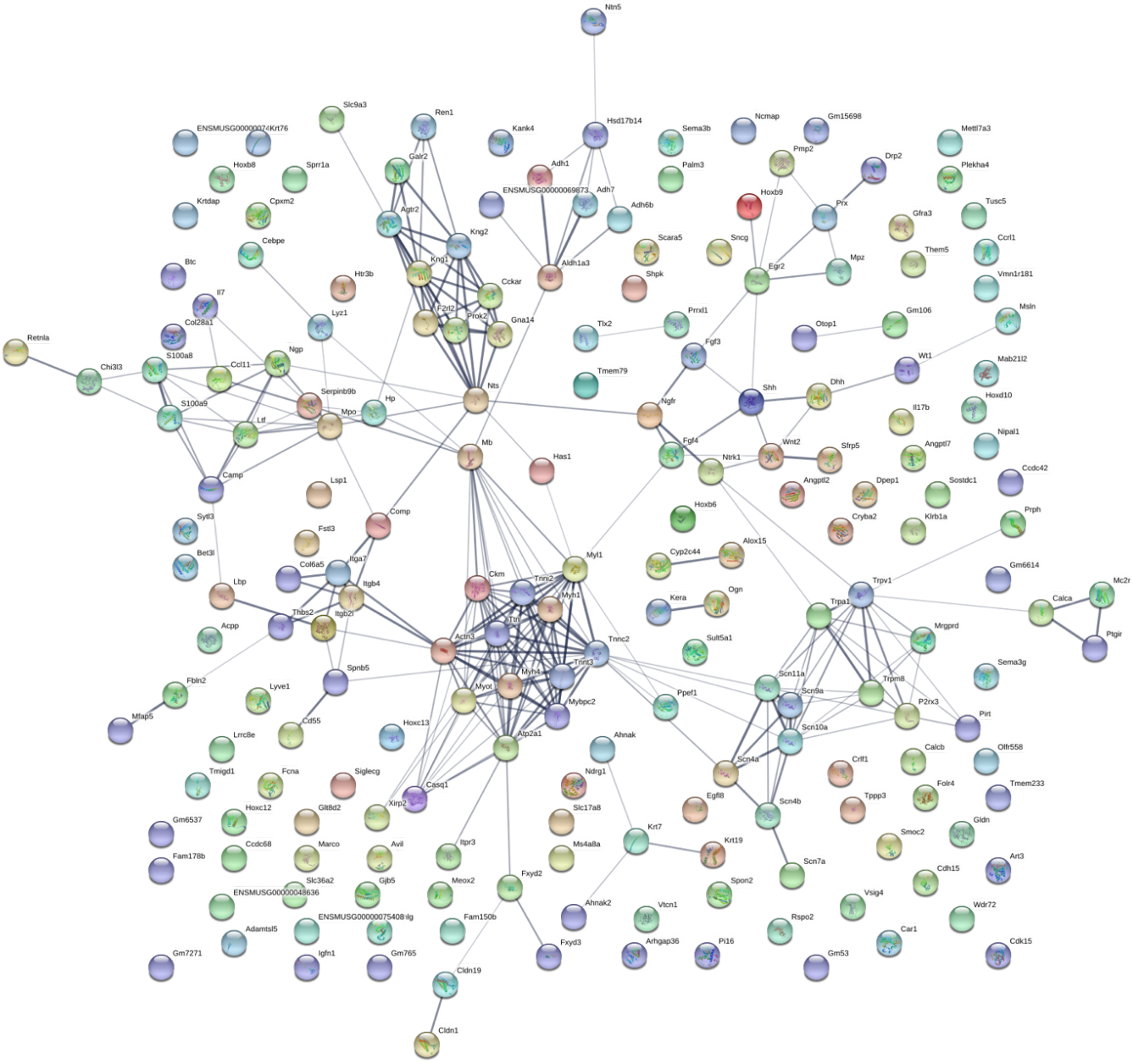


Fig 3.25. STRING network analysis of differentially upregulated genes in spinal tumors compared with brain tumors. The figure shows significant interactions between proteins of the genes in the list (Benjamini-Hochberg adjusted $p < 0.001$, Hypergeometric test), suggesting a functional network of genes are differentially expressed in spinal gliomas. This analysis was performed using the publicly available software http://string-db.org/cgi/network.pl. Color coding: colored nodes are proteins from

CIS genes; connecting lines are known or predicted interactions between proteins; see https://string-db.org for further details.

Biological Process (GO)			
Pathway ID	Pathway Description	Count in Gene Set	False Discovery Rate
GO:0003008	system process	34	9.63E-07
GO:0003012	muscle system process	13	7.76E-06
GO:0016048	detection of temperature stimulus	6	2.15E-05
GO:0009266	response to temperature stimulus	11	2.21E-05
GO:0006936	muscle contraction	11	3.19E-05
GO:0048731	system development	59	8.22E-05
GO:0048513	organ development	49	0.000173
GO:0009408	response to heat	8	0.000529
GO:0048265	response to pain	6	0.000571
GO:0050951	sensory perception of temperature stimulus	5	0.000571
GO:0006935	Chemotaxis	15	0.000591
GO:0009653	anatomical structure morphogenesis	39	0.000591
GO:0035962	response to interleukin-13	3	0.000591
GO:0006937	regulation of muscle contraction	9	0.000628
GO:0009605	response to external stimulus	32	0.000628
GO:0090257	regulation of muscle system process	10	0.000649
GO:0019228	neuronal action potential	5	0.000977
GO:0098655	cation transmembrane transport	17	0.00106
GO:0009582	detection of abiotic stimulus	8	0.00133
GO:0009607	response to biotic stimulus	19	0.00133
GO:0019226	transmission of nerve impulse	6	0.00133
GO:0009581	detection of external stimulus	8	0.00138

Table 3.5. Gene ontology pathways upregulated in *EGFRvIII* spinal tumors compared with brain tumors. DAVID analysis for enriched pathways reveals processes related to the spinal cord are upregulated in spinal tumors, including responses to external stimuli and muscle control. This result is more consistent with spinal gliomas being primary tumors rather than metastases from the brain. Fisher's exact test with multiple test FDR correction used for significance testing (FDR < 0.05 taken as significant). This list was generated using publicly available DAVID analysis software, https://david.ncifcrf.gov/.

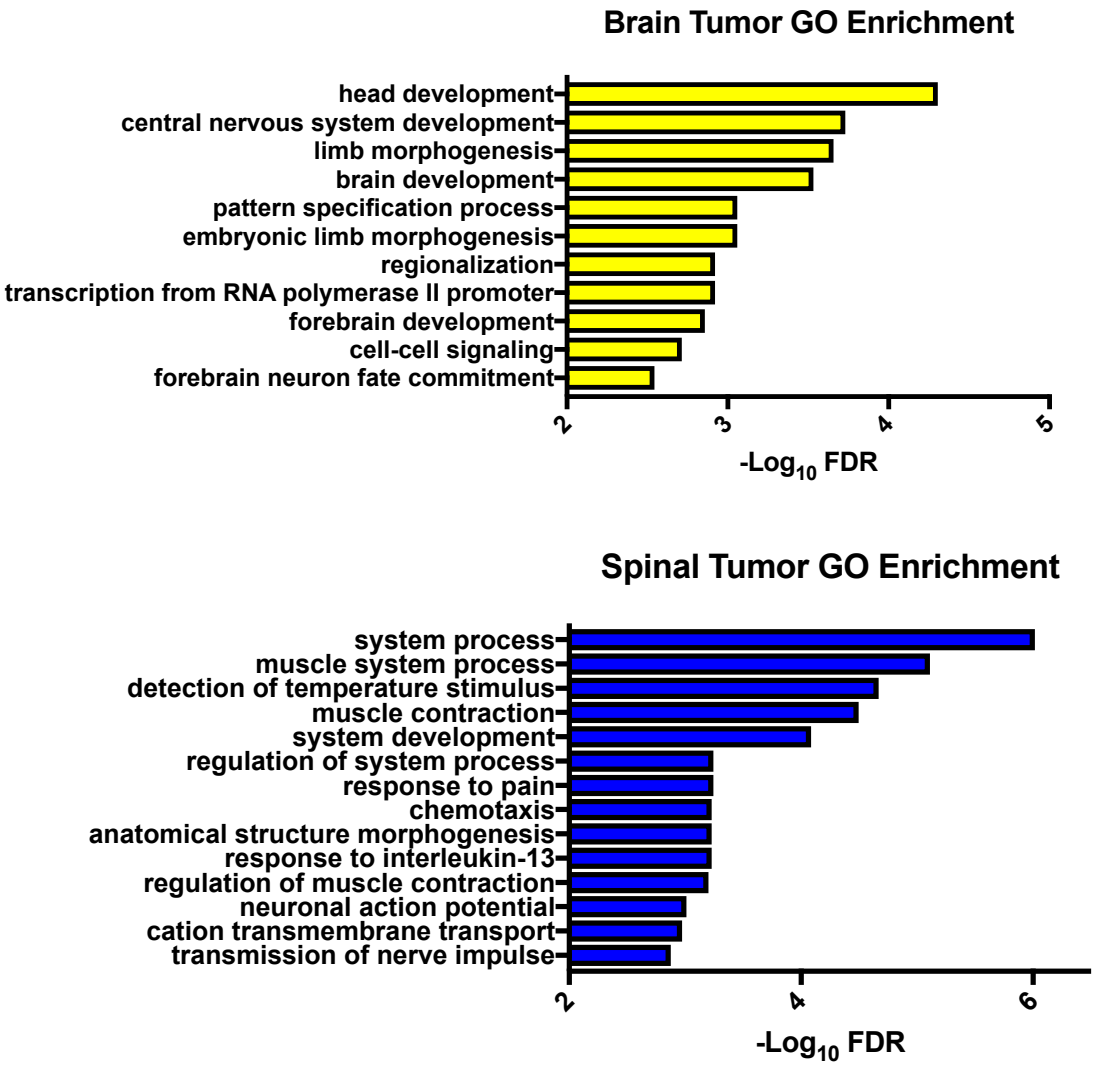
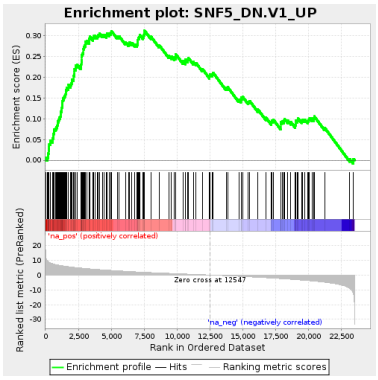
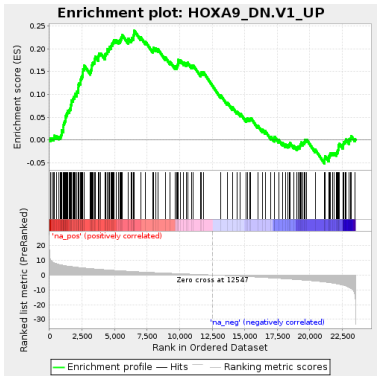


Fig 3.26. Gene ontology (DAVID) analysis of differentially expressed genes in *EGFRvIII; nes*-cre mouse brain and spinal gliomas. In brain tumors, there is significant enrichment for gene sets reflecting brain developmental processes, whereas in spinal tumors the gene sets reflect processes intrinsic to the spinal cord. These data reflect the different tissue origins of these tumors; note the absence of gene sets for oncogenic pathways here, as these are largely shared between the two types of tumor. FDR – false discovery rate. Fisher's exact test with FDR multiple testing correction is used for significance testing (FDR < 0.05 taken as significant).

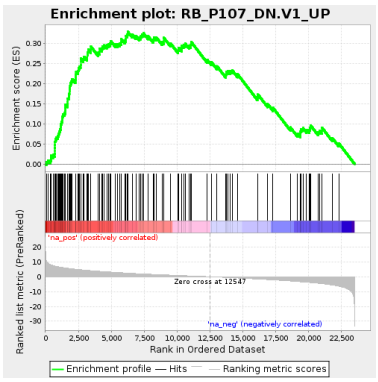
a


Enrichment plot: SNF5_DN.V1_UP
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 12547
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

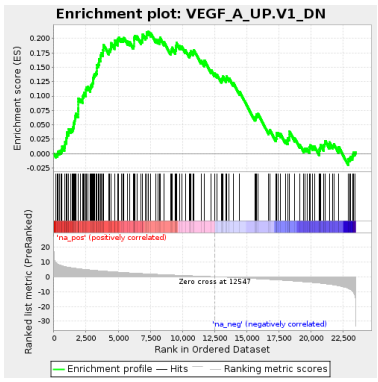
b


Enrichment plot: HOXA9_DN.V1_UP
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 12547
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

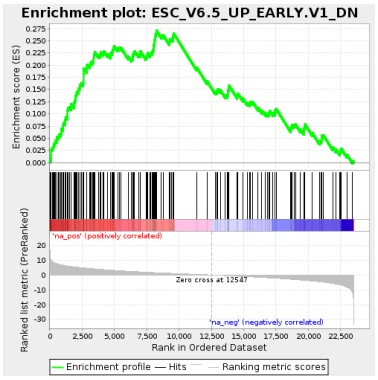
c


Enrichment plot: RB_P107_DN.V1_UP
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 12547
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

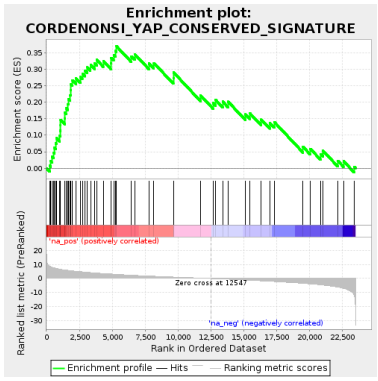
d


Enrichment plot: VEGF_A_UP.V1_DN
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 12547
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

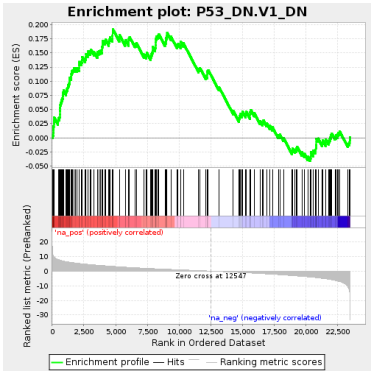
e


Enrichment plot: ESC_V6.5_UP_EARLY.V1_DN
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 12547
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

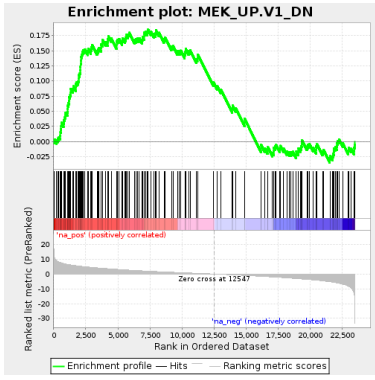
f


Enrichment plot:
CORDENONSI_YAP_CONSERVED_SIGNATURE
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 12547
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

g


Enrichment plot: P53_DN.V1_DN
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 12547
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

h


Enrichment plot: MEK_UP.V1_DN
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 12547
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

i

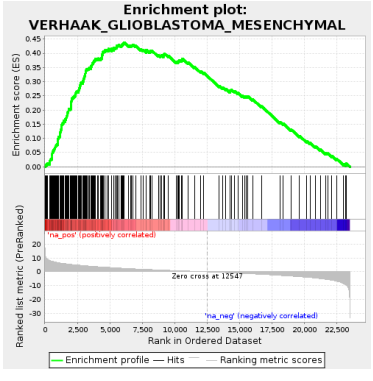
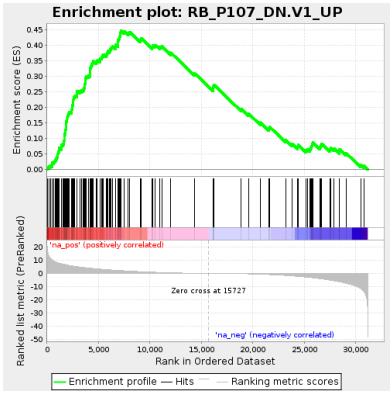
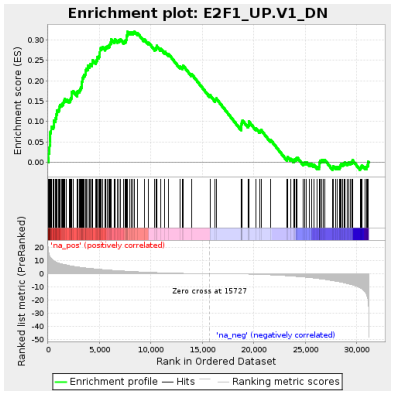

Enrichment plot:
VERHAAK_GLIOBLASTOMA_MESENCHYMAL
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 12547
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

Figure 3.27. Gene set enrichment analysis (GSEA) of *EGFRvIII; nes*-cre brain gliomas defines their key oncogenic pathways. Plots are displayed for a selection of significantly enriched gene sets, including *SNF5*, *HOXA9*, *RB*, *VEGF*, *ESC*, *YAP*, *P53* and *MEK* gene sets (FDR q-value < 0.01, Kolmogorov-Smirnov test). I. Transcriptional profile of these tumors are significantly enriched for the Verhaak human mesenchymal glioblastoma profile (FDR q-value < 0.01, Kolmogorov-Smirnov test).

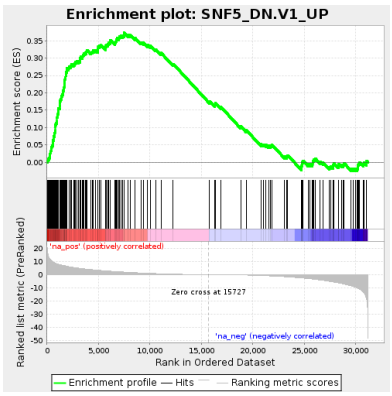
a


Enrichment plot: RB_P107_DN.V1_UP
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 15727
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

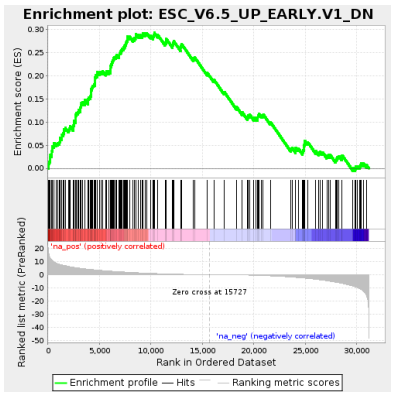
b


Enrichment plot: E2F1_UP.V1_DN
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 15727
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

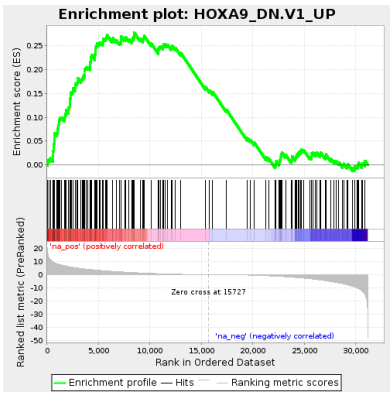
c


Enrichment plot: SNF5_DN.V1_UP
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 15727
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

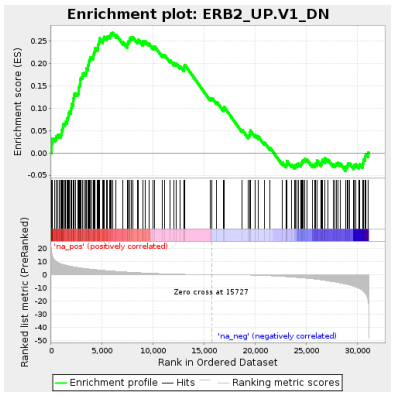
d


Enrichment plot: ESC_V6.5_UP_EARLY.V1_DN
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 15727
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

e


Enrichment plot: HOXA9_DN.V1_UP
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 15727
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

f


Enrichment plot: ERB2_UP.V1_DN
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 15727
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

g

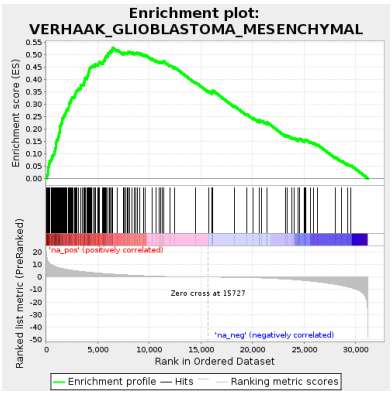

Enrichment plot:
VERHAAK_GLIOBLASTOMA_MESENCHYMAL
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 15727
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile — Hits — Ranking metric scores

Figure 3.28. Gene set enrichment analysis of *EGFRvIII; nes*-cre spinal gliomas defines their key oncogenic pathways. Plots for a selection of significantly enriched gene sets are presented here, including *Rb*, *E2F1*, *SNF5*, *ESC* (embryonic stem cell), *HOXA9*, and *ERBB2* gene sets (FDR q-value < 0.01, Kolmogorov-Smirnov test). G. Transcriptional profile of these tumors are significantly enriched for the Verhaak human mesenchymal glioblastoma profile (FDR q-value < 0.01, Kolmogorov-Smirnov test).

Whole-Exome Sequencing for Mutations and Copy Number Changes

In order to identify somatic mutations and copy number changes acquired following glioma initiation by *EGFRvIII* in our mice, we performed whole-exome sequencing (WES) on 17 tumors (9 brain and 8 spinal gliomas) with matched normal spleen controls from the same mice. Across all tumors, we found 85 significant recurrently mutated genes with mutations in two or more tumors identified by MuSiC[111] (adapted for mouse data- see Materials and Methods for details for mutation calling and determination of significance); most had single-nucleotide variants (SNVs) but some genes exhibited INDELS (insertions or deletions; Fig 3.29, Supplementary Table 1). The median number of exonic mutations per tumor was 29 of which missense mutations were the most common . Amongst the single nucleotide variants, T > C and C > T were the commonest changes. *Sub1* was the most frequently mutated gene (6 mutations in 5/17 tumors, $p=1.1 \times 10^{-16}$, FDR 2.27×10^{-12}, Likelihood ratio test, LRT) displaying INDELs and SNVs, all in splice sites suggesting loss-of-function, Fig 3.29. *Sub1* is a transcriptional regulator whose precise function is unknown, however recent work has demonstrated it is upregulated in many cancers and its knockdown reduced prostate cancer cell invasion *in vitro* and tumor growth *in vivo*, with chromatin immunoprecipitation showing that *Sub1* binds promoter regions of oncogenes *C-MYC* and *PLK1* in prostate cancer [168]. *Trp53*, a known glioma tumor suppressor[169], was the second most frequently mutated gene (5/17 tumors had a *Trp53* missense mutation, all within *Trp53*'s DNA-binding domain; $p=1.13 \times 10^{-12}$, FDR 7.75×10^{-9}, LRT), validating the application of WES to identify relevant collaborative mutations, Fig 3.30. Similarly, *Nf1,* a known genetic driver of brain glioma[170], was found to be mutated in two tumors (p=0.0010, FDR 0.17, LRT).

Other frequently mutated novel genes were *Tead2*, *Nt5c2, Ces1c, Nlrp1b*, *Prex2, Uimc1* and *Itga6*. *Tead2*, a transcription factor in the Hippo pathway, had recurrent mutations across its TEA/ATTS (DNA-binding) domain (4 mutations in 3/17 tumors; $p=2.80 \times 10^{-11}$, FDR 1.15×10^{-7}, LRT), including splice site mutations and one frameshift mutation, suggesting loss-of-function. *TEAD2* is thought to be involved in tumor suppression via interaction with the YAP oncoprotein in the Hippo signalling pathway, restricting cell proliferation and promoting apoptosis [171, 172]. *Uimc1* and *Itga6* had three mutations each ($p=1.39 \times 10^{-7}$ and FDR 1.9 x

10^{-4}, p=2.7 x 10^{-7} and FDR 3.2 x 10^{-4}, LRT, respectively) all of which were INDELS and one of which caused a frameshift in *Itga6* (Fig 3.29).

In contrast to the relatively small number of recurrent mutations, *EGFR*vIII tumors had complex genomes by DNA copy number analysis (Fig 3.40). Significant focal amplifications and deletions, identified by GISTIC2[115], were evident in regions with known cancer genes, for example significant focal *Cdkn2a* deletions (GISTIC q-value = 1.39 x 10^{-5}) were evident and *EGFRvIII* (in *Col1a1* locus, GISTIC q-value = 0.017) was recurrently amplified. Significantly recurrent focal deletions were present in novel putative glioma drivers *Nlrp1b* and *Adgrl2* (GISTIC q-value = 2.92 x 10^{-14} and 2.19 x 10^{-6} respectively, Fig 3.31). Several of the most significantly mutated genes were also in regions with frequent deletions, including *Trp53*, *Tead2* and *Uimc1*, supporting putative tumor suppressive roles (Fig 3.31).

The potential significance and translational relevance of the most frequently mutated and / or focally deleted genes detected in the mouse gliomas were assessed by comparison with human low-grade glioma (LGG) data from 283 patients from The Cancer Genome Atlas (TCGA)[173], using the online tool Cbioportal (see Materials and Methods for further details). This revealed that *TEAD2* is recurrently deleted in 48% of human LGGs in a mutually exclusive manner with *TP53* (Bonferroni-adjusted $p < 0.001$, Fisher's exact test, Fig 3.32), identifying *TEAD2* as a putative cancer gene. Recurrent deletions in previously unknown glioma genes *NT5C2*, *ADGRL2* and *UIMC1* were observed whilst *SUB1, CES1*, *NLRP1* and *ITGA6* were frequently methylated in human LGGs (Supplementary Fig S8). These data help cross-validate the relevance of these novel putative cancer genes in human patients.

These mouse gliomas were all wild-type for *Idh1* (17/17 tumors sequenced), consistent with gliomas in humans in which *IDH1* and *EGFR* mutations tend to be mutually exclusive [174].

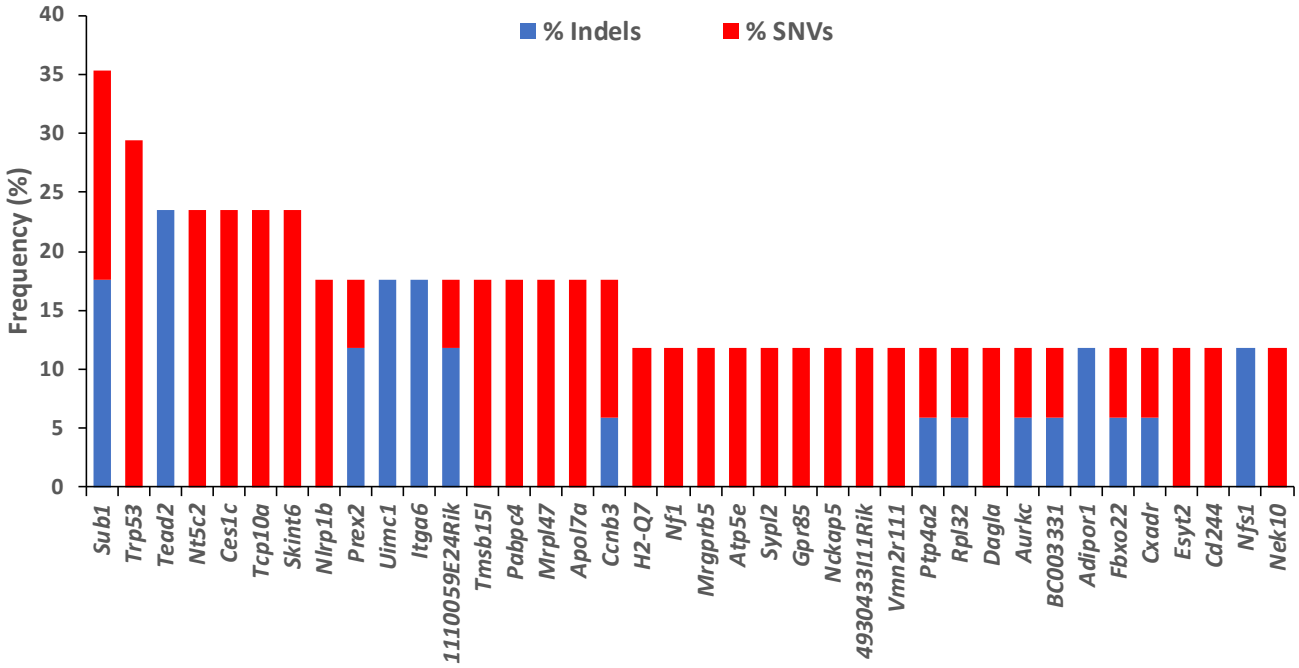


Fig 3.29. Mutational profile of 17 brain and spinal tumors. Genes are ranked according to the frequency of mutations (indels or SNVs). Known glioma drivers include *Trp53* and *Nf1*, and novel ones found mutated are *Sub1* and *Tead2*.

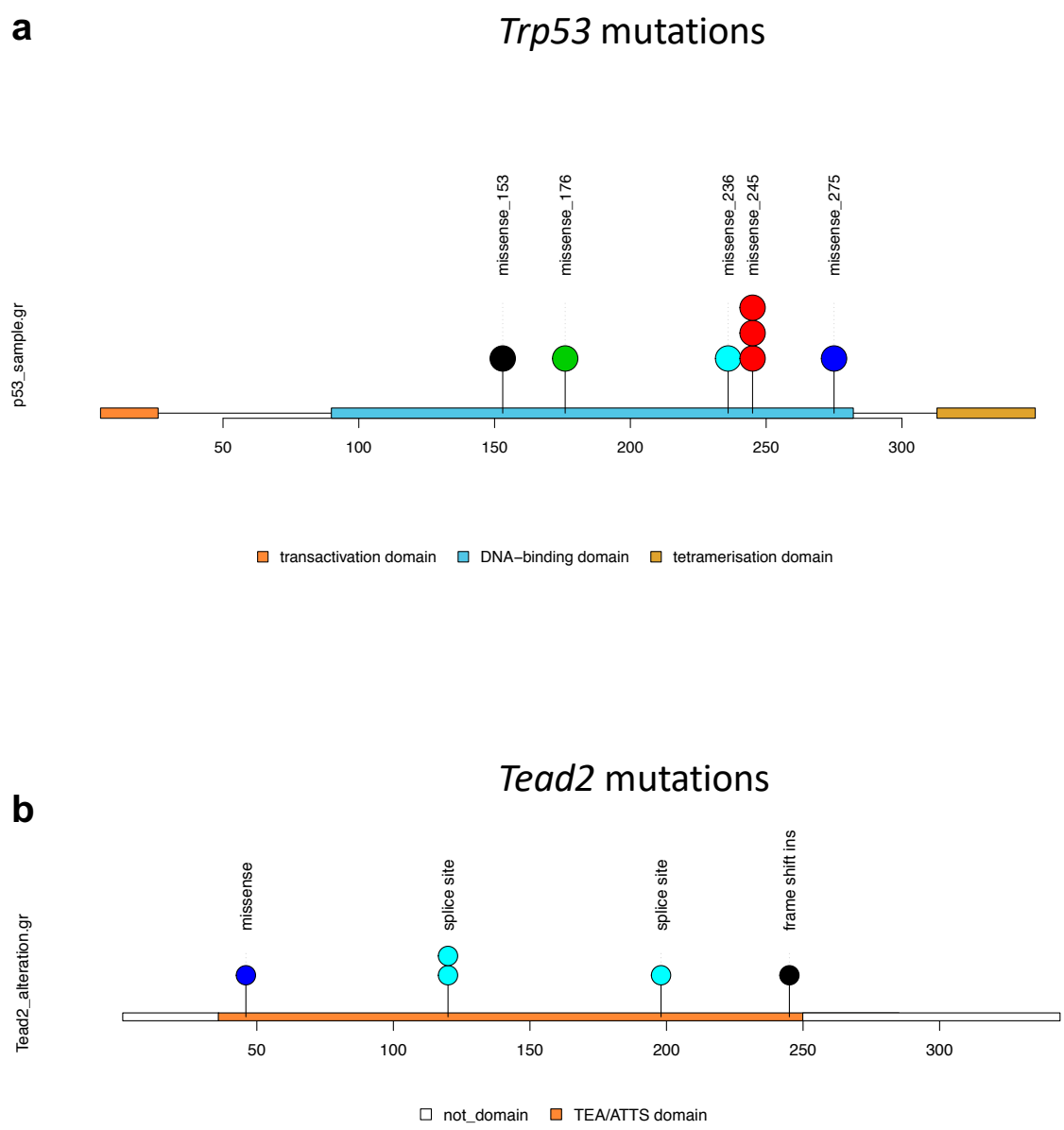


Fig 3.30. Mutations of *Trp53* and *Tead2* in *EGFRvIII*-only and *EGFRvIII*-PB gliomas are in DNA-binding domains. A. Plot outlining the location of *Trp53* mutations across all exome-sequenced mouse tumors. Five *EGFRvIII*-only and two *EGFRvIII*-PB tumors had in *Trp53*, all residing within its DNA-binding domain; 3 occurred in the same location. B. Plot outlining locations of *Tead2* mutations, all residing in the TEA/ATTS domain which is the DNA-binding domain of *Tead2*; 2 mutations were in same splice site location.

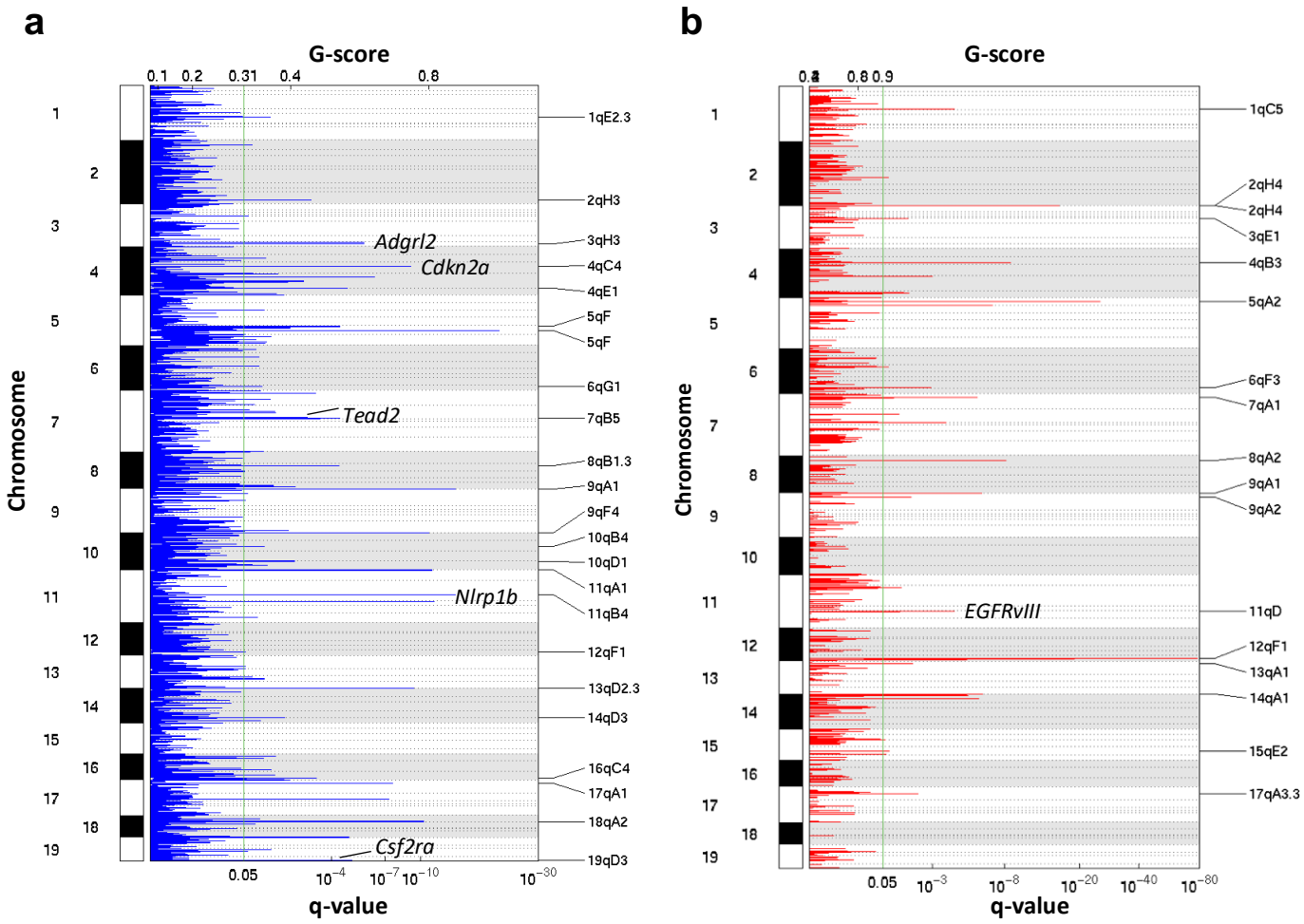


Figure 3.31. Plots showing focal copy number variations across *EGFRvIII*-only and *EGFRvIII*-PB mice. Significant focal deletions as determined by GISTIC2 are displayed in A, and significant focal amplications are displayed in B. Lower x-axis represents q-value (significance at < 0.05) and top x-axis represents the G-score. G-score and p-values are calculated as described in [115].

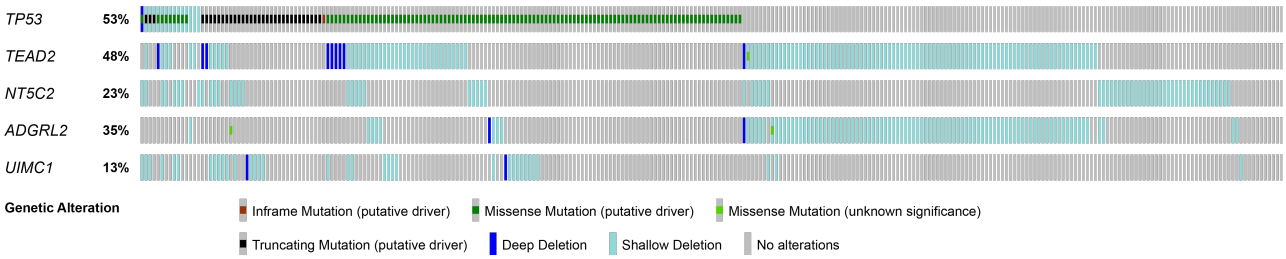


Figure 3.32. Analysis of TCGA data on human LGGs for genetic alterations in the genes most significantly mutated in mouse *EGFRvIII* gliomas. These genes are recurrently mutated in *EGFRvIII*-gliomas in mice, and are seen here to be altered with high frequency in patients with LGG; these alterations in *TP53* and *TEAD2* are mutually exclusive in this patient cohort (Bonferroni-adjusted $p < 0.001$, Fisher's exact test), as are alterations in *NT5C2* and *TP53* (Bonferroni-adjusted $p = 0.003$, Fisher's exact test). These data were analysed using the publicly available software Cbioportal, http://www.cbioportal.org/. Each column represents one tumor.

Transposon Mutagenesis Replaces Genomic Instability in Glioma Progression

Transposons have been successfully used for identifying cancer driver genes[54, 55, 57, 58, 60, 61, 67, 70, 175]. Mobilized *PiggyBac* transposons randomly integrate in the genome and activate and/or disrupt gene expression[69] (see Chapter Four for further discussion of the background to transposon mutagenesis). We performed a conditional *PiggyBac* transposon mutagenesis screen *in vivo* to further identify genes that cooperate with mutant-*EGFR* in gliomagenesis.

As was the case for *EGFRvIII*, To limit transposition to the central nervous system a conditional *PiggyBac* transposase allele was activated by *nes-cre*. An experimental cohort of quadruple transgenic mice carrying conditional *EGFRvIII*, 20 copies of a *PiggyBac* transposon (ATP1S2)[69], a conditional *PiggyBac* transposase and *nes-cre* were generated (*EGFRvIII*-PB, n=47; Fig 3b, see Methods). As controls, we established transgenic mice expressing *EGFRvIII* but lacking transposition (*EGFRvIII*; *nes-cre* = EGFRvIII-only, n=31; and *EGFRvIII*; *nes-cre*; ATP1S2, n=10) and a set with transposition but lacking *EGFRvIII* (transposase; ATP1S2; *nes-cre* = PB-only, n=20). Mean survival times between *EGFRvIII*-PB and *EGFRvIII*-only cohorts were similar (36.2 vs 38.1 weeks, p=0.95, log-rank test, Fig 3.33) and both groups had similar incidences and pathological grades of brain and spinal gliomas (Fig 3.34 and 3.35, Table 3.6). After one year, 10 PB-only mice without any neurological signs were culled to determine if transposon mobilisation in the CNS was sufficient to induce tumors or tumor precursor lesions - histological analysis did not reveal any tumors in the brain or spine of these 10 mice.

Survival Times of Mice

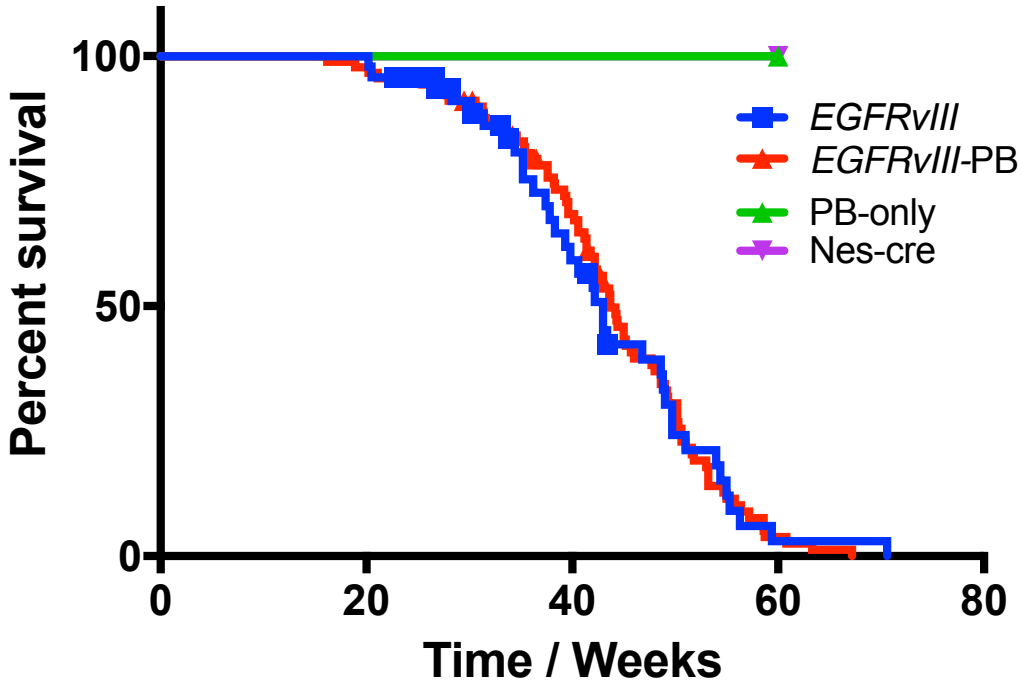


Fig 3.33. Kaplan-Meier curves for survival times of *EGFRvIII*-mice with and without *piggyBac* transposition. Kaplan-Meier survival curves of EGFRvIII-only (n=31) and EGFR-PB (n=47) mice, with no significant difference between them (p = 0.95, log-rank test). No differences in survival or 129 pathology were observed between EGFRvIII; nes-cre and EGFRvIII; nes-cre; ATP1S2 mice. Tumors were not observed in PB-only (TSPB; ATP1S2; nes-cre, n=20) or nes-cre (n=10) mice after 60 weeks.

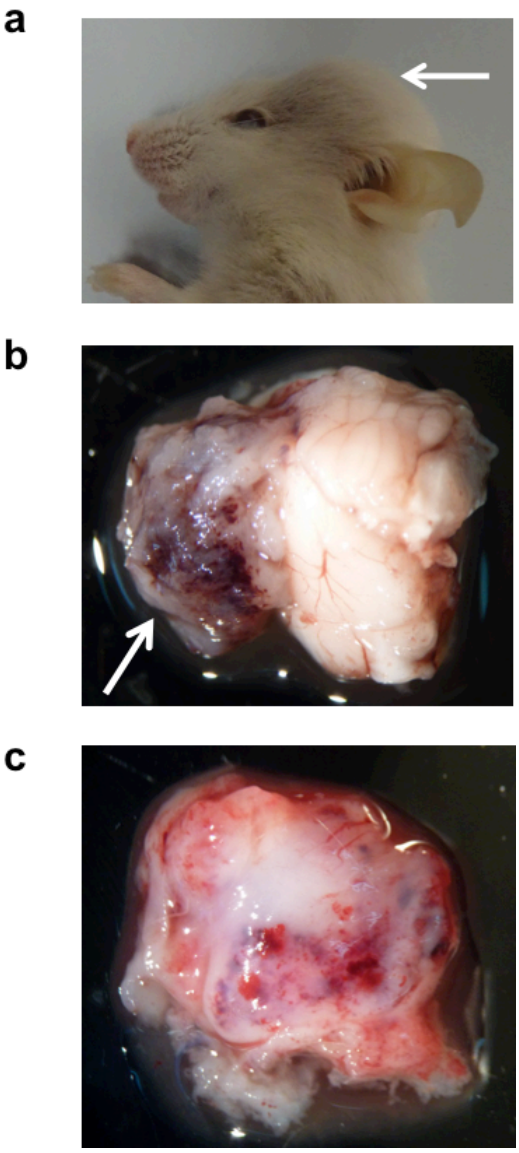


Fig 3.34. Clinical Phenotype of *EGFRvIII*-PB mice. A. Photograph of an *EGFRvIII*-PB mouse with enlarged head due to an underlying brain glioma. B, C. Macroscopic photographs of areas of the brain from the same mouse showing the presence of a tumor on the brain surface.

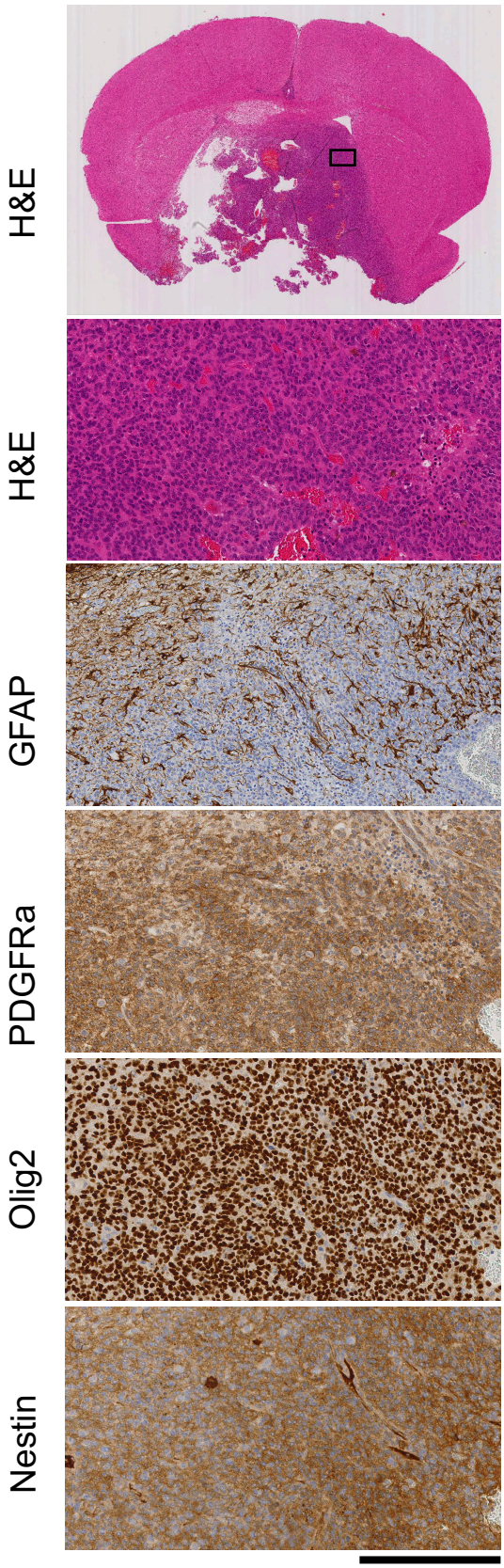

H&E
H&E
GFAP
PDGFRa
Olig2
Nestin

Fig 3.35. H&E and immunostaining profile of a typical grade III brain glioma from an *EGFRvIII*-PB mouse, which demonstrates strong expression of neural stem and transit-amplifying cell markers. The scale bar at the bottom corresponds to 2.8mm for the top panel, and 200μm for all panels below.

Mouse ID	Age (weeks)	Clinical Phenotype	Brain Pathology	Spine Pathology
26.2b	29.5	head greatly enlarged; difficulty breathing; minimal mobilisation	SVZ expansion, widespread leptomeningeal tumor, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
5.2e	30	moderate -severe macrocephaly, pilorection, laboured breathing, hunched, lethargic	subarachnoid spread, large basal tumor, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
36.1b	30.3	moderate macrocephaly, severe ataxia, weak hindlimbs	SVZ growth, possible subarachnoid spread, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
21.2a	30.6	moderate macrocephaly, very lethargic, hyperventilating	widespread subarachnoid tumor growth (grade II glioma), thickened SVZ	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
24.1a	35.5	moderate - severe macrocephaly, lethargy, difficulty breathing; tilted	tumor cell clusters in ventricle, subarachnoid spread, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
22.1b	38.3	acute hindlimb paralysis; mild macrocephaly, head tilt.	small tumor nests in subarachnoid space, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
8.4b	36.5	moderate macrocephaly, rapid breathing (for several weeks), very lethargic / hunched.	expansion of svz, small subarachnoid tumor clusters/ spread, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
5.3c	41.8	moderate macrocephaly, rapid / deep breathing acutely, hunched / immobile.	no SVZ changes, no tumor seen	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
5.5b	36	mild macrocephaly, hyperventilating, very weak hindlimbs (dragging); very big bladder (likely spinal)	intraventricular, subarachnoid tumor growth grade 2/3	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
25.2b	34.2	severe macrocephaly, rapid breathing, slow	intraventricular tumor cell clusters, subarachnoid spread, grade 2/3	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
25.2d	34.2	severe macrocephaly, rapid breathing, all limbs weak (?spinal parenchyma invasion)	tumor spread grade 2/3	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
35.1e	35.3	severe macrocephaly, rapid breathing, severe tilt; hunched	large tumor spread grade 4	widespread subarachnoid/ leptomeningeal spread, minimal

				infiltration into spinal cord but strong infiltration into nerve roots; grade II
38.1e	34.6	moderate body tilt, culled early for tail injury	intraventricular, subarachnoid spread, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
25.3a	31.6	severe macrocephaly, hyperventilating, lethargy	grade 4; basis of brain, frontal, subarachnoidal	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
21.1j	42.2	acute paralysis, urinary retention	grade 2 glioma, no surface attachment	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
5.3b	45.7	moderate macrocephaly, weight loss, lethargy	grade 2 glioma, Ventricular growth, base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
8.5c	37.6	moderate / severe macrocephaly, lethargy	grade 2 glioma, Ventricular growth, base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
49.1b	31.3	moderate macrocephaly, lethargy, hyperventilating	grade 2 glioma, Tumor in lateral ventrical, base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
21.1h	45	acute paralysis (complete) of hindlimbs, prior abnormal gait	grade 2 glioma, small tumor nest in base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
53.1e	29.6	hunched, hyperventilation, severe macrocephaly	grade 2 glioma, small tumor nest in base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
21.2f	40.2	severe ataxia (head bob), walking low	grade 2 glioma, Ventricular growth, base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
22.2b	30	moderate macrocephaly, walking low / weak hindlimbs, big bladder/retention	grade 2 glioma, Ventricular growth, base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
48.1a	37.6	severe macrocephaly, hyperventilating, lethargy	grade 2 glioma, Ventricular growth, base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
35.1D	42.4	Mild hydro, very weak hindlimbs, urine retention	grade 2 glioma, intraventricular tumor and small tumor at the base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II

30.1a	42.2	moderate macrocephaly (hydro), lethargic, weak limbs	grade 2 glioma, intraventricular, small tumor nest at the base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
30.3a	36	severe macrocephaly, slow, hyperventilation	grade 3 possibly 4, large circumscribed extra cerebral tumor	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
48.1b	39.6	moderate body tilt, lethargy	grade 2 glioma, intraventricular tumor, small tumor islands at the base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
49.2e	33.3	moderate body tilt, lethargy	grade 2 glioma, small intraventricular tumor	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
48.1i	39.6	moderate body tilt, hydrocephalus, lethargy, low walk, head bob	grade 2 glioma, very small tumor nests at the base of the brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
21.1b	49.3	moderate body tilt, uncoordinated; no hydrocephalus	grade 4 glioma, large circumscribed extra cerebral tumor	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
32.1d	43.7	severe macrocephaly, lethargy, body tilt, hyperventilation	grade 4, large extra cerebral tumor, glioma at the base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
13.2c	40	hydrocephalus, limb paralysis	grade 2 glioma, small intraventricular tumor, small tumor islands at the base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
21.1a	50.6	hydrocephalus, paralysis, hyperventilation	grade 4 glioma, large extracerebral tumor, tumor islands at the base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
21.2g	43.7	hydrocephalus, inactive / strong body tilt	grade 4 glioma, large tumor at the base of brain with MVP (microvascular proliferation)	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
53.2b	31.4	hydrocephalus, hunched, piloerection, uncoordinated movements	no tumor	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
22.2j	44.3	moderate/severe hydrocephalus, hyperventilation, lethargy	grade 2 glioma, multiple intarventricular tumor islands, widespread tumor cell subseeding	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
13.3a	49.2	moderate hydrocehapalus, uncoordinated / slow	grade 2 glioma, small intraventricular tumor,	widespread subarachnoid/ leptomeningeal spread, minimal

			small tumor islands at the base of brain	infiltration into spinal cord but strong infiltration into nerve roots; grade II
35.1b	45	moderate hydrocephalus, very weak hindlimbs (walking low)	grade 2 glioma, small intraventricular tumor, small tumor islands at the base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
20.1j	30	head tilt, hydrocephalus, lethargy	hydrocephalus, no tumor	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
25.3b	40.6	head tilt, hydrocephalus, lethargy	ventricular tumor and tumor islands at the base of brain, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
20.3d	41.4	moderate hydrocephalus, head tilt, lethargy	grade 2 glioma, small tumor islands at the base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
48.1g	23.9	head tilt, hydrocephalus, lethargy	grade 2 glioma, ventricular tumor and tumor islands at the base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
21.1k	41	hydrocephalus, abnormal gait, lethargy	grade 2 glioma, small tumor islands at the base of brain	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
22.2l	46	hydrocephalus, paralysis of hindlimbs	ventricular tumor and tumor islands at the base of brain, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
25.3f	42.7	severe hydrocephalus, body tilt, lethargy	ventricular tumor and tumor islands at the base of brain, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
51.2h	35.4	moderate hydrocephalus, hyperventilation, slow	ventricular tumor and tumor islands at the base of brain, grade II glioma	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II
35.1c	43	walking low, urogenital staining	no tumor	widespread subarachnoid/ leptomeningeal spread, minimal infiltration into spinal cord but strong infiltration into nerve roots; grade II

Table 3.6. Clinical and pathological details of all *EGFRvIII*-PB mice. As with *EGFRvIII*-only mice, brain tumors showed heterogeneity in pathology, but spinal tumors were homogenous.

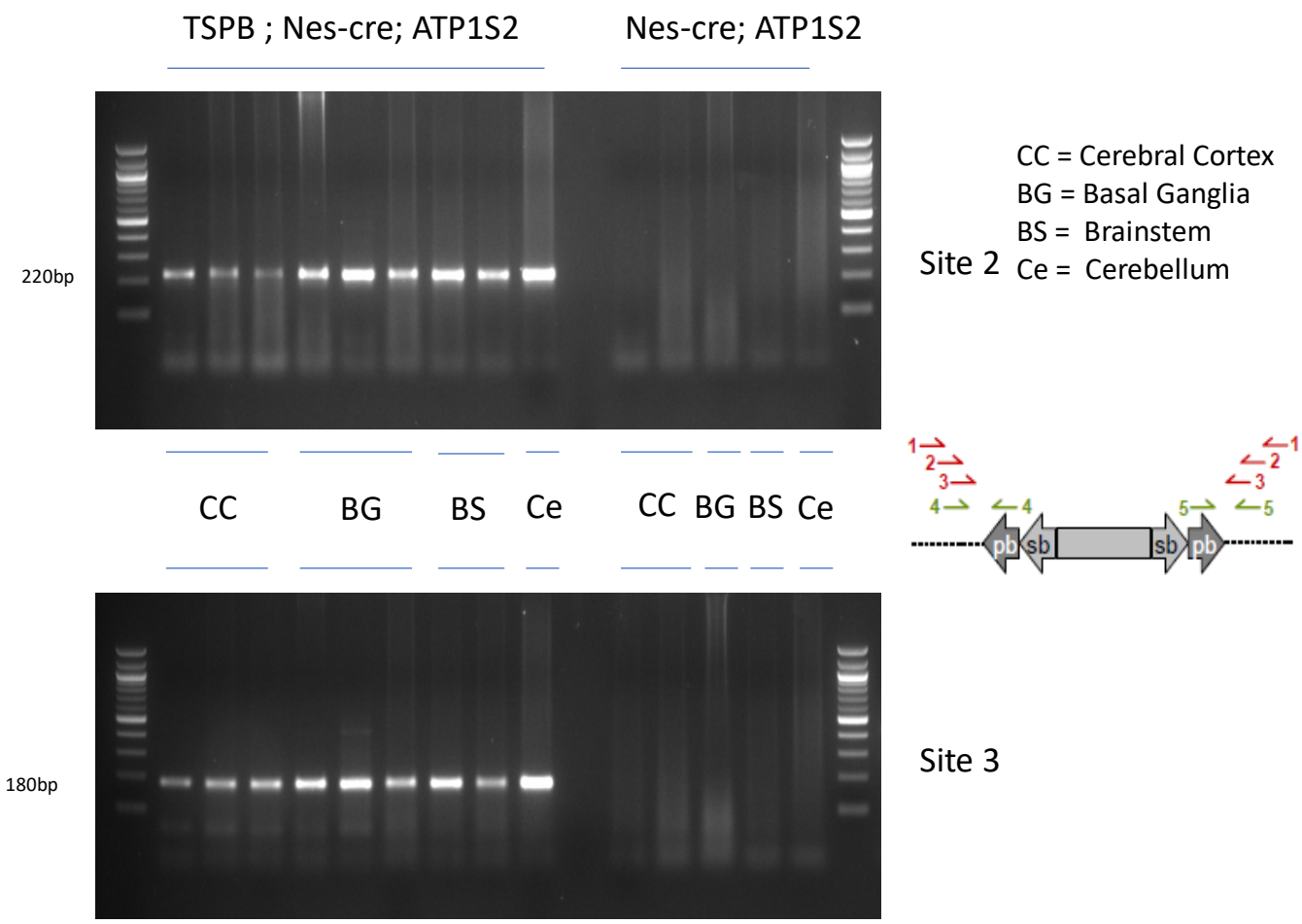


Fig 3.36. PCR to identify mobilisation of the *piggyBac* transposon in the brain. Two independent PCR experiments were performed (sites 2 and 3 as shown here), and only mice with the transposase allele showed jumping of the transposon.

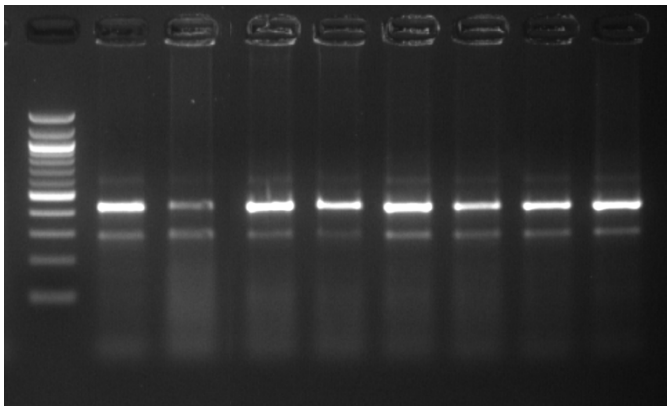
To identify mobilisation of the *piggyBac* transposon in the brain of the relevant mice, PCRs were performed as described in the Methods section. Fig 3.36 demonstrates that in mice carrying both the transposon and transposase alleles (in addition to nestin-cre), specific bands were produced that signify transposon mobilisation: a 220bp band for PCR site 2 and 180bp for PCR site 3. These bands were present for all sites of the brain tested in these mice, including the cerebral cortex, basal ganglia, brainstem, and cerebellum. In contrast, mice

lacking the transposase allele did not produce these bands, indicating that the transposon was not able to mobilise in the brain tissue of these mice.

DNA from brain tissue of the same mice in the previous experiment was also tested for non-mobilisation of the *piggyBac* transposon through a PCR as described in the Methods section. The results shown in Fig 3.37 are for PCR non-mobilisation. For mice carrying the transposon (and nestin-cre); the specific 423bp band was always present, indicating transposon non-mobilisation. This also occurred even for mice that also carry the transposase allele, indicating that mobilisation of the transposon is never complete and there is always an element of non-mobilisation. Multiple brain sites were tested (cerebral cortex, basal ganglia, brainstem and cerebellum), and the non-mobilisation band was present for all sites.

TSPB ; Nes-cre; ATP1S2

423bp



Nes-cre; ATP1S2

423bp

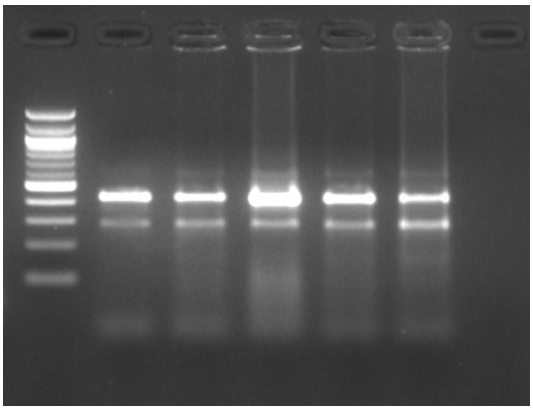


Fig 3.37. PCR for detecting non-mobilisation of the *piggyBac* transposon. This experiment was performed on mice with and without the transposase (mouse genotypes are displayed above) – both sets of mice had an element of non-mobilisation, implying that not all transposons jump even in the presence of the transposase.

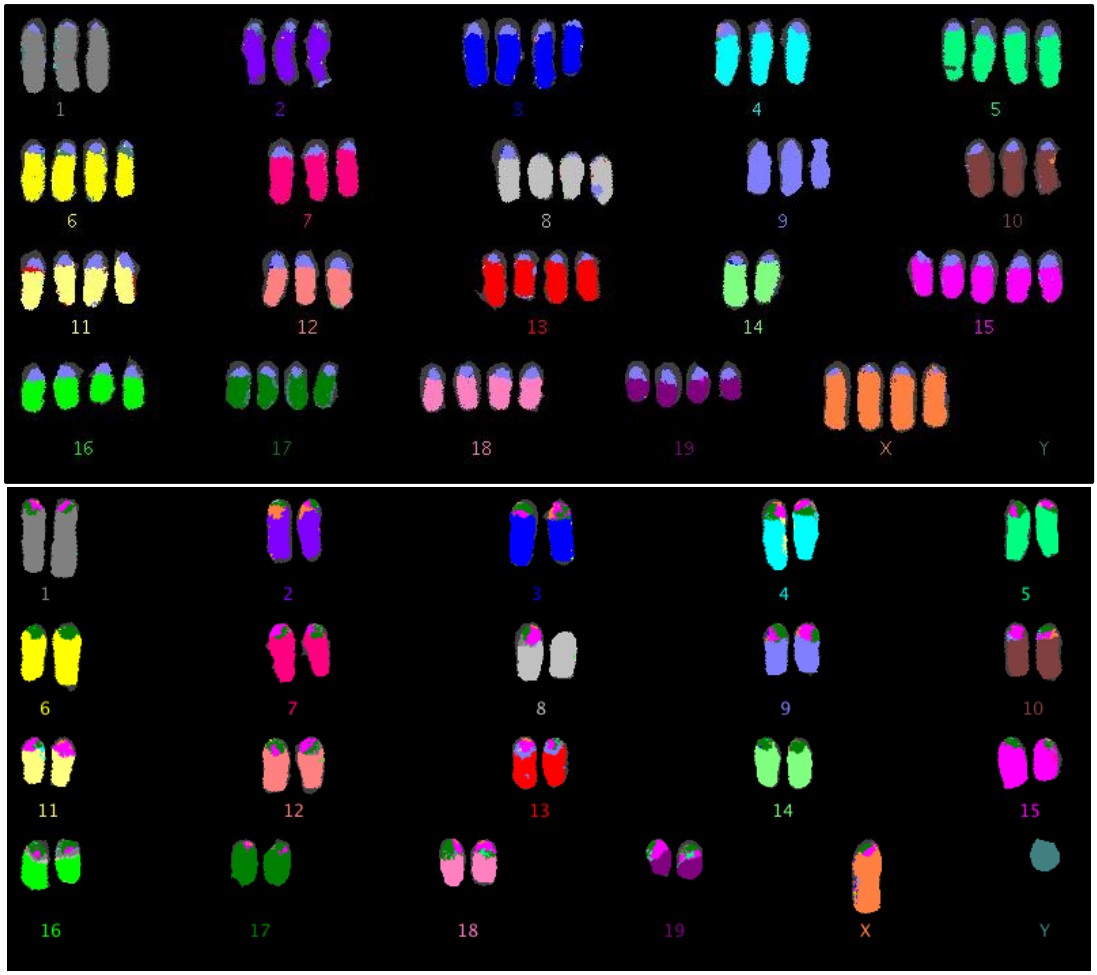
Most previous successful transposon mutagenesis forward genetic screens have found that mice with transposition tend to survive for significantly less time than those lacking transposition. We hypothesized that the absence of reduction in survival times of the *EGFRvIII*-PB mice may reflect genetic instability in the *EGFRvIII*-only mice, perhaps through oncogene-induced replicative stress[176], that is similar in consequence to the transposon mutagenesis in *EGFRvIII*-PB mice. To test this, FISH cytogenetic analysis was conducted; primary cultures were established from these tumors and cells were used from these for FISH (SKY). Chromosomal aberrations, typically appearing in subsets of metaphases rather than all metaphases, were counted as described in the Methods for 3 *EGFRvIII*-only tumors and 6 *EGFRvIII*-PB tumors. This analysis revealed a significantly higher frequency of chromosomal aberrations in *EGFRvIII*-only tumors compared to *EGFRvIII*-PB tumors (19 vs 6.4 mean number of chromosomal aberrations, $p = 0.013$, unpaired two-tailed t-test; Fig 3.38, 3.39).

Whole-exome sequencing of 20 brain and spinal gliomas from *EGFRvIII*-PB mice confirmed these had substantially less complex tumor-genomes with fewer copy number changes than *EGFRvIII*-only tumors (Fig. 3.40). Nevertheless, whole chromosome 11 amplification was still common as well as focal amplifications of *EGFRvIII* (*Col1a1* locus) and localized deletions in *Cdkn2a, Nlrp1b* and *Adgrl2* in tumors arising from both cohorts. GISTIC2 analysis shows these alterations occur significantly more frequently than expected by chance (q-value < 0.05; Fig 3.31), suggesting they provide a selective advantage for tumor progression. The wild-type *EGFR* gene was not amplified however.

Whole-exome sequencing analysis revealed that while the median number of mutations was similar between the cohorts, their mutational profiles differed substantially. The top 5 mutated genes identified in the *EGFRvIII*-PB tumors were *Obscn*, *Hspg2*, *Rrbp1*, *Rpgrip1* and *Atp5o* which have unknown functions in cancer (Fig 3.41). Although, the frequency of mutations in these genes was high (70-40%), *Obscn* and *Hspg2* are particularly large genes (so more likely to harbor mutations) and contained many synonymous changes, suggesting they were passengers. Nevertheless, in the *EGFRvIII*-PB cohort there were low-frequency mutations in a subset of the putative drivers we previously identified in *EGFRvIII*-only tumors, including frequent splice site mutations in *Sub1* and *Nt5c2*, and mutations in *Nlrp1b, Trp53,*

Tead2, Uimc1 and *Itga6* (Fig 3.42). Of note, three tumors across both cohorts had a mutation and focal deletion in *Nlrp1b,* suggesting selection for complete loss of *Nlrp1b* which may occur via multiple mechanisms.

Together, these results suggest that *PiggyBac* mutagenesis substitutes for genetic instability thus reducing copy number variation in *EGFRvIII*-PB tumors and highlights the relevance of the transposon-mediated genetic alterations for glioma progression. Mapping and analysis of *piggyBac* insertion sites in these tumors are detailed in the next Chapter.



EGFRvIII - only

EGFRvIII-PB

Fig 3.38. Representative karyotype of *EGFRvIII*-only and *EGFRvIII*-PB brain tumors, showing polyploidy in the non-PB tumor.

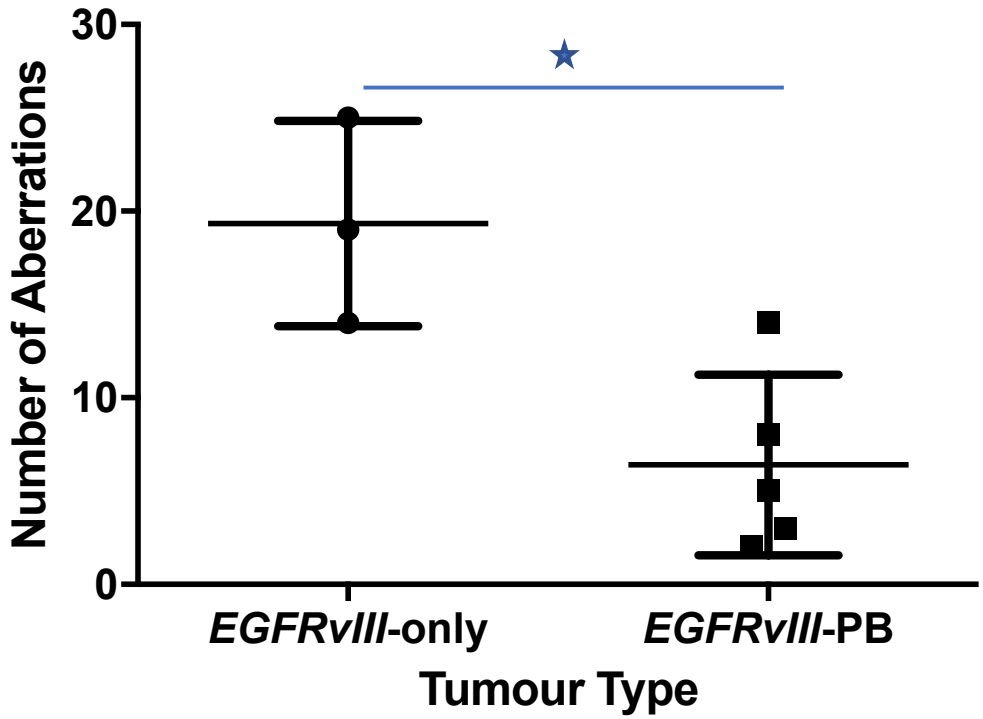
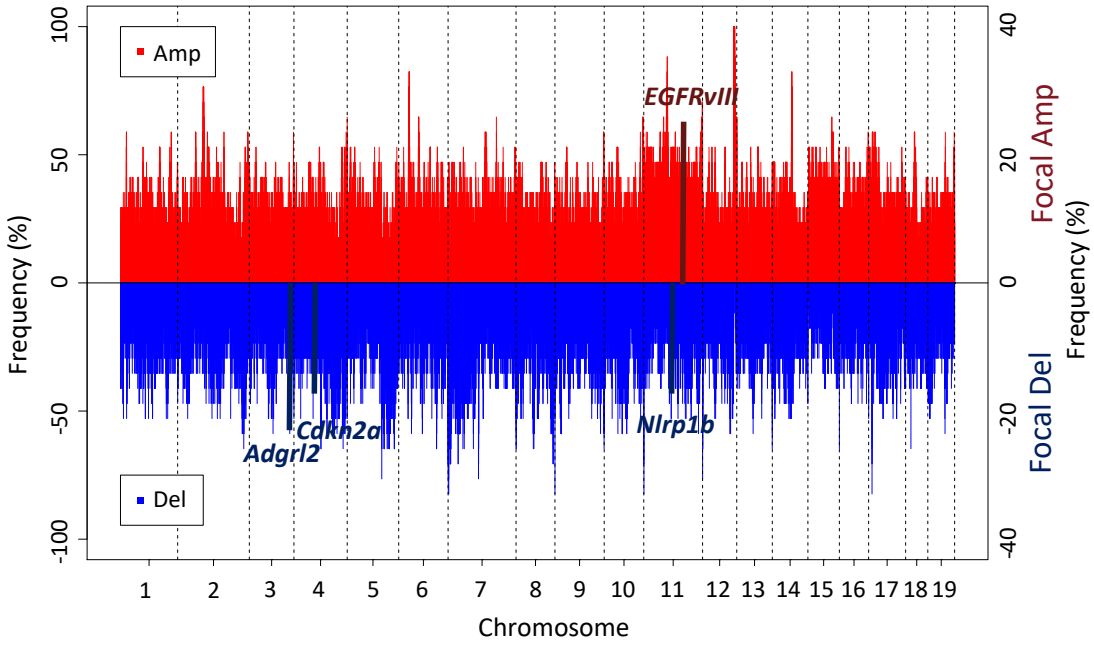
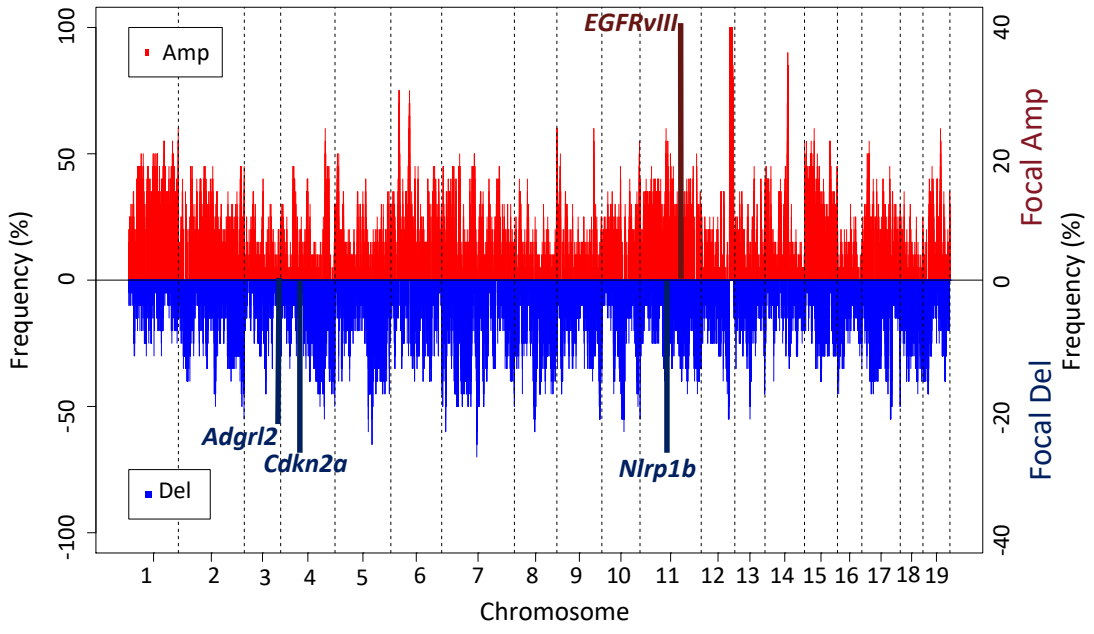


Fig 3.39. Chromosomal aberrations in *EGFRvIII*-only and *EGFRvIII*-PB tumors (n=3 and n=5 tumors respectively; mean chromosomal aberrations 19 vs 6.4, p = 0.013, unpaired two-tailed t-test; plots show mean +/- standard deviation). * denotes significance at level $p < 0.05$.



***EGFRvIII*-only**



***EGFRvIII*-PB**

Fig 3.40. Copy number analysis from whole-exome sequencing data of mouse gliomas. Top panel if from *EGFRvIII*-only mice and demonstrates there are more copy number alterations, particularly gains, in this cohort compared with *EGFRvIII*-PB mice in the bottom panel. Each row represents an independent tumor and each column is a chromosome. Red lines are copy number gains (amplifications) and blue lines are copy deletions.

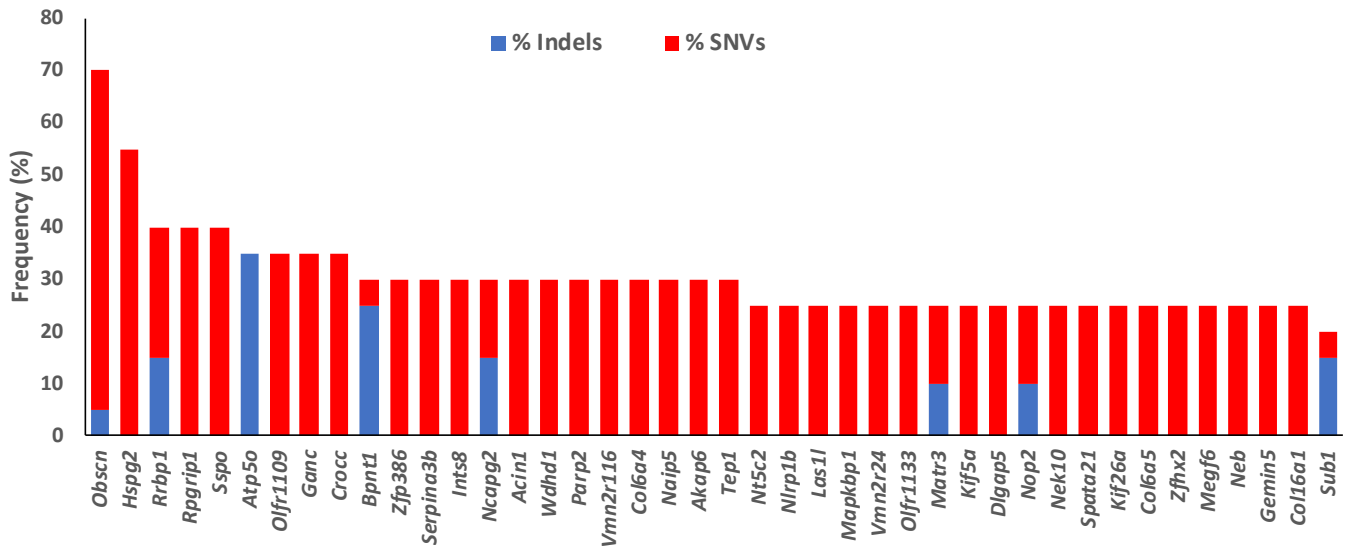


Fig 3.41. Mutational profile of 20 *EGFRvIII*-PB brain and spinal tumors from whole-exome sequencing.

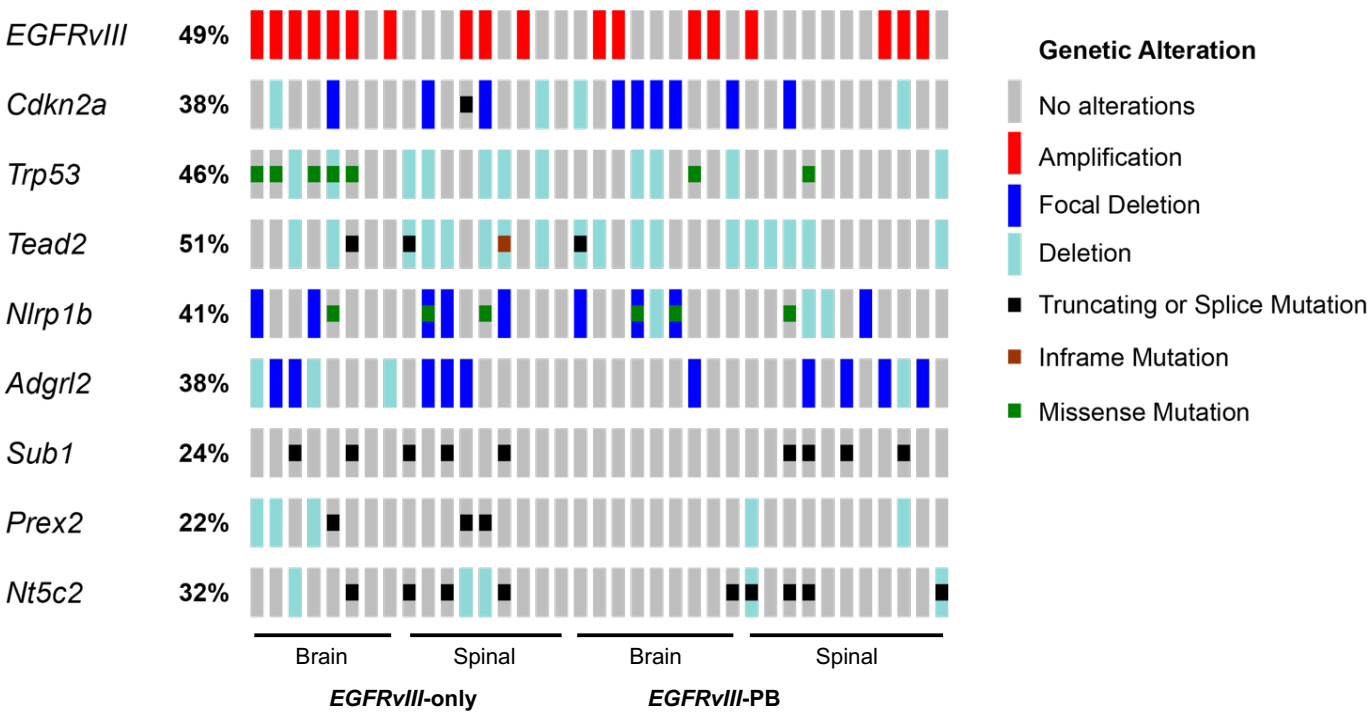


Fig 3.42. Key cancer genes identified, either as significantly mutated from MuSiC or copy number altered from GISTIC2, across all mouse brain and spinal tumors in both cohorts; each column represent one tumor.

Discussion

EGFRvIII can initiate gliomagenesis *in vivo*

In this study, we have demonstrated that *EGFRvIII* when expressed in the central nervous system using the *nestin*-cre driver, initiates glioma formation in the brain and spinal cord. In our work, the earliest glioma-like lesions were observed in the subventricular zone and at the brain surface, and later gliomas were observed both in the brain and spinal cord with 100% penetrance.. Other studies in mice have concluded that *Egfr* activating mutations alone are insufficient to generate gliomas *in vivo*, and instead they require cooperation with other mutations such as loss of tumor suppressors *Pten* and *Cdkn2a* [92, 96, 128]. In our study, microneoplasias were observed from 12 weeks of age and gliomas were observed with median latency of 36 weeks, suggesting *EGFRvIII* is sufficient to initiate gliomagenesis *in vivo*. Although here the mice only carried a single mutation (in *EGFR*), the latency for formation of gliomas was rather long. Similarly, there was a low incidence (~10%) of glioblastomas. These findings suggest that *EGFRvIII* can be the first tumor-driving genetic event in gliomagenesis but there is a requirement for activation or suppression of additional cooperative pathways for tumor progression. RNA-sequencing analysis of the *EGFRvIII*-induced tumors showed differential expression between these and normal brain and spinal cord samples, with significant enrichment for neural development pathways and oncogenic pathways such as P53 signalling and the MAPK pathway. Although *EGFRvIII* starts the tumorigenic process by stimulating cellular proliferation, activation in these other pathways must later become important in tumor formation and can at least partially explain the delay between *EGFRvIII* is first expressed and when a tumor is fully developed. Additionally, during glioma formation there is somatic acquisition by tumor cells of genetic alterations in other oncogenes and tumor suppressor genes, as shown in our whole-exome sequencing data from *EGFRvIII*-driven gliomas. This likely occurs because *EGFRvIII* is a potent stimulator of cell proliferation, and mitosis is associated with a low but non-zero rate of errors in DNA replication; if these errors occur in cancer genes, then this will set up a Darwinian natural selection process whereby mutations providing a survival advantage to the cells will be selected for, enabling evolution

of the tumor. Indeed, *EGFRvIII* has been previously associated with genetic instability *in vitro [177]*.

Microneoplasias as precursors to gliomas have been described in mice previously, for example the work by Jacques and colleagues demonstrated such lesions in mice SVZ in the presence of *Trp53* and *Pten* double homozygous loss [97]. In their study, the investigators micro-injected adenovirus expressing cre under control by the GFAP-promoter into the ventricles of $Trp53^{-/-}$; $Pten^{-/-}$ mice, which then developed microneoplasias in the SVZ followed later by high-grade gliomas, leading the authors to conclude that SVZ type-B neural stem cells (which express GFAP) were the origin of gliomas in their model.

It is worth considering why previous studies in mice did not find that *EGFRvIII* alone induced gliomas, unlike in our study. There are likely multiple reasons for this discrepancy:

1. The relatively long latency for tumor formation in our model demanded long observation times, whereas some previous studies had shorter observation times of their mice expressing *EGFRvIII*. For example, Holland and colleagues histologically assessed their mouse brains carrying this mutation after 10 weeks [92], in contrast the median latency of tumor formation in our model is 36 weeks. Similarly, Zhu and colleagues observed their mice for around only 30 weeks after induction of EGFRvIII expression [96].
2. Several studies targeted expression of *EGFRvIII* from a particular brain region. For example, Zhu et al injected cre (carried by adenovirus) into the striatum of their mice [96], and Holland et al used the replication competent ALV splice acceptor (RCAS) viral system for *EGFRvIII* gene transfer into the frontal lobes of mice. In contrast the tumors seen in our mouse model were typically in the ventricles and / or surface of the brain / spinal cord, rather than deep parenchymal lesions; the *nes-cre* driver used here targets neurogenic niches.
3. Previous studies on *EGFRvIII* in mice did not report assessment of the spinal cord, whereas mice in this study developed spinal gliomas with 100% penetrance.

4. Some studies induced expression of *EGFRvIII* in 'adult' mice, for example Klingler and colleagues employed a tetracycline-inducible *EGFRvIII* with the mutant allele induced from mice aged 4 weeks[128], and Zhu et al induced expression of the allele with cre from mice aged over 3 months[178]. Other studies have suggested that the age at which the genetic lesions occur affect tumor latency – Llaguno and colleagues from Luis Parada's laboratory showed that the latency for glioma formation from neural progenitors in mice with *Nf1*, *Trp53* and *Pten* was substantially reduced when these genetic alterations occurred from embryonic stages compared with 4 week old mice[179]. The *nes*-cre driver employed in our model induces recombination from embryonic day 13.5, and this cre line was similarly employed in the study by Llaguno et al.
5. The mice in this study are of mixed background, which may influence cancer phenotypes. However, the complete penetrance of gliomas in our mice provides strong evidence for *EGFRvIII* being the key initiating drive for these tumors. Interestingly, a recent systematic review of many published studies found no evidence for inbred mice (lacking genetic heterogeneity) having greater stability in phenotypic traits compared with outbred mice [180].

The cell of origin of gliomas is not the focus of our work, but it is worth considering this issue and how future studies can help elucidate this question for our model. Tumors in our mice may originate from a single source such as the SVZ, or the origins may be multi-focal. The latter explanation is more likely given that there are some cases where there are tumors on the brain surface without well-developed neoplastic lesions in the SVZ. The origin of these tumors is unclear. It is possible that the brain surface tumors therefore arise for neural stem cell or neural precursors much like the SVZ gliomas; indeed, a recent study has demonstrated meningeal neural stem cells in the developing central nervous system in mice, and these cells have a matching transcriptomic profile to neural stem cells from the SVZ [181]. Further work will be needed to shed light on whether these meningeal NSCs are the origin of a subset of gliomas in these mice and in humans. Such studies are likely to use cellular lineage-tracing

approaches to directly interrogate the fate of the meningeal neural stem cells that carry oncogenic mutations such as *EGFRvIII*. Knowledge of the origin of gliomas, even if dependent on genetic context, may be important clinically for treatment. For example, understanding that *EGFRvIII*-driven gliomas arise from the subventricular zone and brain surface with subarachnoid seeding would suggest that treatments that fail to tackle these regions would not be sufficient to treat the cancer: removal of the primary tumor will still lead to recurrence because tumor cells would be left behind in the SVZ and / or subarachnoid space. Therefore, it may be sensible to add adjunctive treatments targeted to these regions if they are thought to be the origin of tumors in individual cases, such as using radiotherapy to the SVZ and intrathecal or intraventricular chemotherapy to eradicate tumor cells in the ventricular system that would otherwise lead to distal tumor recurrence.

Nevertheless, our data demonstrates that *EGFRvIII* can be the initiating driver mutation for gliomas *in vivo*. At first glance, these observations appear to be at odds with the observation in many human glioblastomas with same *EGFR* mutation that show heterogeneous expression of the mutant EGFR protein across the tumor (suggesting this mutation was a late acquisition in clonal evolution) [182]. However, these observations may be reconciled by the notion that although not all cells in the tumor express the protein, all cells may carry the mutation and some cells may simply downregulate expression of the EGFRvIII protein. This is particularly likely to be the case late in tumor evolution when other somatic mutations can act as tumor drivers making the initiating *EGFRvIII* less important for tumor progression. Several publications have described internalisation of the EGFRvIII protein in gliomas [183, 184], supporting this possibility. Our work is consistent with a study suggesting that EGFRvIII cells from GBMs have cancer stem-cell properties such as self-renewal and tumor-initiating capacity [185], although serial transplantation of tumor cells into mice is required to demonstrate this from our tumors. Our findings are also in keeping with the observations in humans that the presence of EGFR alterations in low grade gliomas is poor prognostic indicator for survival and that this mutation is much more common in glioblastomas

compared with low grade tumors, suggesting EGFR alterations are driving glioma progression in humans as well.

As mentioned previously, our understanding of the timing and prevalence of *EGFRvIII* is confounded by significant intratumor heterogeneity in gliomas. Tumor heterogeneity may even be partially maintained by EGFR itself which has been reported to actively drive genetic heterogeneity through a cytokine circuit involving IL-6 [186]. A recent study has demonstrated that most patients with who had *EGFRvIII* in their primary GBM maintain this mutation in their recurrent GBM, suggesting this mutation may be still be, at least partially, driving gliomagenesis at recurrence [187]. Studies have been conflicting on the timing of acquisition of *EGFR* mutations and amplifications in gliomagenesis. A recent report found that *EGFRvIII* was present throughout different sites of GBMs in a small patient cohort, despite not being expressed throughout the same tumors suggesting there are mechanisms for downregulating expression of the mutant receptor; moreover, it was found that demethylation of *EGFRvIII* led to re-expression of the protein implying that epigenetic mechanism underpin control of its expression [188]. These results would be consistent with our work here showing *EGFRvIII* can be an initiating event in a subset of gliomas. In contrast however, another study concluded that the *EGFR* amplification occurs early in GBM patients but that *EGFRvIII* occurred later as there was substantial intratumoral heterogeneity in its expression; there were only three patients sampled in this study, limiting the strength of these findings however [189]. Previous work also suggested that glioma cell subpopulations develop an *EGFR* amplification first and later develop a mutation in the gene [190] ; it is more difficult however to draw conclusions about the timing of these genetic alterations from studying fully formed 'end-stage' GBMs in human patients as opposed to transforming normal neural cells with these events as we have done here.

Interesting differences between the brain gliomas, including the high-grade ones, discovered in these *EGFRvIII*-mice and human gliomas is that the mouse tumors tend to have little

invasion into the brain parenchyma beyond the tumor margins; in contrast human gliomas tend to widely infiltrate brain parenchyma and can be situated at virtually any location. These differences may reflect unique features of tumors that are initiated by *EGFRvIII* as opposed to tumors triggered by different initiating mutations; it could be that distinct additional mutations need to be acquired at later stages for successful invasion into brain parenchyma. An alternative explanation for this difference is that mouse gliomas may be generally less invasive than their human counterparts.

Relationship to human brain gliomas

In human gliomas the initiating event in *IDH1*-wild-type tumors is poorly understood, yet activating mutations in *EGFR* (including *EGFRvIII*) are particularly common in these gliomas. Our intention therefore was not to model a histologically low-grade glioma versus a glioblastoma (GBM), but rather to determine the precise effect of introducing an activating *EGFR* mutation into the CNS and study the resulting phenotype and genetics. It is true that the majority of the gliomas generated in this model appear histologically low-grade, whereas in humans the majority with *EGFR* mutations are histologically GBM. Likely explanations for this include:

1. In humans very recent work has demonstrated histologically low-grade appearing, *IDH1*-wildtype astrocytomas are in fact representing *IDH1*-wildtype, early forms of GBM with their corresponding molecular features (particularly *EGFR* amplification) and poor prognosis [159, 191]. Von Deimling's group showed In a large cohort of human patients with histologically low-grade gliomas (LGGs) (n=544) that were reclassified as glioblastomas (GBMs) based on methylation profiling, *EGFR* amplification was found in 36% (n=196)[191]. They further demonstrated *EGFR* amplification as the single parameter of those tested with the highest specificity (>99%) for upgrading histologically *IDH1*-wt LGGs to *IDH1*-wt GBMs. Although this study analysed *EGFR* amplification than *EGFRvIII* specifically (which requires further study), this is the most likely subtype of glioma we have modelled with *EGFRvIII* in

mice. Consistent with this, all of our mice succumbed to their tumors by approximately 1 year of age, all tumors were *Idh1*-wildtype and transcriptomic analysis of mouse tumors showed significant enrichment for the human mesenchymal GBM gene set.

2. Some tumors in humans may first acquire tumor suppressor losses (such as *CDKN2A*, *PTEN* and other genes identified in this study) and subsequently acquire an *EGFR* mutation leading to transformation to a high-grade histological phenotype.

Eye Lesions

The *EGFRvIII* / nestin-cre mice developed eye abnormalities suggestive of underlying vascular pathology: there was a high incidence of bleeding (typically minor, but occasionally more major haemorrhages occurred) and histology showed neovascularisation of the retina and a reactive proliferation in the ciliary body. There was also apoptosis of the lens. This constellation of pathologies has not been previously reported to be induced by *EGFRvIII* expression in the eye. As discussed above, it is thought that expression of a strong oncogene such as *EGFR* may trigger a cell to activate cell death programs as a protective mechanism against tumor formation. Apoptosis of the lens therefore may be explained in this way. *EGFR* has been linked to angiogenesis, and this may be an additional reason as to why it is oncogenic given that cancers require development of their own blood supply in order to progress; the neovascularisation of the eye in these mice may be a reflection of the pro-angiogenic characteristic of *EGFRvIII*. Further characterisation of these eye lesions is outside the scope of this study.

Spinal Gliomas

It is certainly of interest that mice conditionally expressing *EGFRvIII* under *nestin*-cre control develop spinal gliomas. Very few models of spinal tumors have been published in the literature, and the genetic basis of these tumors is poorly understood not only because these

are rare tumors but also because surgical options are often constrained by the tumor invading the spinal cord limiting the extent of resection and therefore availability of material for DNA sequencing. This is the first time to our knowledge that it has been demonstrated that *EGFR* activating mutations can drive spinal gliomagenesis *in vivo*. Although histologically classified as benign, these tumors despite being rare are highly clinically relevant because they often lead to severe neurological impairments such as paralysis due to their eloquent location, and these clinical signs were observed in the mice of this study. Human spinal astrocytomas tend to strongly invade the spinal cord parenchyma; the tumors in the mice here tend to weakly invade the parenchyma but more strongly invade the nerve roots, based on histopathological analysis. This difference in pathology may be due to the species differences. The minimal invasion into CNS parenchyma seems to resemble optic pathway gliomas in humans, which are benign tumors but fervently invade the optic nerve leading to significant neurological deficits (blindness). These spinal gliomas are most likely primary tumors in the mice because in many cases there are only small and early tumor lesions in the brain but larger and more pervasive spinal tumors, making these tumors less likely to be a result of metastasis from the brain tumors. Conversely, in cases where the mice had aggressive high-grade gliomas in the brain, the spinal tumors were often low grade, in keeping with these being independent primary spinal tumors. RNA-sequencing also identified different transcriptomic profiles for brain compared with spinal gliomas in these mice, reflecting the locations in which these tumors arose. The unique location of these tumors (leptomeningeal) and benign histology share do not reflect all spinal gliomas in humans, but they are similar to a subset of tumors – paediatric leptomeningeal low-grade gliomas.

Although the frequency and nature of *EGFR* alterations in human spinal tumors remains to be determined in larger genomic studies than the few smaller scale studies thus far conducted, *EGFR* amplification and expression has been detected in a small cohort of disseminated paediatric spinal LGGs[160]. Although our mice have *EGFRvIII* as the driver, these tumors may possibly be generated by other mechanisms for increased EGFR signalling including alternative *EGFR* mutations, amplification and / or overexpression. Further work

overexpressing the wild-type *EGFR* in mice will help to address this question. In human patients, germline *NF1*-loss predisposes to spinal glioma[192] and a study of spinal gliomas detected *CDKN2A* deletion and loss of heterozygosity at 10q23 (containing *PTEN*)[193].

We hope our novel mouse models of spinal gliomas will provide further opportunities for insights into the pathogenesis of this disease and development of therapeutics *in vivo*.

Molecular alterations

Solid tumors typically contain a number of chromosomal copy number changes reflective of chromosomal instability due to loss of mitotic fidelity in DNA replication. Chromosomal instability and aneuploidy correlates with a worse prognosis and resistance to treatment [194]. Chromosomal analysis of the *EGFRvIII*-driven brain gliomas using FISH revealed a number of chromosomal aberrations. FISH is a useful technique in enabling detection of large translocations, duplications and deletions. These tumors substantial inter-tumor heterogeneity in their chromosomal changes. Certain alterations occurred in multiple tumors however, including amplification of chromosomes 11 and 15 (either through whole chromosome gains or through Robertsonian translocations), indicating that there must be a selective advantage conferred to cells carrying these chromosomal changes. One possible explanation for chromosome 11 gains is that the *EGFRvIII* transgene is on the *Col1a1* locus, which is situated on this chromosome; moreover, the mouse wild-type *EGFR* gene is itself located on this same chromosome. Whole-exome sequencing also revealed there was amplification of the *Col1a1* locus (where *EGFRvIII* transgene is situated). Gain of chromosome 11 therefore leads to copy number gains of the *EGFRvIII* transgene, amplifying the oncogenic EGFR signalling, suggesting the 'dose' of oncogenic EGFR is important for tumorigenesis. It is unclear what the reason is behind selection for chromosome 15 amplification in the gliomas from this study; although *C-myc* is a prominent proto-oncogene located in this chromosome, *C-myc* itself is not amplified according to the copy number analysis of these tumors.

We examined tumor-genomic evolution through whole-exome-sequencing and RNA-seq. Given the number of exonic mutations in each tumor was modest, genetic drivers could be discerned based on recurrent mutations as well as the impact of these on gene function. The exome-sequencing data also demonstrated that additional driver mutations are needed for glioma expansion after *EGFRvIII* mutation and amplification initiate tumorigenesis. The most frequently mutated genes *Sub1, Trp53* and *Tead2* had loss-of-function mutations in more than 20% of samples. A number of recurrently mutated genes were observed in frequently deleted regions and/or they had focal deletions, including *Tead2*, *Uimc1* and *Nlrp1b*. *Cdkn2a* and *Adgrl2* had recurrent focal deletions too. Correlation of the mouse with human glioma genetic data suggested many of the mutated and deleted genes are also altered in patients, such as recurring deletions of *TEAD2* and methylation of *NLRP1*.

To further characterise these spinal gliomas at a molecular level and to determine if they are related to the brain tumors, we compared their transcriptomic profiles from the RNA-sequencing data with those of brain gliomas from *EGFRvIII* mice. This comparison revealed that genes representing brain development were significantly upregulated and enriched in the brain tumor cohort, whereas genes involved in spinal cord processes such as sensory perception and regulation of motor activity were upregulated and enriched in the spinal tumors. These data point towards different and independent origins of these two types of tumor: the brain gliomas have an origin from cells in the brain, and the spinal gliomas originate from the spinal cord. This result is certainly in keeping with the spinal tumors being primary gliomas rather than metastases from the brain, despite that these two types of tumor often co-occurred in the same mice.

Genomic copy number analysis revealed that these tumors had very complex genomes with significant copy number changes throughout. The extent of these alterations make it hard to discern causal alterations by this method, aside from the focal changes described. Given their genomic complexity, the transcriptomes of mouse tumors exhibited many changes from normal tissue. Recurrent amplification of *EGFRvIII* was observed, suggesting strong selection for increased expression. RNA-seq analysis confirmed the endogenous mouse *Egfr* was upregulated in all tumors, implying that *EGFRvIII* signalling involves collaboration with the

endogenous gene.

Genomic instability includes both structural and numerical chromosomal abnormalities, and is key driving force and also a hallmark of cancer [195-197]. By using conditional *PiggyBac* insertional mutagenesis the expected reduction in mouse survival[70] was not observed - one explanation for this is that the chromosomal instability observed in the absence of transposition is able to provide an adequate reservoir of additional mutations to facilitate oncogenesis. Supporting this is the obvious difference in ploidy observed in tumors from the two cohorts, with reduced ploidy in tumors with PB transposition compared to those without. A previous study from David Largaespada's laboratory reported similar findings with *Sleeping Beauty* transposition in osteosarcomas [58]. A difference in the spectrum of mutations was also apparent from exome sequencing data: in the absence of the transposon the most frequently mutated genes are more plausible cancer genes, such as *Trp53*. Together, these data imply that *piggyBac* mutagenesis replaces the need for genomic instability in providing secondary molecular alterations needed to drive gliomagenesis. The reduced copy number variation in tumors with *piggyBac* mutagenesis can potentially greatly simplify interpretation of the cancer genomes: as the transposon integrations in cancer genes will be clonally selected for, sequencing for and identifying common integration sites should provide comprehensive information on the functional cancer genes at play in these tumors. This will be the focus of my next Chapter.

Deeper understanding of the molecular basis of spinal gliomas is needed in order to advance treatment options, which are currently limited. Demonstration that *EGFR* is a driver of spinal gliomagenesis suggests these tumors may be amenable to therapeutic targeting with EGFR inhibitors that have had some success in lung and colorectal cancers for example. It remains to be seen however what the frequency of *EGFR* mutations is in human spinal gliomas, and this can only occur if concerted efforts are made to collect sufficient material for DNA sequencing in a large cohort of patients likely from multiple clinical centres.

Study Limitations

One potential limitation of this work is the control used for our RNA-seq experiment; I used SVZ tissue dissected from a *nes*-cre mouse lacking *EGFRvIII.* A more accurate control would be the cell of origin of the gliomas we produced in our model, such as the neural stem cell. However, as mentioned earlier, our aim was not to study the glioma cell of origin in this investigation and thus the cell of origin is currently unknown for our model; using cerebral cortex, although admittedly will contain a mixture of cell types, avoids making an assumption about the cell of origin here. Future work is warranted to answer this question; one possibility is to compare the RNA-seq data from the brain gliomas with gene expression profiles of different 'normal' brain cell types , including neural stem cells, oligodendrocyte precursor cells, astrocytes and neurons , in order to determine the most similar normal cell type as this may point towards the cell of origin. Alternatively, lineage tracing studies such as mosaic analysis with double markers as utilized by Liu and colleagues (see Introduction, [101]) can greatly help in identifying the cell of origin in the model we have generated here.

A potential limitation is that one of the EGFR antibody we used for immunohistochemical staining of tumors was not specific for the *EGFRvIII* mutation, so may not accurately define the proportion of tumor cells carrying this mutation. The monoclonal antibody we used for immunohistochemistry was manufactured using the human EGFR protein (purified from A431 cells, see https://www.abcam.com/egfr-antibody-31g7-ab218383.html and[96]) as an immunogen, with specificity for human over mouse EGFR, but further work would be needed to assess any cross-reactivity with mouse EGFR protein. The antibody cannot be staining human EGFR-wild type in these tumors - we only introduced the human *EGFRvIII* transgene in these mice (not the human EGFR-wild type gene). We confirmed this by PCR genotyping with primers spanning the junction between exons 1 and 8 that is created upon deletion of exons 2 -7, yielding the expected 670 bp fragment for *EGFRvIII.* However, to further confirm protein expression of EGFRvIII in tumor cells, we also sourced an EGFRvIII monoclonal antibody and this confirmed expression of the recombinant protein in glioma cells.

Another limitation of this work is that, although there are recurrent mutations in certain genes in the tumors generated from our model, further functional studies are required in order to definitively prove that some of these are driving tumorigenesis. Such work may include, for example, CRISPR knockout or siRNA knockdown of individual genes in cell lines or directly in mice to demonstrate these accelerate glioma development. However, the presence of recurrent genetic alterations, including many in known glioma tumor suppressor genes (*Trp53* and *Cdkn2a*), provides strong statistical support for the notion that, at least some , of these mutations are contributing to gliomagenesis.

One challenge in interpreting pathology of the brain of mice is distinguishing between the normal rostral migratory stream (RMS) and small tumors in the same location (such as SVZ), as these can look similar. The RMS is a specialised migratory route by which neural precursors migrate from the SVZ to the olfactory bulb in mice and certain other mammals. However, we overcame this challenge by taking advantage of an expert neuropathologist with substantial experience in interpreting normal and pathological mouse pathologies, examining mouse brains in exactly the same way (see Materials and Methods) with four coronal slices per brain, and blinding the pathologist to mouse genotype. Given that all *EGFR*-mutant mouse brains displayed microneoplasias or full tumors, many of which were also in locations the RMS does not reside (such as third ventricle and multiple locations on the brain surface), many mice had multiple such lesions and control mice had none, and the fact that older mice had full tumors in the same location (suggesting these had progressed from the earlier precursors), make compelling arguments that the phenotypes we are observing are truly tumors rather than normal variants of the RMS.

Another limitation of the study is that, although RNA-sequencing was performed and gene set enrichment analysis demonstrated enrichment for key oncogenic signatures, downstream signalling pathways in our tumors were not studied in detail. In particular, activation of specific signalling proteins, such as Akt, Erk, Mek, and Stat3, was not tested, largely because the focus of our work here is the role of genetic alterations in tumorigenesis. However, investigating the downstream signalling proteins that are activated in these tumors will be important in future work to further understand the molecular driving forces of gliomagenesis.

Conclusions

In this Chapter, I have demonstrated *EGFRvIII* can initiate gliomagenesis in the brain and spinal cord, with the long latency for tumor formation reflecting the need for additional secondary molecular alterations to be present as well. Through whole-exome sequencing, we identified significantly recurrent mutations present in these tumors, suggesting a landscape of genes contribute to tumorigenesis. Chromosomal instability was observed in these tumors; however, in the presence of *piggyBac* transposition this instability was reduced, implying *piggyBac* provides secondary mutations needed for gliomagenesis instead. Such mutations will be the focus of the next Chapter.

Chapter Four: Genome-Wide *PiggyBac* Transposon Screen for Genetic Drivers Co-Operating with *EGFRvIII* for Gliomagenesis *in vivo*

Abstract

EGFR is recurrently mutated or amplified in gliomas, in addition to many other cancers, and represents a clinically important therapeutic target. However, the genes that cooperate with *EGFR* in driving gliomagenesis are poorly understood. It is also unclear whether such driver genes differ between brain and spinal gliomas. Here, we performed an *in vivo* genome-wide screen using *piggyBac* transposon mutagenesis in mice carrying the *EGFRvIII* mutation. Sequencing of 96 resulting brain and spinal gliomas identified 281 significant common integration site (CIS) genes. The top CIS genes included known *EGFR*-cooperative partners and established glioma drivers such as *Cdkn2a*, *Pten* and *Nf1*, highlighting the validity of this approach. Brain and spinal gliomas shared a CIS-genetic profile, showing these tumors share truncal drivers such as *Cdkn2a* and *Pik3r1.* Several of the top CIS genes are novel mutated genes involved in neural differentiation, such as *Sox6*, *Sox5* and *Tcf12*; analysis of large-scale human glioma sequencing data shows that many of these genes are also recurrently altered in human tumors, implicating them as tumor suppressor genes. Expression levels of *SOX6* and *TCF12* are also prognostic for glioma patients, highlighting the clinical relevance of these genes.

Introduction

Large-scale sequencing projects, including The Cancer Genome Atlas for example, have been invaluable in establishing the genetic landscapes of gliomas. These have identified a number of recurrent driver mutations in genes such as *EGFR*, *TP53*, *PTEN*, *RB*, as explained in the Introduction to this thesis. However, these studies demand complementary detailed functional analyses of the biology behind the contribution of these genes to gliomagenesis. Moreover, there are great numbers of transcriptional and epigenetic alterations found in these tumors whose role in tumorigenesis are unknown. Although computational methods for determining which genes are true genetic drivers of cancer rather than passengers are improving, very large numbers of human tumors are needed for doing so for genetic changes that occur less frequently [198]. Complementary approaches are therefore required to prioritise which genetic and epigenetic alterations are important in driving tumorigenesis. A fruitful approach over the last decade in enabling better interpretation of human sequencing data is through *in vivo* transposon mutagenesis forward genetic screening, typically in mice. *PiggyBac* transposition has recently been developed as a conditional *in vivo* screening approach, Fig 4.1; this has had success in identifying genes that contribute to pancreatic carcinogenesis [70]. Although it is a powerful cancer screening platform, conditional *piggyBac* mutagenesis has not been previously applied to central nervous system tumors. *PiggyBac* has a tendency to insert into open chromatin regions, which also gives the advantage of enabling identification of non-coding regions that may contribute to cancer. For example, a *Cdkn2a*-cis regulatory regions was identified as a contributor to pancreatic cancer due to *piggyBac* insertions.

An additional advantage of conducting transposon mutagenesis screens in mice is that it allows for identification of genes that cooperate with a known cancer gene. By predisposing to cancer initiation using a mouse with tumorigenic allele, such as a *Trp53* mutation, and then crossing in transposon and transposase alleles, one can sequence for and map the transposon insertion sites of resulting tumors to elucidate the genetic driver mutations that cooperate with the predisposing mutation in cancer progression.

Therefore, in order to assign functional roles to alterations in human gliomas, we have conducted a transposon-based insertional mutagenesis screen in mice that allows identification of functional driver mutations for gliomas *in vivo*.

Most DNA transposons have a tendency for local hopping, that is excision and reintegration of the transposon in a neighbouring region of the same chromosome. A study which examined the properties of PB in mouse embryonic stem (ES) cells in some detail demonstrated that PB local transposition frequency is substantially less than that of SB, which is certainly advantageous when conducting unbiased genome-wide screens [199]. The same study showed that the distance of local transposition of PB (100kb) is also much lower than that of SB (5MB) in mouse ES cells. Interestingly, other studies that examined the characteristics of PB in *Drosophila* and in the mouse did not find any local hopping [199, 200]; this difference may be due to the fact that the Wang et al study transiently expressed PBase in mouse ES cells, whereas the *in vivo* studies has constitutive expression of PBase or continuous induction in the germ line allowing for multiple rounds of transposition that may mitigate any local hopping by enabling distant transposition to occur. Wang et al also found that the reintegration rate of excised PB transposons was around 40% in mouse ES cells. A later study examined other features of PB in mouse ES cells: Li et al determined that although PB has a clear preference for inserting into sites TTAA, it also can insert into other regions containing TA within a broader GC rich context but not necessarily being TTAA. It inserts into sites other than TTAA with a frequency of 2%, and such sites include CTAA and ATAA; the only absolute requirement for PB insertion being for the central TA [201]. These insertions in non-TTAA sites introduce nucleotide mismatches, and these are repaired with host cell DNA repair pathways. Importantly this study also demonstrated that PB integrates into expressed genes; this is particularly useful in cancer forward genetic screens as compared to pure cancer genome sequencing wherein non-expressed genes such as olfactory genes can acquire many passenger mutations that make it more difficult to identify true driver mutations. Open chromatin structures are needed for PB insertions, which generally do not occur in heterochromatin.

The efficiency with which PB can be excised and reintegrated in the genome, combined with the low rate of local hopping, make it particularly useful for genome-wide screening when PB

is engineered to contain gene-trapping cassettes. This underpins our decision to use this tool for performing an unbiased screen for drivers of gliomas *in vivo* in a conditional mouse model.

As discussed in the previous chapter, we have demonstrated that *EGFRvIII* is sufficient to initiate glioma formation from the SVZ and brain surface in mice. However, the latency for tumor formation in this model is long, and the tumors lack the ability to invade the brain parenchyma. Moreover, these gliomas have features consistent with low-grade gliomas. These findings suggest that additional genetic events are required in order to transform *EGFRvIII*-driven gliomas into invasive and / or high-grade gliomas. As a way to identify such co-operative genetic events, we have performed an *in-vivo* genome-wide screen using *piggyBac* transposon mutagenesis in mice. Having previously shown that *EGFRvIII* can initiate gliomagenesis not only in the brain but also in the spinal cord, we aimed to demonstrate the cooperative genetic events needed to drive glioma formation in the spinal cord and compare these with tumor formation in the brain. Genomic studies focused on the driver events in spinal gliomas are limited for multiple reasons, including the relative rarity of the disease, the small size of the tumors placing a limit on the availability of material for genetic sequencing, and the difficulty associated with complete resection of tumor due to its dangerous location. As such, we have a relatively limited understanding of the genetics of this disease and much of our understanding has come from sequencing of brain gliomas and speculating on the relevance of these finding to their spinal counterparts [162]. In this study, we have established a landscape of putative genetic driver events for both *EGFR*-mutant brain and spinal gliomas, allowing a direct comparison to be performed of the genetics of these two important diseases.

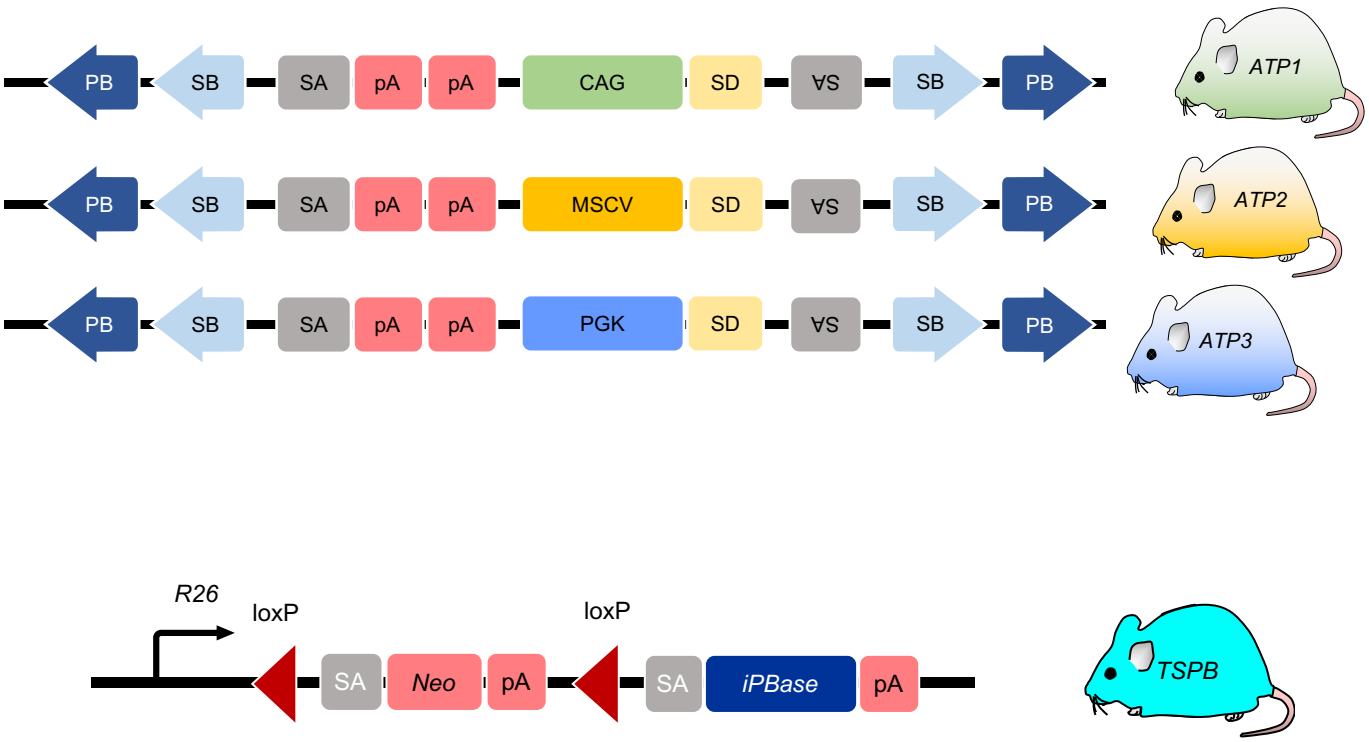


Figure 4.1. A conditional *piggyBac* system for *in vivo* forward genetic screening. Top panel shows the different ATP *piggyBac* transposon mouse lines generated by Rad et al [70]. These lines differ in their promoters for driving expression of the transposon; ATP1 contains a CAG promoter, and this is the line we employ in this study, in particular the ATP1-S2 tranposon line, with 20 copies per cell. CAG promoter = cytomegalovirus early enhancer element, promoter/first intron/ first exon of chicken beta-actin gene, and splice acceptor of rabbit beta-globin gene. MSCV = murine stem cell virus promoter. PGK = phosphoglycerate kinase promoter. Bottom panel shows the configuration for the conditional *piggyBac* transposase allele, targeted to *Rosa26* (TSPB). The *LoxP* sites and the contained stop cassette (reflected by the polyA tail that terminates transcription) are removed upon expression of cre, leading to expression of the downstream transposase sequence. SA = splice acceptor; SD = splice donor; CAG = CAG promoter; SB = Sleeping Beauty; PB = *PiggyBac* inverted repeats; iPBase = insect version of the *PiggyBac* transposase. The transposon can activate gene transcription if it inserts in the same orientation as the gene, usually in a 5' position. Gene inactivation can occur if the transposon inserts in the body of the gene as a consequence of gene trapping which can occur in either orientation because of the presence of two splice acceptors and poly(A) (pA) sites. Figure has been adapted from [122].

Aims of Study

In the previous Chapter, we established a novel mouse model of brain and spinal gliomas driven by the expression of *EGFRvIII* by the nestin promoter in the CNS, and used whole-exome sequencing to identify secondary molecular alterations acquired during tumorigenesis. When *piggyBac* transposition is introduced into this model, also driven by the nestin-cre expression, there is a reduction in large genomic changes, suggesting that transposition substitutes for genomic instability during tumor progression. The implication of this is that *piggyBac* transposition is providing the secondary molecular changes to drive oncogenesis instead of large chromosomal aberrations that provide these in the absence of transposition. Indeed, similar findings have been noted for *Sleeping Beauty* transposition in mouse models of other cancers [58].

In this Chapter, I have sequenced and mapped the insertion sites of *piggyBac* from 96 gliomas, with the aims of identifying common integration sites (CIS) of the transposon that are likely to represent functional drivers of gliomagenesis. To support these data, I analyse RNA-sequencing data from these tumors to identify gene-transposon fusion transcripts, as direct evidence of transposon insertions affecting their target gene expression. Lastly, I compare the CIS with publicly available datasets of gliomas from large cohorts of human patients to determine the frequencies of alterations in CIS genes in these patients.

Results

Common Insertion Sites Across All Gliomas

Gene	Chromosome	Start	End	Total IS	Total samples	Total reads
Cdkn2a	4	89096058	89489079	99	47	233719
Nf1	11	79287193	79619803	47	29	18100
Ppp1r14c	10	3309887	3724803	43	29	3155
Pten	19	32678851	32905851	30	20	110519
Sox6	7	115548304	115871683	29	19	8971
Map7	10	20073797	20308872	22	20	15183
Adgrl3	5	81048981	81303956	20	19	6741
Asap1	15	64110005	64371418	20	17	2096
Sox5	6	143942931	144227797	20	19	9252
Ccna2	3	36430026	36675616	19	19	10065
Csmd3	15	47486085	47660361	18	18	21497
Exosc9	3	36447497	36668433	18	18	9996
Spred1	2	117042029	117209386	18	13	3570
Clcn3	8	60850884	61027199	17	14	1925
Ctnnd2	15	30484570	30707255	17	15	9065
Pik3r1	13	101593163	101847492	17	16	2457
Ust	10	8261283	8476769	17	16	2330
Snx29	16	11314621	11548051	16	15	10056
Dmd	X	84744328	85004890	15	15	14998
Slc8a1	17	81501469	81695955	15	15	5217
Tcf12	9	71862595	72068316	15	13	1263
Zfat	15	67719615	67820188	15	11	14253
Zfhx4	3	5239819	5286895	15	8	12633
Csnk1g3	18	53839007	54013921	14	13	805
Nova1	12	46671946	46867497	14	14	11444
Nrip1	16	76237345	76451322	14	14	10157
Phlda1	10	111429778	111566905	14	14	4974
Tnr	1	159624066	159831238	14	13	2494
Asb16	11	102211607	102341521	13	10	16049
Epn2	11	61453428	61639298	13	12	1885
Nav3	10	109682551	109747074	13	13	20475
Ptprj	2	90471620	90658666	13	13	2422
Qk	17	10222528	10426738	13	13	1014
Tcf4	18	69435503	69610417	13	12	1290
Tmub2	11	102200050	102358455	13	10	16049

Table 4.1. Table of the top 35 significant common integration sites for *piggyBac* across 96 brain and spinal gliomas. 'Start' and 'End' refer to the chromosomal locations of the common insertion

site. Total IS refers to the total number of insertion sites; in some cases the number of IS exceeds to the total number of samples because some samples have multiple different insertions within a gene.

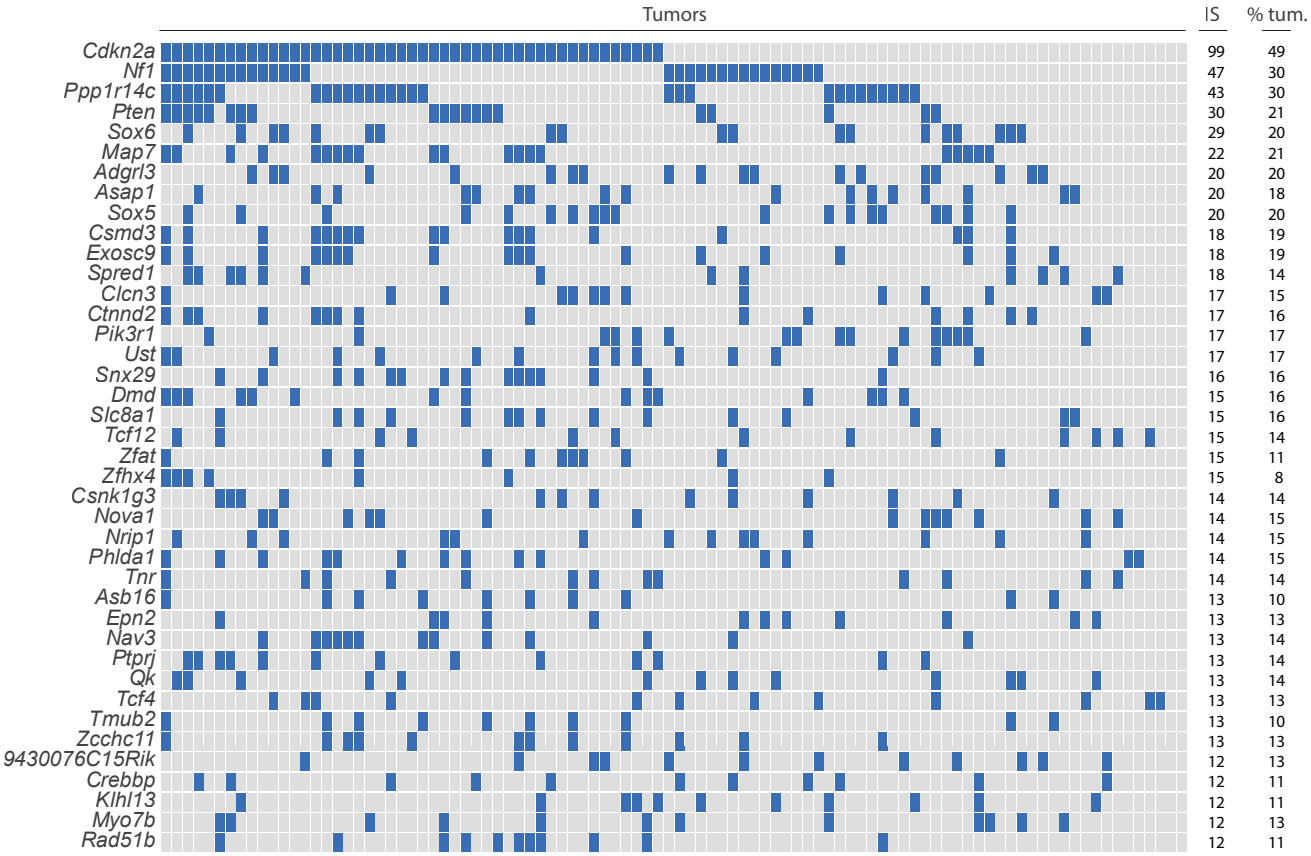


Fig 4.2. *PiggyBac* transposition identifies 281 known and novel genes cooperating with mutant-*EGFR* in brain and spinal gliomas. Oncoprint showing the top CIS transposon genes across all 96 brain and spinal gliomas (Bonferroni adjusted $p < 0.01$ for each gene, Gaussian Kernel Convolution analysis). 'IS', total number of insertion sites; '% tum.', percentage of tumors with an insertion in corresponding gene. The most well-known brain glioma tumor suppressors are amongst the top 4 genes (*Cdkn2a*, *Nf1*, and *Pten*). Novel glioma genes include *Sox6*, *Spred1*, and *Tcf12*. Each column is one tumor and each row is one gene, ranked according to the number of insertions present per gene across all tumors. A blue square indicates the presence of a *PiggyBac* insertion for a particular gene in a given tumor; a grey square indicates the absence of such an insertion.

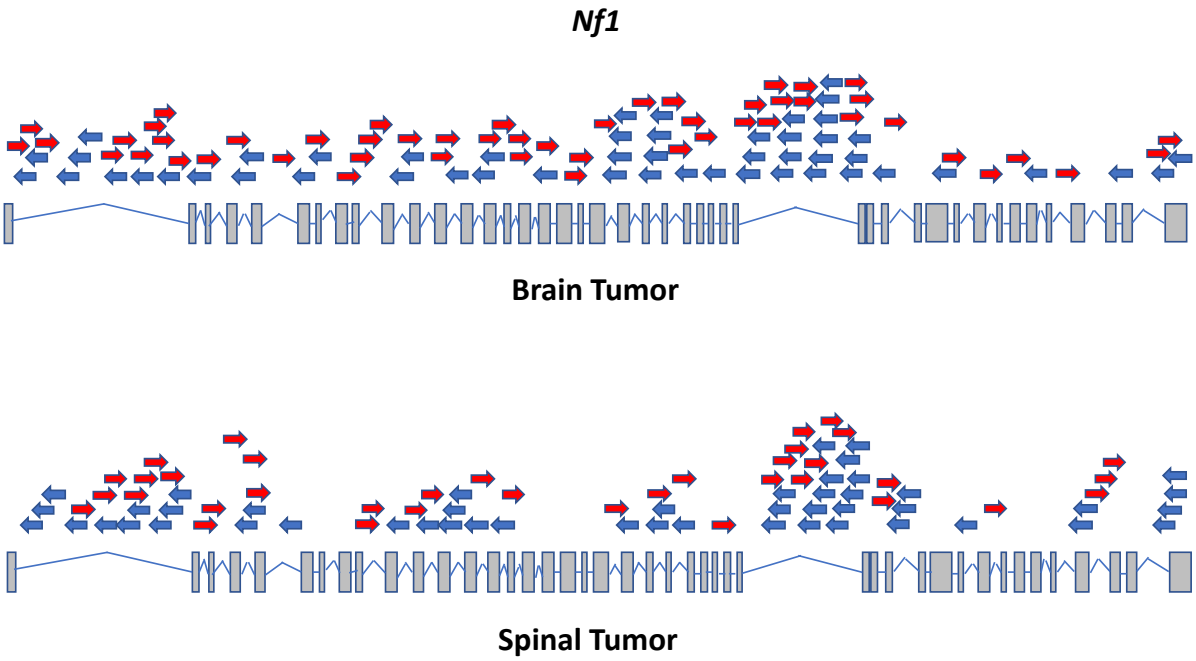


Fig 4.3. The position of all transposon insertions across *Nf1* (a known brain tumor driver) in *EGFRvIII*-PB brain and spinal gliomas, showing a gene-disruption insertion pattern. Blue arrow = antisense orientation; red arrow = sense orientation with respect to gene direction.

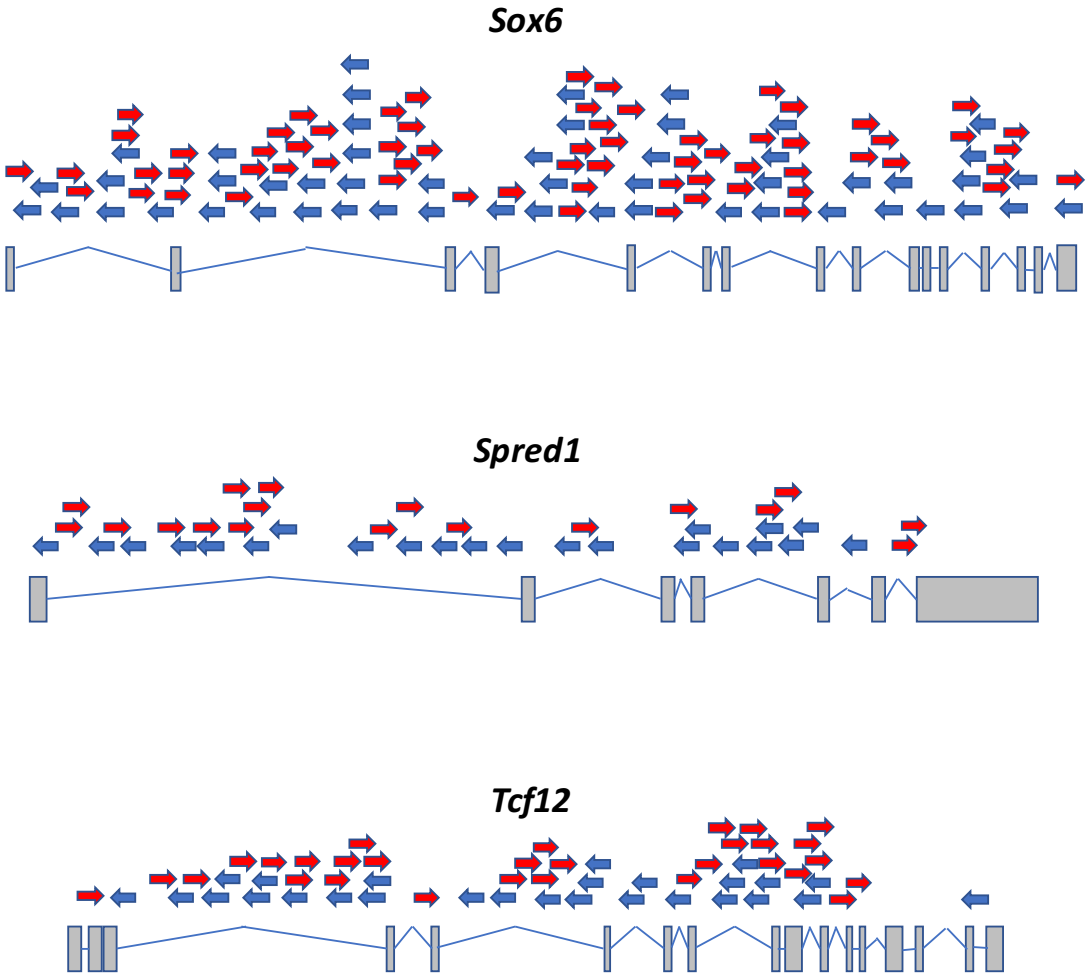


Fig 4.4. Novel putative glioma drivers, *Sox6*, *Spred1*, and *Tcf12* also have disruptive insertional patterns. These figures show all *PiggyBac* insertions in brain tumors.

Transposon Mutagenesis Identifies *EGFR*-Cooperating Driver Genes

To identify genetic driver mutations with *piggyBac*, common integration sites (CIS – genes into which the *piggyBac* transposon has inserted more frequently than expected by chance, $p < 0.01$, Gaussian kernel convolution method with Bonferroni correction for multiple testing) were identified by transposon-host PCR[122] and sequence analysis (quantitative insertion site sequencing, QI-seq). Gaussian kernel convolution was used to identify CIS from 46 brain and 50 spinal tumors[70]. In total, 281 significant CIS were ranked according to the number of insertions across all tumors (Fig 4.2, Table 4.1). A full list of all CIS identified in the combined brain and spinal glioma set is provided in the Supplementary Table 6. The CIS are ranked according to the number of insertion sites identified in these regions. As can be observed, the CIS are on many different chromosomes, and generally have more than 1000 reads each. Analysis of integrations sites in brain and spinal tumors from the same mouse for 5 consecutive mice revealed in no case was there 1 shared transposon integration in a CIS gene, confirming these tumors arose independently.

The highest-ranked CIS was *Cdkn2a*, followed by *Nf1 (*Fig 4.3). Loss-of-function mutations of *CDKN2A* and *NF1* have been observed drivers in a range of human gliomas including GBM[178, 202]. These are therefore important positive controls, highlighting the validity of our screen in identifying *EGFRvIII*-cooperative glioma driver genes. Interestingly, *Spred1,* a paralog of *Nf1*, whose product acts as negative regulator of the Ras pathway[203], ranked within the top 10 CIS and exhibited a disruptive *PiggyBac* insertional pattern, suggesting *Spred1* acts as a novel tumor suppressor in glioma (Fig 4.4). In humans, germline mutations in *NF1* cause a neurofibromatosis syndrome, with features such as café-au-lait spots, axillary freckling and frequently optic gliomas [170, 204]. Germline mutations of *SPRED1* have been described more recently to cause Legius syndrome, which has some of the skin features of neurofibromatosis but lacks many other features including glioma formation [205].

Genes involved in the PI3K-AKT oncogenic pathway were also identified including known tumor suppressor genes in GBM such as *Pten[206]* and *Pi3kr1[207]* as well as novel genes such as *Prex2* and the protein tyrosine phosphatases *Ptpro* and *Ptprj*, all with inactivating transposon insertional patterns. The glioma oncogene and PI3K-AKT activator, *Pdgfrα[208]*, was also a CIS, with an insertional pattern consistent with gene activation (Fig 4.5). This

supports the validity of our transposon screen in identifying both tumor suppressor genes and oncogenes.

Several top CIS genes known from their function in nervous system development were not previously recognized as tumor suppressors. *Sox6* and its paralog, *Sox5,* are expressed in a mutually exclusive pattern during brain development[209] - both were identified as CIS. *Tcf12* and *Tcf4,* transcription factors implicated in neurogenesis[210], were also identified as CIS. *Nav3,* a gene belonging to the neuron navigator family predominantly expressed in the nervous system, was identified as a CIS too. *NAV3* has recently been implicated as a tumor suppressor gene in breast cancer, and deletions in this gene have so far been found in a few human gliomas [211]. Their inactivating transposon insertion patterns suggest tumor suppressor roles for these genes (Fig 4.4).

Other genes of interest included *Rad51b*. *RAD51B* may have a role in sensing DNA-damage, as experiments showed that overexpression of this gene causes a delay in cell cycle G1 progression and increased apoptosis in response to DNA damage, and thus may function similarly to *TP53* ([212]. This gene was an important CIS in our screen, implicating loss of *Rad51b* as a driver of glioma progression.

To explore the evolutionary mechanisms underlying brain gliomas in our mouse model, we sampled three independent sites in each of two tumors and performed QI-seq, Fig 4.6a, b. Shared (clonal) insertions between all regions for each tumor identified putative truncal drivers. Tumor A; *Map7*, *Exosc9* and *Nav3* and tumor B; *Adgrl3*, *Begain* and *Pdgfra*.). With the exception of clonal *Pdgfra* insertions in one tumor, transposon insertions in MAPK and PI3K pathway genes (including *Nf1*, *Pten*, *Pik3r1* and *Ptprj*) were subclonal in these tumors, implying these were late evolutionary events. There were also distinct (subclonal) insertions in each region revealing intratumor heterogeneity, as observed in patients. While the biological plausibility of some of the less frequently mutated genes cannot be adequately assessed from this small sample set, these data implied alternative branching tumor evolutionary routes following an initiating *EGFRvIII* mutation in individual tumors, Fig 4.6c, d.

Further studies to analyse more tumors with multi-region sequencing for *piggyBac* insertions are warranted to help identify functional drivers and evolutionary routes for oncogenesis.

To understand how the CIS genes may interact with one another, we performed a protein network analysis using all of the 281 significant CIS genes from the brain and spinal tumors using STRING (see Materials and Methods). This analysis demonstrated these CIS genes are significantly more functionally connected to each other than would be expected by chance (Benjamini-Hochberg adjusted $p = 4.88 \times 10^{-13}$, Hypergeometric test). 253 functional connections could be drawn between this set of CIS genes. The network of CIS genes can be seen in Fig 4.7, which shows that the most well-known human glioma driver genes, *Cdkn2a*, *Pten* and *Nf1* are centrally located with the most connections with other CIS genes, along with other novel putative glioma genes we identified in this screen such as *Spred1*, *Tcf12*, *Rad51b* and *Dmd*. Collectively, these findings show that PB mutagenesis enriches for mutations that affect functionally interacting proteins in gliomagenesis.

Gene ontology analysis revealed enrichment for multiple pathways in our CIS gene list with low false discovery rates (FDR < 0.02, Fisher's exact test). Pathways which are particularly enriched are those related to neurogenesis and cell differentiation, including neural differentiation, and those related to cancer processes such as regulation of cell migration, cell metabolism and phosphatidylinositol-mediated signalling, Fig 4.8, Table 4.2.

Known oncogenic pathways were frequently altered by transposon mutagenesis and / or spontaneous genetic changes in these tumors, including not only the PI3K-Akt and Ras pathways, but also the Wnt pathway, chromatin regulators, stem cell and neural differentiation pathways, and DNA-damage response pathways; these results complement the transcriptomic profile of *EGFRvIII*-mutant gliomas showing similar oncogenic pathways are active and cooperate with EGFR. Previous studies have proposed roles for several of these pathways in gliomas [213-215]. This gives further credence to our CIS genes as an entire set being highly relevant for cancer formation, and specifically CNS cancer formation.

Altogether, through our transposon-based screen we have identified known and novel

putative cancer genes and pathways driving *EGFR*-mutant gliomas.

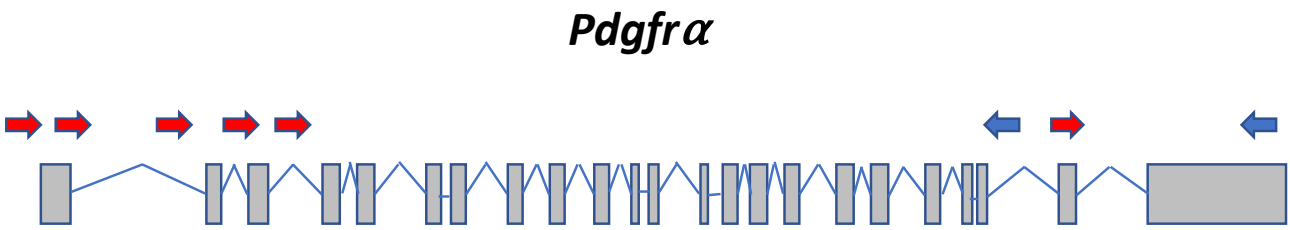


Figure 4.5. Insertional pattern consistent with *Pdgfrα* gene activation in brain tumors. *PiggyBac* transposons from all *EGFRvIII*-PB gliomas are largely at the start of the gene in the forward orientation, with only two at the last exons of the gene (likely to be of lesser functional significance), suggesting the transposons are driving transcriptional activation.

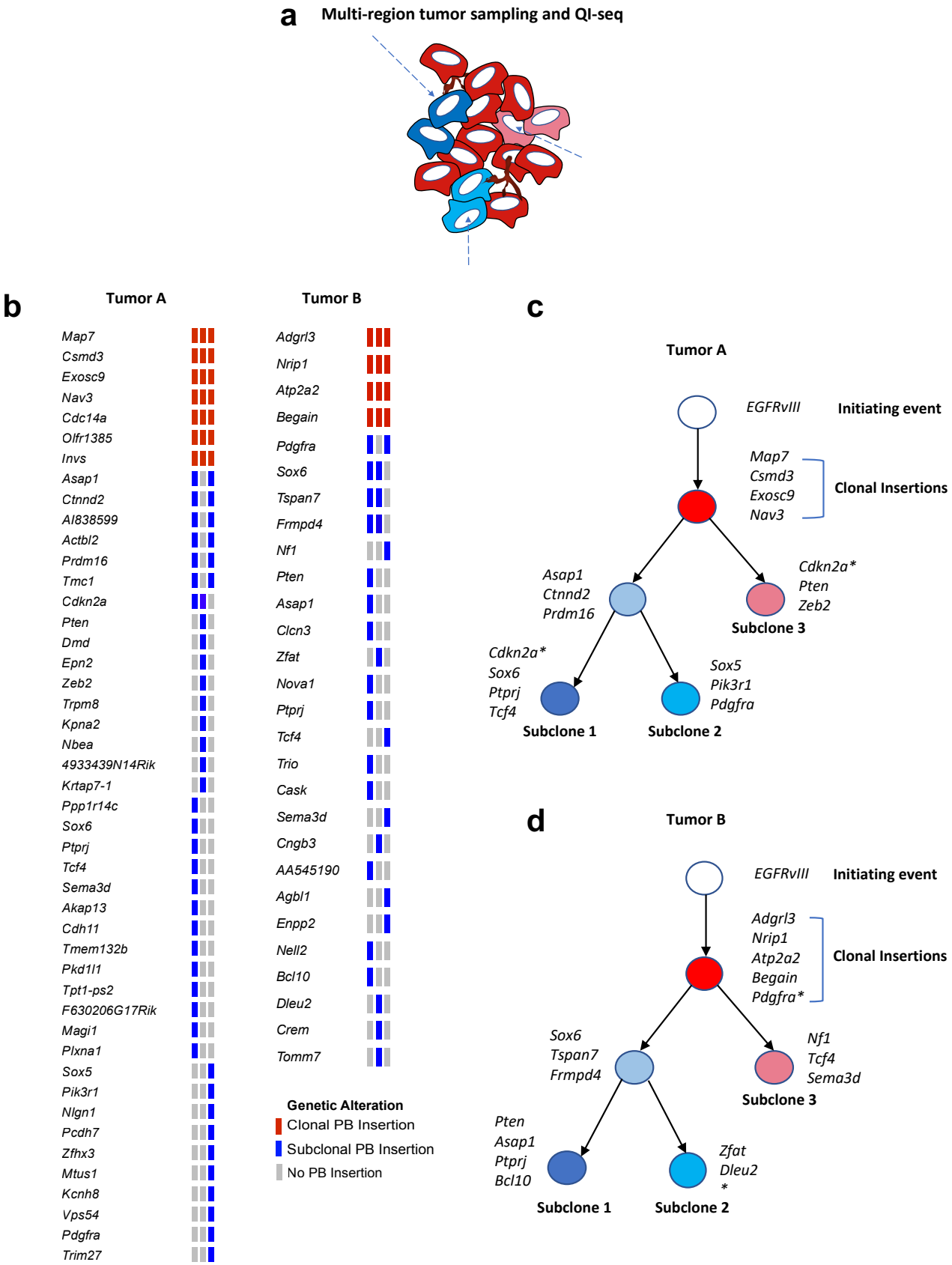

a
Multi-region tumor sampling and QI-seq
b
Tumor A
Map7
Csmd3
Exosc9
Nav3
Cdc14a
Olfr1385
Invs
Asap1
Ctnnd2
AI838599
Actbl2
Prdm16
Tmc1
Cdkn2a
Pten
Dmd
Epn2
Zeb2
Trpm8
Kpna2
Nbea
4933439N14Rik
Krtap7-1
Ppp1r14c
Sox6
Ptprj
Tcf4
Sema3d
Akap13
Cdh11
Tmem132b
Pkd1l1
Tpt1-ps2
F630206G17Rik
Magi1
Plxna1
Sox5
Pik3r1
Nlgn1
Pcdh7
Zfhx3
Mtus1
Kcnh8
Vps54
Pdgfra
Trim27
Tumor B
Adgrl3
Nrip1
Atp2a2
Begain
Pdgfra
Sox6
Tspan7
Frmpd4
Nf1
Pten
Asap1
Clcn3
Zfat
Nova1
Ptprj
Tcf4
Trio
Cask
Sema3d
Cngb3
AA545190
Agbl1
Enpp2
Nell2
Bcl10
Dleu2
Crem
Tomm7
Genetic Alteration
Clonal PB Insertion
Subclonal PB Insertion
No PB Insertion
c
Tumor A
EGFRvIII
Initiating event
Map7
Csmd3
Exosc9
Nav3
Clonal Insertions
Asap1
Ctnnd2
Prdm16
Cdkn2a*
Pten
Zeb2
Subclone 3
Cdkn2a*
Sox6
Ptprj
Tcf4
Sox5
Pik3r1
Pdgfra
Subclone 1
Subclone 2
d
Tumor B
EGFRvIII
Initiating event
Adgrl3
Nrip1
Atp2a2
Begain
Pdgfra*
Clonal Insertions
Sox6
Tspan7
Frmpd4
Nf1
Tcf4
Sema3d
Subclone 3
Pten
Asap1
Ptprj
Bcl10
Zfat
Dleu2
*
Subclone 1
Subclone 2

Fig 4.6. *EGFRvIII*-PB gliomas display intratumor heterogeneity, and PB insertions identify their evolutionary routes. A. Overview of the experiment: two gliomas were sampled from three independent regions each, and their DNA was subjected to QI-seq to determine their insertions. Only insertions in CIS genes (determined to be significant across all 96 tumors) were included in this analysis. B. The insertional patterns from tumor A (a low-grade glioma based on histopathology) and tumor B (a glioblastoma based on histology) from all three regions are displayed on this oncoprint. Clonal PB insertions (found in all regions of the tumor) are coloured red and subclonal ones (found in some regions of the tumor but not all) coloured blue. C. Tumor A shows branching evolution, with truncal clonal insertions in genes including *Map7*, *Csmd3*, *Nav3* and *Exosc9*. *Subclones 1 and 3 have different *Cdkn2a* insertions (ie within different positions in this gene), implying these arose later and independently in evolution. D. Tumor B similarly shows branching evolution, with distinct clonal and subclonal PB insertions. *Subclones 1 and 3 have the same *Pdgfra* insertion (at the same locus), but subclone 2 does not suggesting *Pdgfra* was likely a truncal insertion that subclone 2 later lost due to continued PB transposition.

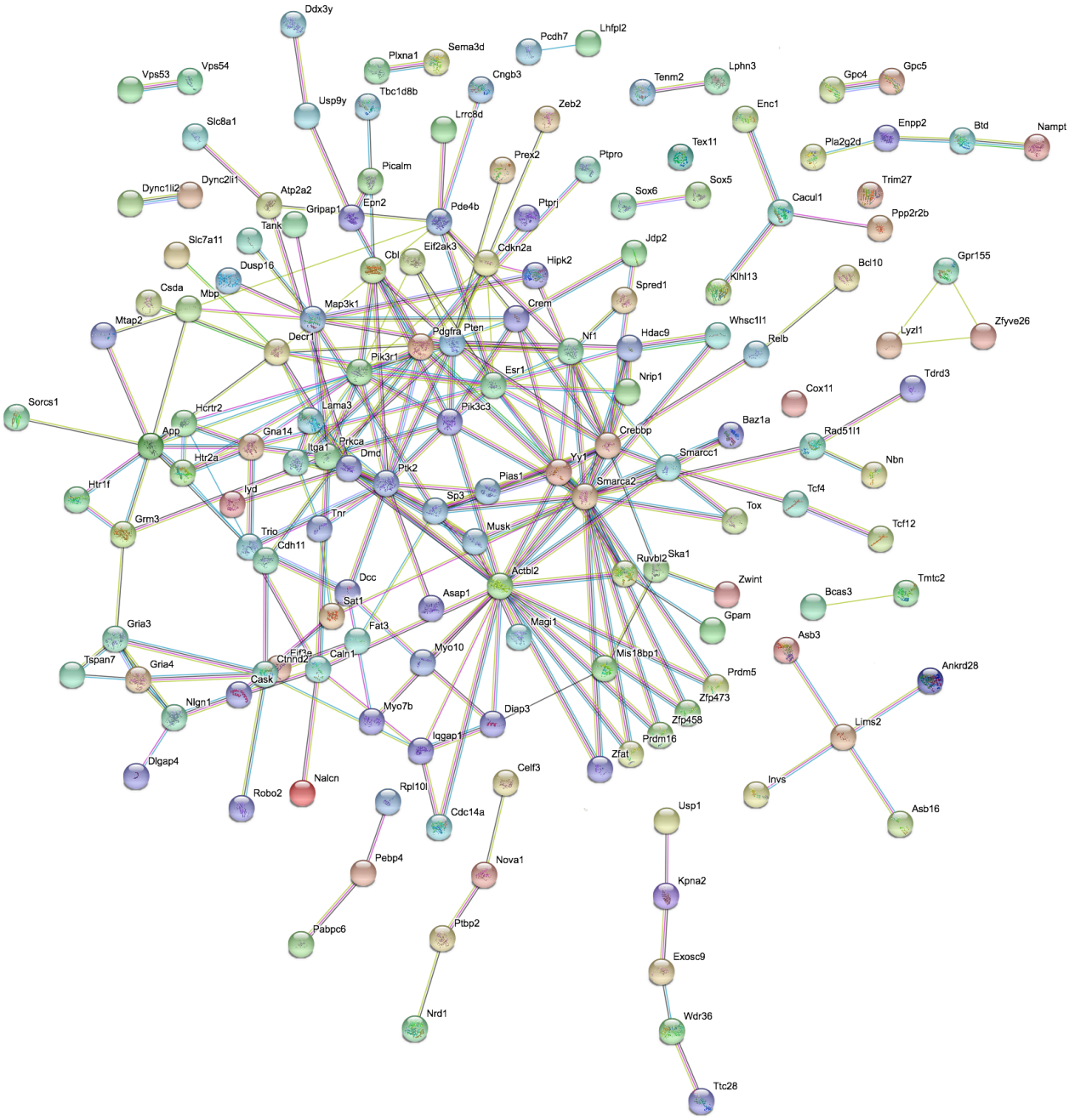


Fig 4.7. Network analysis of all interacting CIS transposon genes. An analysis, performed using STRING, to determine the functional connectivity between CIS genes demonstrates there are 253 interactions between their proteins, showing *PiggyBac* mutagenesis has identified mutations in functionally interacting proteins (Benjamini-Hochberg adjusted p = 4.88 x 10^{-13}, Hypergeometric test). Color coding: colored nodes are proteins from CIS genes; connecting lines are known or predicted interactions between proteins; see https://string-db.org for further details.

	Biological Process (GO)		
Pathway ID	**Pathway Description**	**Count in Gene Set**	**False Discovery Rate**
GO:0032501	multicellular rganismal process	99	2.90E-06
GO:0065007	biological regulation	142	2.90E-06
GO:0044707	single-multicellular organism process	96	3.70E-06
GO:0050789	regulation of biological process	136	3.70E-06
GO:0050794	regulation of cellular process	130	6.72E-06
GO:0044767	single-organism developmental process	90	7.53E-06
GO:0044699	single-organism process	152	8.56E-06
GO:0007275	multicellular organismal development	79	7.56E-05
GO:0048856	anatomical structure development	79	0.000164
GO:0048869	cellular developmental process	67	0.000164
GO:0030154	cell differentiation	64	0.00021
GO:0001953	negative regulation of cell-matrix adhesion	6	0.000251
GO:0048519	negative regulation of biological process	75	0.000251
GO:0048523	negative regulation of cellular process	71	0.000251
GO:0001952	regulation of cell-matrix adhesion	9	0.000265
GO:0009653	anatomical structure morphogenesis	47	0.000545
GO:0044763	single-organism cellular process	135	0.000545
GO:0051171	regulation of nitrogen compound metabolic process	66	0.000594
GO:0048731	system development	68	0.000634
GO:0019219	regulation of nucleobase-containing compound metabolic process	62	0.000973
GO:0050793	regulation of developmental process	46	0.00111
GO:0051252	regulation of RNA metabolic process	58	0.00115
GO:0007399	nervous system development	42	0.00131
GO:0048518	positive regulation of biological process	82	0.00131
GO:0051239	regulation of multicellular organismal process	49	0.00131
GO:0022008	neurogenesis	33	0.00251
GO:0051173	positive regulation of nitrogen compound metabolic process	39	0.00251
GO:0048522	positive regulation of cellular process	73	0.00255
GO:1903506	regulation of nucleic acid-templated transcription	55	0.00289
GO:0010810	regulation of cell-substrate adhesion	10	0.00342
GO:0031324	negative regulation of cellular metabolic process	44	0.00342
GO:0009892	negative regulation of metabolic process	47	0.0036
GO:0010629	negative regulation of gene expression	33	0.0036
GO:0016043	cellular component organization	70	0.0036
GO:0031344	regulation of cell projection organization	17	0.0036
GO:0043549	regulation of kinase activity	21	0.0036
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	37	0.0036
GO:0051270	regulation of cellular component movement	21	0.0036
GO:0031323	regulation of cellular metabolic process	81	0.00395
GO:0006355	regulation of transcription, DNA-templated	54	0.00418
GO:0030334	regulation of cell migration	19	0.00427
GO:0009889	regulation of biosynthetic process	62	0.00432
GO:0051254	positive regulation of RNA metabolic process	34	0.00432
GO:0009891	positive regulation of biosynthetic process	38	0.00474
GO:0021955	central nervous system neuron axonogenesis	5	0.00474
GO:0031326	regulation of cellular biosynthetic process	61	0.00474
GO:0045892	negative regulation of transcription, DNA-templated	28	0.00474
GO:0045893	positive regulation of transcription, DNA-templated	33	0.00474

	Biological Process (GO)		
Pathway ID	**Pathway Description**	**Count in Gene Set**	**False Discovery Rate**
GO:0048015	phosphatidylinositol-mediated signaling	6	0.00474
GO:0071840	cellular component organization or biogenesis	71	0.00474
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	30	0.00474
GO:0051253	negative regulation of RNA metabolic process	29	0.00487
GO:0080090	regulation of primary metabolic process	78	0.00543
GO:0010468	regulation of gene expression	61	0.00555
GO:2000026	regulation of multicellular organismal development	35	0.00555
GO:0021954	central nervous system neuron development	7	0.00559
GO:0031328	positive regulation of cellular biosynthetic process	37	0.00559
GO:2000112	regulation of cellular macromolecule biosynthetic process	57	0.00559
GO:2000739	regulation of mesenchymal stem cell differentiation	3	0.00609
GO:0010556	regulation of macromolecule biosynthetic process	58	0.0062
GO:0040012	regulation of locomotion	20	0.00626
GO:0006357	regulation of transcription from RNA polymerase II promoter	36	0.00631
GO:0048812	neuron projection morphogenesis	14	0.00635
GO:0065008	regulation of biological quality	49	0.00666
GO:0051240	positive regulation of multicellular organismal process	31	0.00667
GO:0044260	cellular macromolecule metabolic process	87	0.00674
GO:0010605	negative regulation of macromolecule metabolic process	42	0.0068
GO:0060255	regulation of macromolecule metabolic process	77	0.00696
GO:0048858	cell projection morphogenesis	18	0.00736
GO:0071391	cellular response to estrogen stimulus	5	0.00741
GO:0031325	positive regulation of cellular metabolic process	50	0.00785
GO:0000902	cell morphogenesis	22	0.00818
GO:0051489	regulation of filopodium assembly	5	0.00818
GO:0032879	regulation of localization	42	0.00823
GO:0051172	negative regulation of nitrogen compound metabolic process	31	0.00958
GO:0010557	positive regulation of macromolecule biosynthetic process	34	0.0108
GO:0019222	regulation of metabolic process	86	0.0124
GO:0023052	signaling	61	0.0124
GO:0010604	positive regulation of macromolecule metabolic process	47	0.0126
GO:0048468	cell development	33	0.0126
GO:0097105	presynaptic membrane assembly	3	0.0126
GO:0032989	cellular component morphogenesis	23	0.0129
GO:0030155	regulation of cell adhesion	17	0.0133
GO:2000171	negative regulation of dendrite development	4	0.0153
GO:0009719	response to endogenous stimulus	27	0.0154
GO:0050803	regulation of synapse structure or activity	10	0.0155
GO:0050896	response to stimulus	85	0.0155
GO:0007154	cell communication	62	0.0156
GO:0061000	negative regulation of dendritic spine development	3	0.016
GO:0009893	positive regulation of metabolic process	56	0.0163
GO:0071392	cellular response to estradiol stimulus	4	0.0169
GO:0044700	single organism signaling	60	0.0178
GO:0048699	generation of neurons	28	0.0178
GO:0051128	regulation of cellular component organization	40	0.0178
GO:0090304	nucleic acid metabolic process	55	0.0178

	Biological Process (GO)		
Pathway ID	**Pathway Description**	**Count in Gene Set**	**False Discovery Rate**
GO:0098609	cell-cell adhesion	16	0.0187
GO:0051345	positive regulation of hydrolase activity	19	0.0188
GO:0051491	positive regulation of filopodium assembly	4	0.0188
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	20	0.0194
GO:0051716	cellular response to stimulus	72	0.0194
GO:0021953	central nervous system neuron differentiation	9	0.0196

Table 4.2. Gene ontology (DAVID) analysis for biological processes enriched in the CIS gene list from all gliomas. Fisher's exact test with FDR correction used for significance testing (FDR < 0.05 deemed significant). This list was generated using publicly available DAVID analysis software, https://david.ncifcrf.gov/.

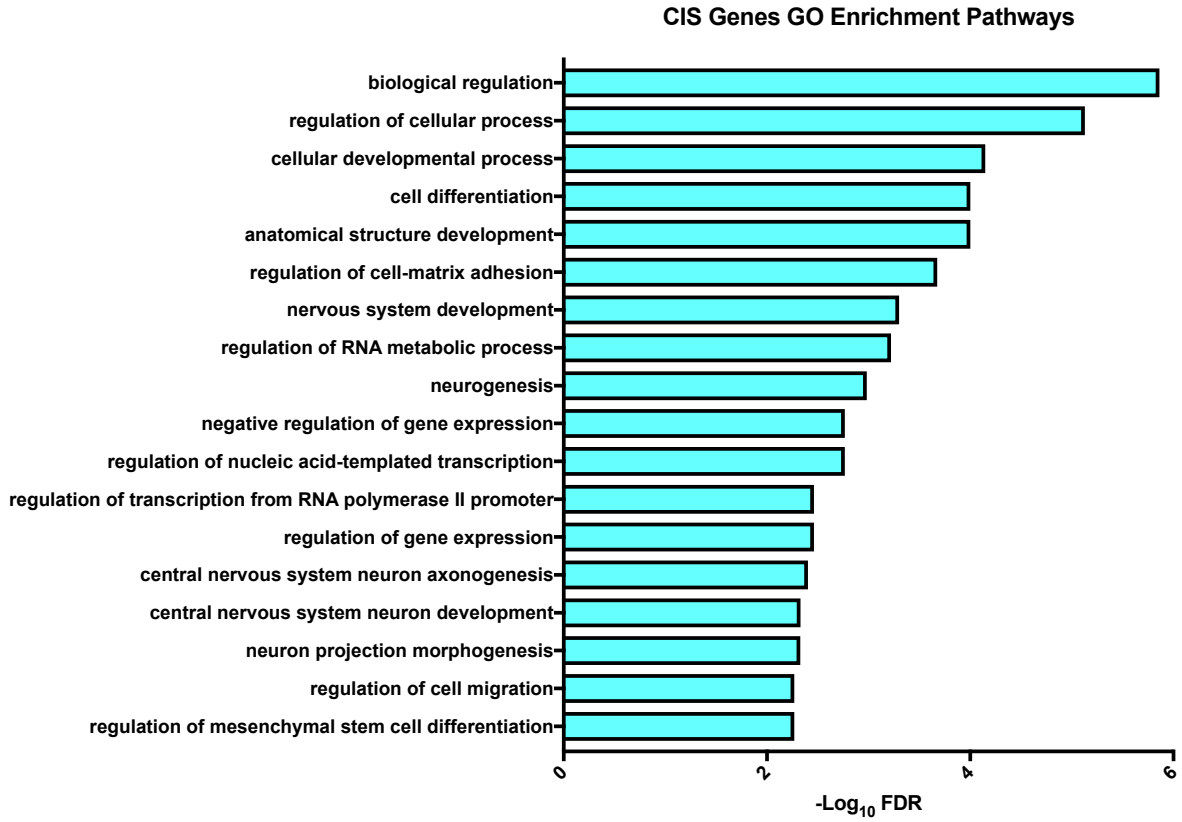


Fig 4.8. DAVID gene ontology (GO) analysis of all 281 glioma CIS genes shows significant enrichment for pathways including neurogenesis and mesenchymal stem cell differentiation, suggesting these pathways are important in driving *EGFR*-mutant gliomagenesis (FDR = false discovery rate). Fisher's exact test with FDR multiple testing correction used as the statistical test.

Comparison of CIS in brain and spinal gliomas

Of the 281 CIS genes, 206 (73%) were shared by both brain and spinal tumors, Fig 4.9. The affected genes include known tumor suppressors underlying multiple types of human gliomas, such as *Cdkn2a*, *Nf1*, and *Pik3r1*, as well as several putative tumor suppressors such as *Sox6*, *Tcf12* and *Spred1*. However, the frequency of insertions in particular shared genes differed between brain and spinal tumors. For example, *Pten* had significantly more insertions in spinal tumors than in brain tumors (22 vs 8 insertions respectively, p= 0.008, Fisher's exact test). Conversely, *Sox6* has significantly more insertions in brain tumors compared with spinal tumors (26 vs 3 insertions, respectively, $p < 0.0001$, Fisher's exact test; Fig 4.10 and Fig 4.11). Other CIS occurred uniquely in each tumor type, for example *Pdgfra* had activating insertions in brain but not spinal tumors (4 and 0 insertions, respectively). Although the CIS genes with lower frequency insertions require further characterization to confirm their tumor-type specificity, collectively these results show there is a shared core set of driver genes for both brain and spinal gliomas.

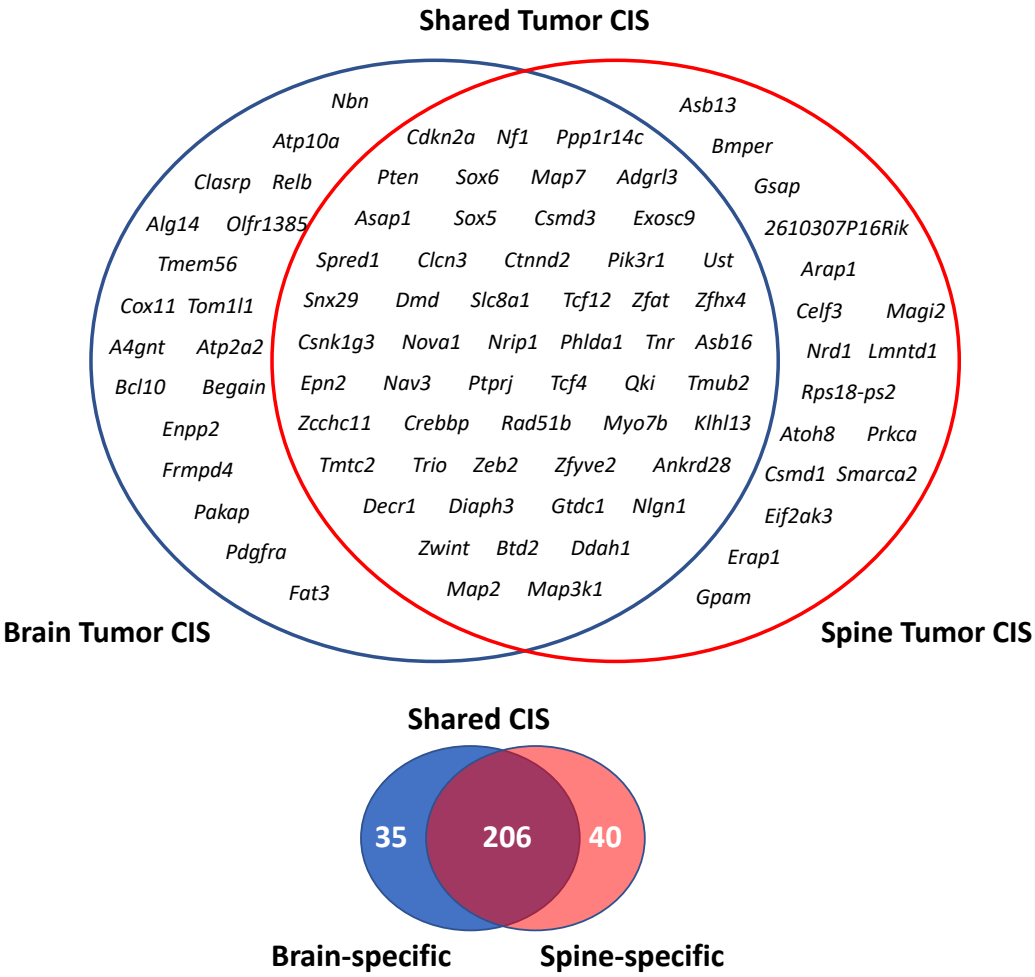


Fig 4.9. Brain and spinal gliomas share a core set of drivers. Upper Venn diagram shows the top genes from each tumor cohort, with core drivers including genes such as *Cdkn2a*, *Pten* and *Sox6.* Lower Venn diagram shows amongst all transposon CIS genes, brain and spinal cord tumors share 206 genes (with at least one insertion in each tumor type), and there are 35 brain glioma-specific CIS genes and 40 spinal glioma-specific CIS genes.

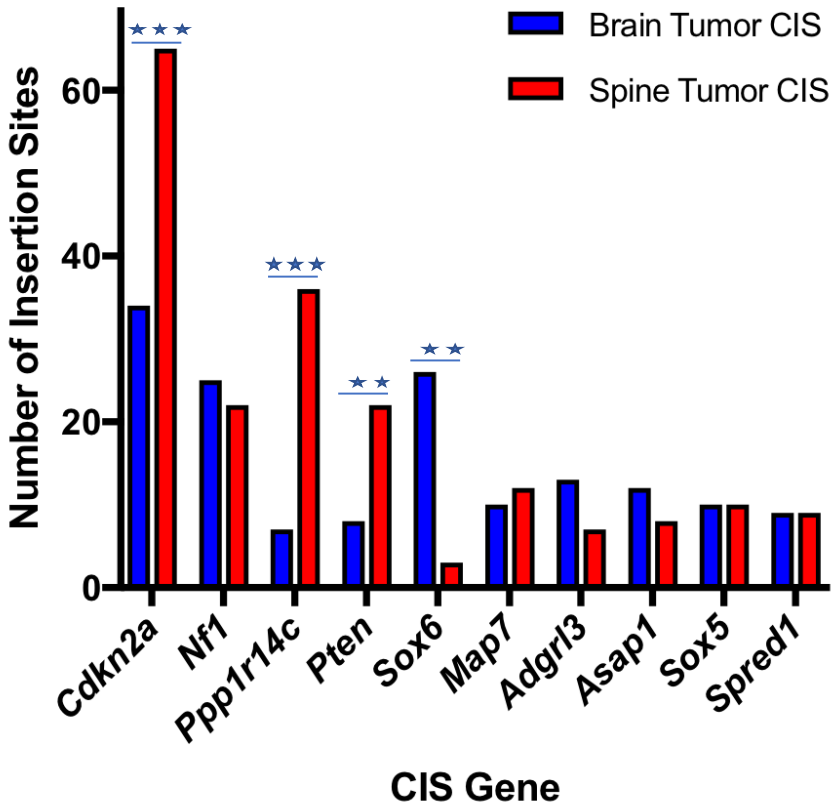


Fig 4.10. Bar plot comparing number of insertions between brain and spinal tumors for the top 10 CIS genes. *Cdkn2a*, *Ppp1r14c* and *Pten* have significantly more insertions (normalized for number of tumors analyzed) in spinal than brain tumors, and *Sox6* has more insertions in brain tumors (Fisher's exact test, $p < 0.05$). ** means $p < 0.01$; *** means $p < 0.001$.

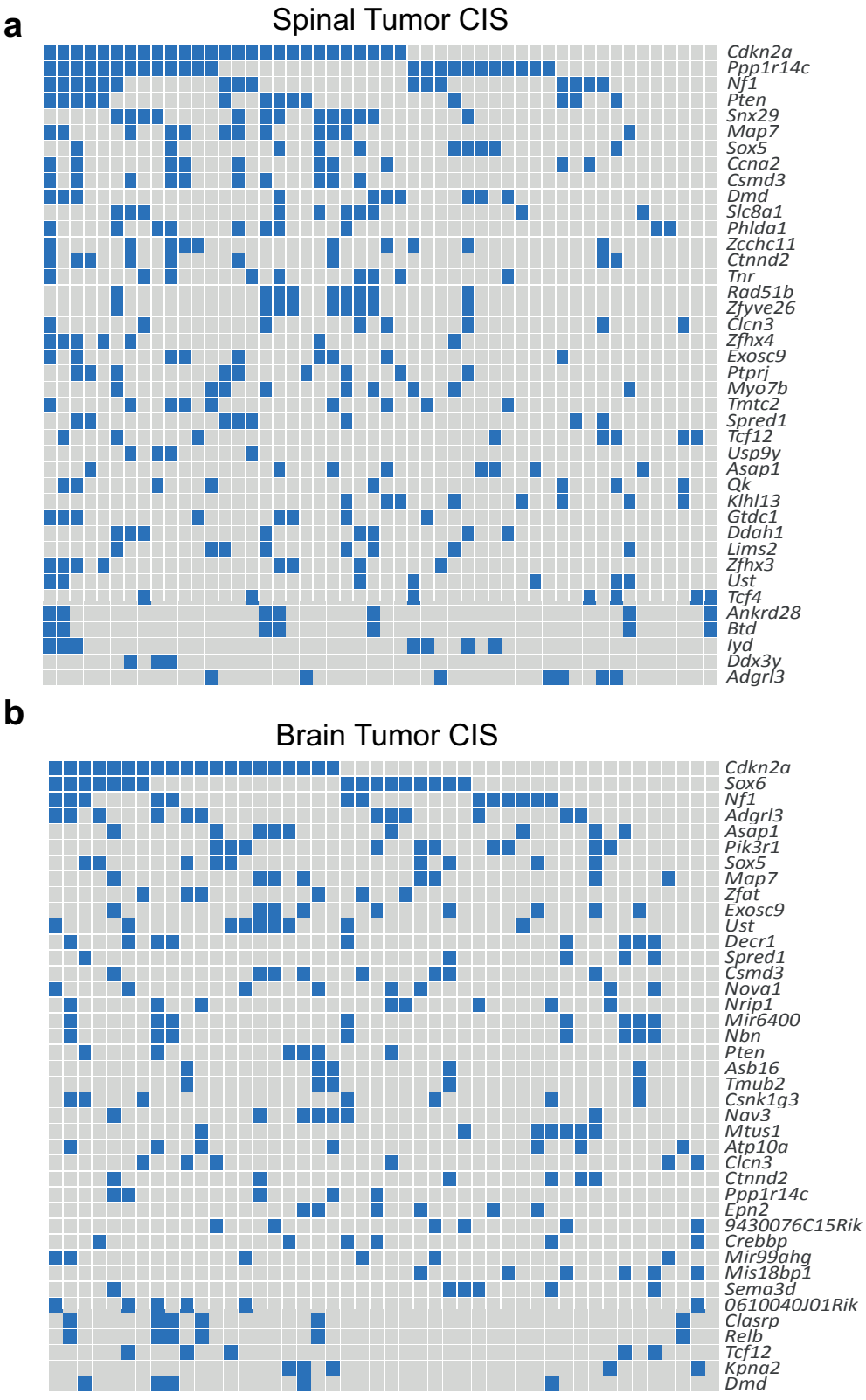

a
Spinal Tumor CIS
Cdkn2a
Ppp1r14c
Nf1
Pten
Snx29
Map7
Sox5
Ccna2
Csmd3
Dmd
Slc8a1
Phlda1
Zcchc11
Ctnnd2
Tnr
Rad51b
Zfyve26
Clcn3
Zfhx4
Exosc9
Ptprj
Myo7b
Tmtc2
Spred1
Tcf12
Usp9y
Asap1
Qk
Klhl13
Gtdc1
Ddah1
Lims2
Zfhx3
Ust
Tcf4
Ankrd28
Btd
Iyd
Ddx3y
Adgrl3
b
Brain Tumor CIS
Cdkn2a
Sox6
Nf1
Adgrl3
Asap1
Pik3r1
Sox5
Map7
Zfat
Exosc9
Ust
Decr1
Spred1
Csmd3
Nova1
Nrip1
Mir6400
Nbn
Pten
Asb16
Tmub2
Csnk1g3
Nav3
Mtus1
Atp10a
Clcn3
Ctnnd2
Ppp1r14c
Epn2
9430076C15Rik
Crebbp
Mir99ahg
Mis18bp1
Sema3d
0610040J01Rik
Clasrp
Relb
Tcf12
Kpna2
Dmd

Fig 4.11. *PiggyBac* mutagenesis identifies *EGFRvIII* cooperative genes in brain and spinal tumors. A. Oncoprint showing the top CIS genes for spinal tumors, ranked according to the total number of insertions. B. Oncoprint for the top CIS genes in brain tumors. Note that *Pten* ranks very highly in spinal tumors but ranks lower in brain tumors (not seen in this oncoprint), where in contrast there are some alternative drivers ranking highly such as *Sox6* and *Pik3r1*.

Correlation with Human Genetic Data

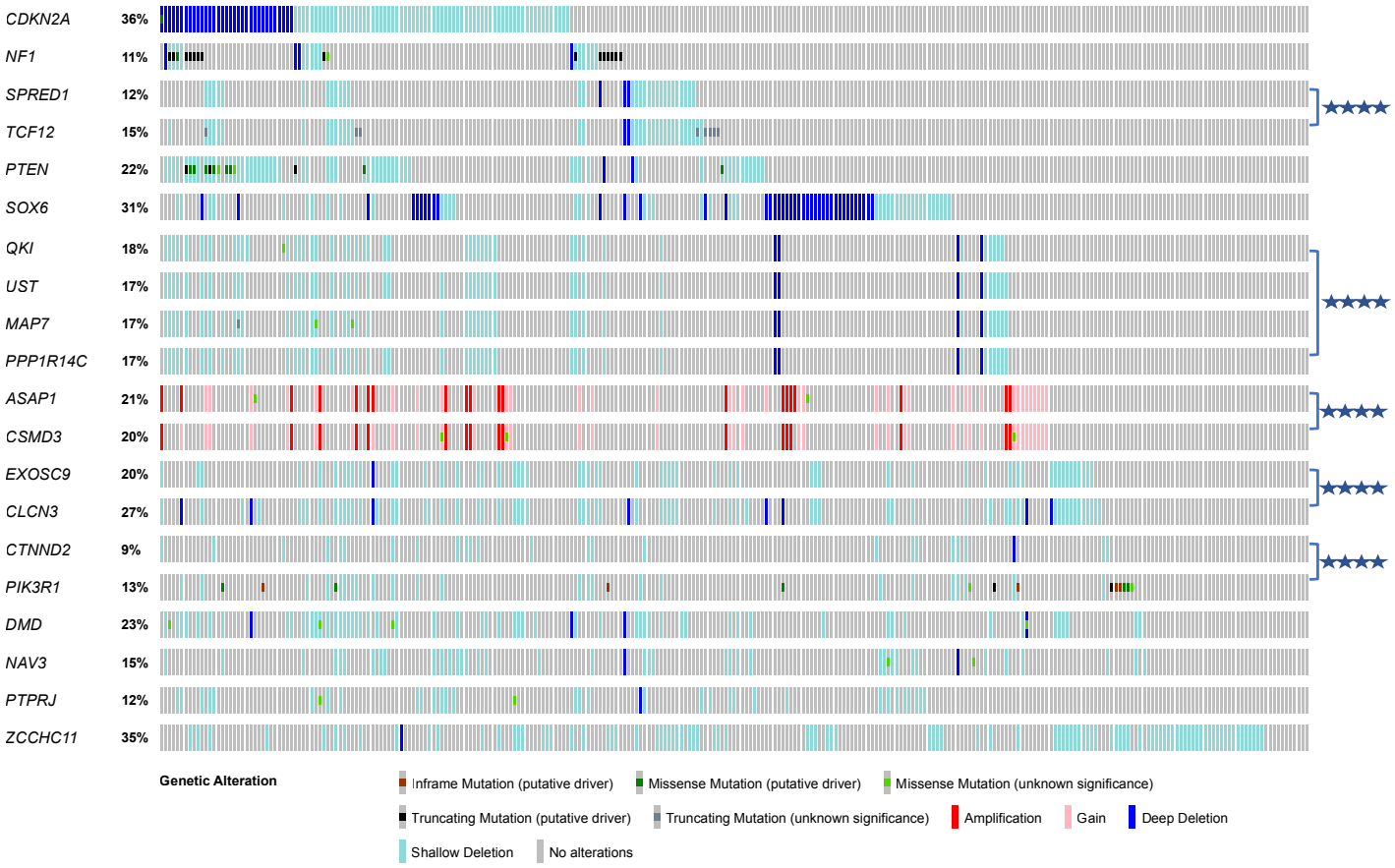


Fig 4.12. Top *PiggyBac* CIS genes are recurrently altered in human low-grade brain gliomas. Patient data was analyzed from The Cancer Genome Atlas (TCGA) dataset (n=283), for cross-comparison of the main CIS genes in mouse brain and spinal tumors. The frequency of alterations of CIS genes observed in patient samples is indicated. Functionally similar genes (*NF1* and *SPRED1*) and co-deleted / co-amplified genes have been grouped together. *TCF12* and *SPRED1* are co-deleted (chromosome 15q), as are *QKI*, *UST*, *PPP1R14C* and *MAP7* (chromosome 6p), as well as *EXOSC9* and *CLCN3* (chromosome 4q). *ASAP1* and *CSMD3* (chromosome 8q) are co-amplified in human tumors. From these 20 top CIS genes, there are 28 gene pairs with significantly co-occurring alterations in human low grade gliomas, many of which are on

neighbouring chromosomal locations; 8 pairs had mutually exclusive alterations (Bonferroni-corrected p-value < 0.05, Fisher's exact test); for simplicity, only the key co-occurring alterations are highlighted here. **** denotes p < 0.0001, Fisher's exact test. These data were analysed using the publicly available software Cbioportal, http://www.cbioportal.org/.

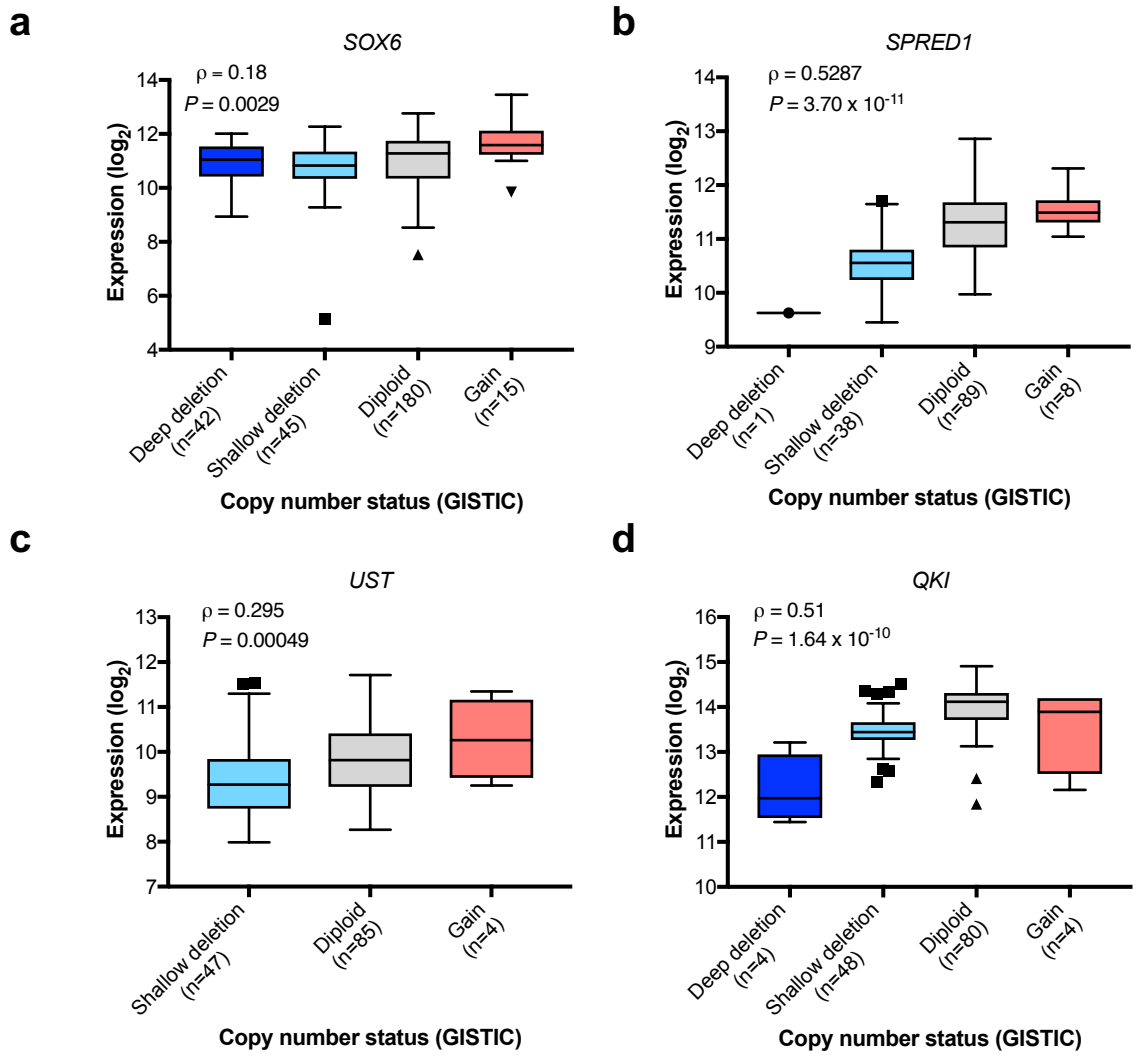


Fig 4.13. Deletions in putative tumor suppressors are associated with reduced gene expression. A – D. Correlation of expression levels of *SOX6* in LGGs (A), *SPRED1* (B), *UST* (C) and *QKI* (D) in human patient GBMs with their respective copy number levels using the entire TCGA human datasets (RNA-seq data available for n=282 LGGs and n=136 GBMs) in order to provide adequate sample sizes. Boxes span the third (Q3) quartile to the first (Q1) quartile (interquartile range, IQR), with the line at the median; whiskers extend to Q3 + 1.5 x IQR and Q1 – 1.5 x IQR. Outliers are plotted as individual points. Spearman's rank correlation was used to calculate correlation coefficients (ρ) and *P* values. The number of patients / tumors (n) is stated for each sub-category.

These correlations suggest deletions of these genes result in loss of their expression, supporting their roles as putative tumor suppressors in this context.

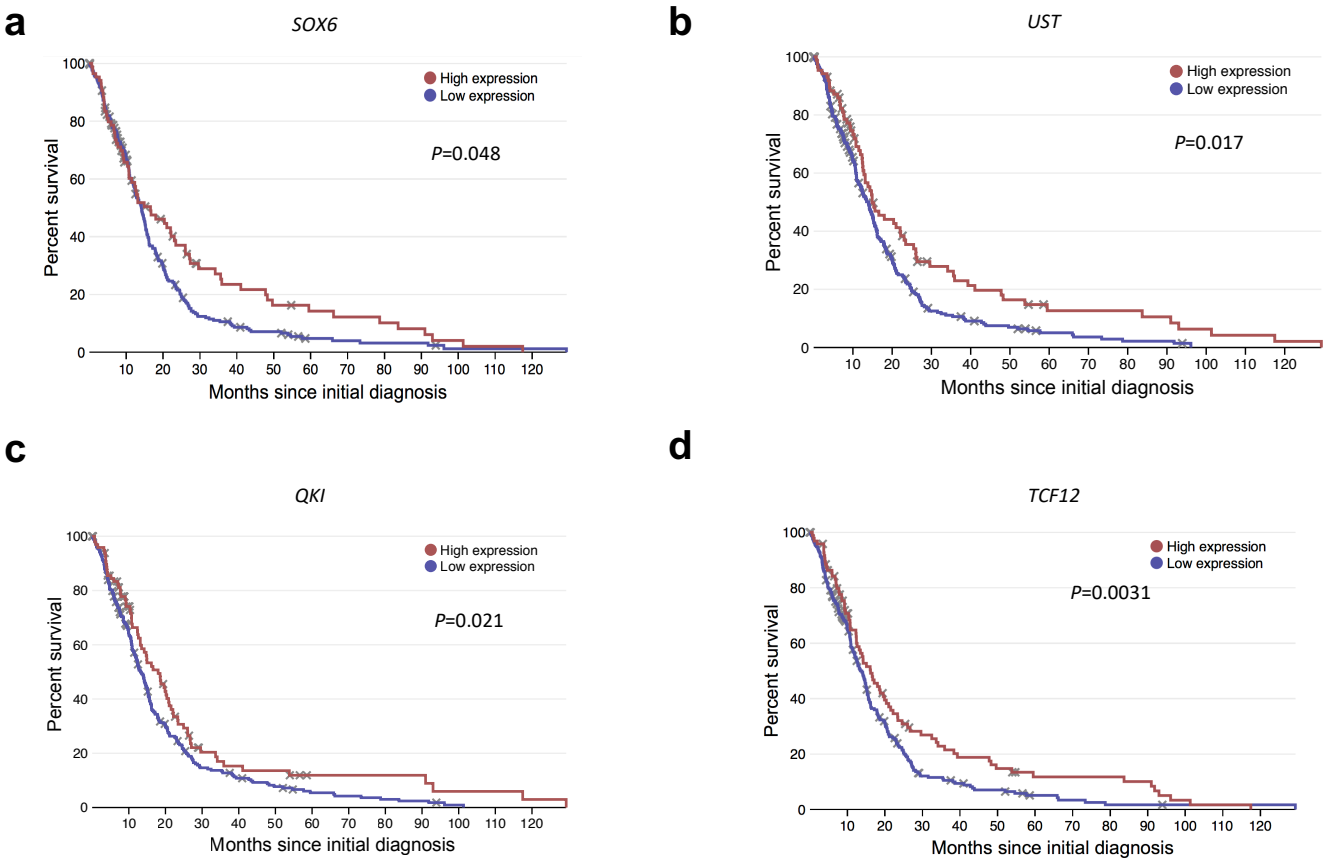


Fig 4.14. A-D. Kaplan-Meier plots of GBM patient survival in relation to expression levels of key CIS genes *SOX6* (a), *UST* (b), *QKI* (c) and *TCF12* (d). *P* values were calculated using the log-rank test comparing the top 30% of expression level with the lower 70% for each gene. The entire TCGA GBM dataset was used (n=273 patients with survival data), to ensure a sufficient sample size with survival data; analyses were performed using the open web interface 'Project Betastasis' (www.betastasis.com).

To assess the clinical relevance of the candidate glioma driver genes, we decided to perform a comparative genomic analysis of our mouse data with data from human patient tumors. To do this, we analysed the frequency with which genetic alterations occur in our top CIS genes in 283 human brain LGGs and 273 GBMs from TCGA datasets[173]. Apart from the known brain glioma tumor suppressors, *CDKN2A*, *NF1* and *PTEN* (all of which of course have previously been established as genetically altered in gliomas), we found *SPRED1* is deleted (heterozygous or homozygous) in 12% of LGGs and 27% of GBMs; and *TCF12* deletions and /or truncating mutations are present in 15% of LGGs and 23% of GBMs. On closer inspection, *SPRED1* and *TCF12* are mostly co-deleted ($p < 0.001$, Fisher's exact test) likely as part of a 15q deletion[216]. *SOX6* is deleted with high frequency: 31% of LGGs and 18% of GBMs, Fig 4.12). Moreover, deletions in these genes associate with correspondingly lower gene expression, Fig 4.13. These data imply deletions of these genes result in loss of their expression, supporting their roles as tumor suppressors.

Other top CIS genes in our dataset, *QKI, UST, PPP1R14C*, and *MAP7,* all map to chromosome 6q and are frequently co-deleted and correspondingly downregulated in human LGGs (Bonferroni-adjusted $p < 0.001$, Fisher's exact test; Fig 4.12). Chromosome 6q is deleted in many human solid cancers including melanomas, and it is interesting that several groups have previously noted the high frequency with which chromosome 6q occurs in gliomas (14% in grade 2 astrocytoma, 38% in anaplastic astrocytomas and 37% in glioblastomas) whereas loss of 6p is limited [216-222]. These observations have led researchers to suggest that many tumor suppressor genes are located in 6q, yet none have been conclusively identified [223]. In our mice all four of these genes had recurrent *piggyBac* insertions across their sequence (implying gene disruption), suggesting these represent multiple new putative tumor suppressors in this region. Similarly, *EXOSC9* and *CLCN3* are co-located on human chromosome 4q and both had disruptive transposon insertions in mice. These data illustrate the utility of *PiggyBac* in pinpointing the cancer drivers hidden within large copy number altered regions, Table 4.3.

To further understand their clinical relevance, we analysed TCGA GBM dataset for correlation of gene expression with patient survival: lower expression of *SOX6*, *UST*, *QKI* and *TCF12* all significantly correlated with shorter patient survival ($p < 0.05$, log-rank test, Fig 4.14).

Gene A	Gene B	Neither	A Not B	B Not A	Both	Log Odds Ratio	Fisher's exact test, p-Value	Bonferroni adjusted p-Value	Tendency
ASAP1	CSMD3	217	4	0	62	>3	<0.001	<0.001	Co-occurrence
UST	PPP1R14C	234	0	0	49	>3	<0.001	<0.001	Co-occurrence
UST	MAP7	233	2	1	47	>3	<0.001	<0.001	Co-occurrence
MAP7	PPP1R14C	233	1	2	47	>3	<0.001	<0.001	Co-occurrence
QKI	UST	231	3	2	47	>3	<0.001	<0.001	Co-occurrence
QKI	PPP1R14C	231	3	2	47	>3	<0.001	<0.001	Co-occurrence
QKI	MAP7	230	5	3	45	>3	<0.001	<0.001	Co-occurrence
EXOSC9	CLCN3	206	1	19	57	>3	<0.001	<0.001	Co-occurrence
SPRED1	TCF12	240	1	10	32	>3	<0.001	<0.001	Co-occurrence
SOX6	PTPRJ	193	55	3	32	>3	<0.001	<0.001	Co-occurrence
SOX5	NAV3	222	19	17	25	2.844	<0.001	<0.001	Co-occurrence
CDKN2A	PTEN	163	57	19	44	1.89	<0.001	<0.001	Co-occurrence
MAP7	RAD51B	197	19	38	29	2.068	<0.001	<0.001	Co-occurrence
UST	RAD51B	196	20	38	29	2.012	<0.001	<0.001	Co-occurrence
PPP1R14C	RAD51B	196	20	38	29	2.012	<0.001	<0.001	Co-occurrence
CDKN2A	QKI	167	66	15	35	1.776	<0.001	<0.001	Co-occurrence
CSMD3	ZCCHC11	128	57	93	5	-2.114	<0.001	<0.001	Mutual exclusivity
ASAP1	ZCCHC11	125	60	92	6	-1.996	<0.001	<0.001	Mutual exclusivity
QKI	RAD51B	193	23	40	27	1.734	<0.001	<0.001	Co-occurrence
CLCN3	DMD	177	42	30	34	1.564	<0.001	<0.001	Co-occurrence
CDKN2A	UST	166	68	16	33	1.616	<0.001	<0.001	Co-occurrence
CDKN2A	PPP1R14C	166	68	16	33	1.616	<0.001	<0.001	Co-occurrence
CTNND2	PIK3R1	233	12	25	13	2.312	<0.001	<0.001	Co-occurrence
CDKN2A	MAP7	166	69	16	32	1.571	<0.001	<0.001	Co-occurrence
SOX6	ZCCHC11	112	73	84	14	-1.364	<0.001	0.001	Mutual exclusivity
NAV3	ZCCHC11	146	39	95	3	-2.135	<0.001	0.002	Mutual exclusivity
EXOSC9	DMD	187	32	38	26	1.386	<0.001	0.004	Co-occurrence
NF1	MAP7	218	17	33	15	1.763	<0.001	0.005	Co-occurrence
PTPRJ	ZCCHC11	152	33	96	2	-2.344	<0.001	0.006	Mutual exclusivity
NF1	UST	217	17	34	15	1.728	<0.001	0.007	Co-occurrence
NF1	PPP1R14C	217	17	34	15	1.728	<0.001	0.007	Co-occurrence
CDKN2A	DMD	155	64	27	37	1.2	<0.001	0.008	Co-occurrence

CSMD3	PTPRJ	204	44	17	18	1.591	<0.001	0.008	Co-occurrence
NF1	QKI	216	17	35	15	1.695	<0.001	0.01	Co-occurrence
NF1	PTEN	205	15	46	17	1.62	<0.001	0.01	Co-occurrence
SOX5	CTNND2	226	32	13	12	1.875	<0.001	0.013	Co-occurrence
CDKN2A	ZCCHC11	104	81	78	20	-1.111	<0.001	0.013	Mutual exclusivity
TCF12	DMD	197	22	44	20	1.404	<0.001	0.02	Co-occurrence
SOX5	PIK3R1	216	29	23	15	1.581	<0.001	0.02	Co-occurrence
ASAP1	PTPRJ	200	48	17	18	1.484	<0.001	0.022	Co-occurrence
SPRED1	DMD	203	16	47	17	1.524	<0.001	0.023	Co-occurrence
NF1	RAD51B	201	15	50	17	1.516	<0.001	0.027	Co-occurrence
SOX5	ZCCHC11	146	39	93	5	-1.603	<0.001	0.04	Mutual exclusivity
PTEN	ZCCHC11	132	53	88	10	-1.262	<0.001	0.046	Mutual exclusivity
CTNND2	NAV3	227	14	31	11	1.75	<0.001	0.048	Co-occurrence
CTNND2	ZCCHC11	161	24	97	1	-2.671	<0.001	0.05	Mutual exclusivity

Table 4.3. Analysis of TCGA low-grade glioma dataset for the top CIS genes shows that many of these genes are recurrently altered in a significant co-occurring or mutually exclusive manner with one another. Fisher's exact test was used to determine significance, with Bonferroni corrected p-value < 0.05 taken as the significance level. The log odds ratio for each gene pair is displayed, reflecting how strongly gene A is associated with the presence or absence of gene B. These data were analysed using the publicly available software Cbioportal, http://www.cbioportal.org/.

Effects of Transposon Insertions on Tumor Transcriptomes

To produce direct evidence of *PiggyBac* insertions affecting their predicted target genes, we used performed paired-end RNA-sequencing of 36 gliomas from *EGFRvIII*-PB mice and implemented IM-Fusion to detect gene-*PiggyBac* fusion transcripts [121], Fig 4.15, Fib 4.16. IM-Fusion is a novel method described to detect RNA-seq reads with sequences from both a transposon and an endogenous gene.

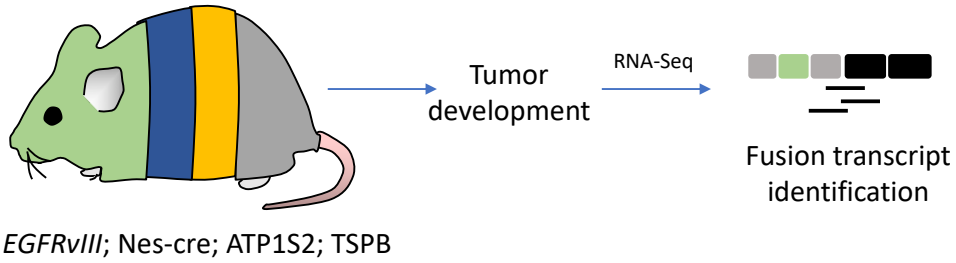


Fig 4.15. Effects of PB insertions on glioma transcriptomes. A. RNA-seq was performed on tumors from *EGFRvIII*-PB mice (n=36), with IM-Fusion[121] analysis of the data to identify fusion transcripts.

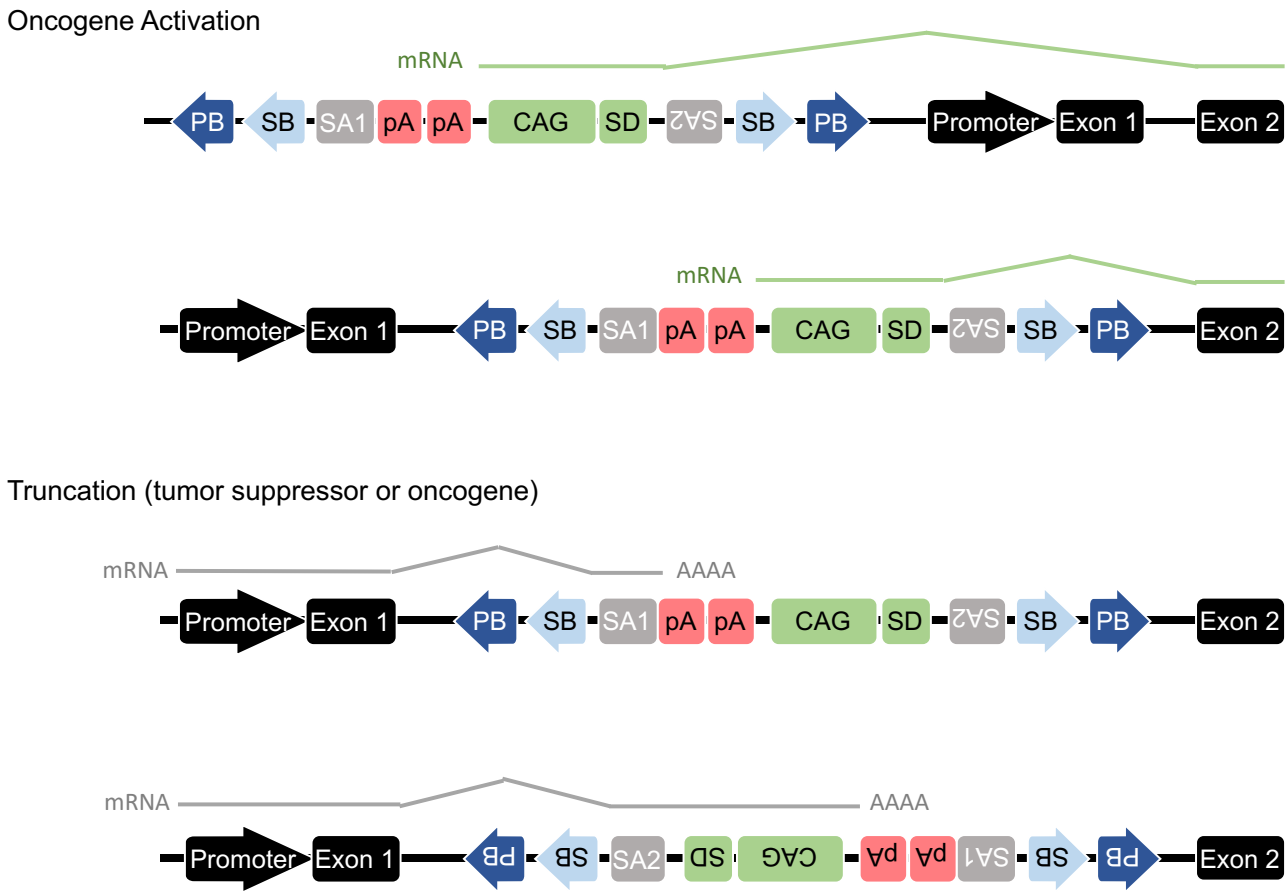


Fig 4.16. Overview of the effect ATP1-S2 transposons on the transcriptome: the transposon can insert in the sense orientation upstream of a gene's promoter or in an early intron, driving gene transcription through the transposon's promoter and splice donor (SD). Alternatively, it can cause transcript termination by inserting in an intron in either sense or antisense orientation because of its two splice acceptors (SA1 = CbASA; SA2 = En2SA) and bi-directional polyA sites; transcript termination can have the effect of inactivating tumor suppressor genes, but also potentially activating an oncogene if there are downstream inhibitory domains for the protein that are removed.

Whereas transcriptomes from *EGFRvIII*-only gliomas had no read counts supporting gene-transposon fusions, *EGFRvIII*-PB gliomas had fusion transcripts for 737 genes in total, of which 80 overlapped with CIS genes detected by DNA-sequencing, Fig 4.17. Moreover, the top CIS genes were more likely to be validated by fusion transcripts: 16 of the top 20 CIS genes had supporting fusion transcripts from at least one tumor, including *Cdkn2a*, *Nf1*, *Pten*, *Sox6*, *Sox5*, *Spred1* and *Tcf12,* Fig 4.18. All fusion transcripts detected with the carp-beta-actin splice acceptor (CbASA) and splice donor (SD) contained *PiggyBac* in the sense orientation, and all those with Engrailed-2 exon-2 splice acceptor (En2SA) contained *PiggyBac* in the antisense orientation, suggesting the transposon was functional in all cases. There were significantly more fusion transcripts containing the first *PiggyBac* splice acceptor (CbASA) than its splice donor (and second splice acceptor, En2SA; $p < 0.0001$ in both cases, t-test). These data imply transcript termination was the predominant effect in the transposon insertional landscape of mutant-*EGFR* gliomas. Of the genes with the most fusion transcript sequencing reads containing PB splice donor (implying activating insertions, see Fig 4.16), *Rad51b* was also a CIS gene (Fig 4.19); its fusion transcripts found in two tumors imply a putative oncogenic role, supporting data demonstrating *RAD51* inhibition radio-sensitizes gliomas by reducing DNA repair[224]. These transcriptomic signatures of *piggyBac* support the functional effects of the identified CIS genes on gliomas. All fusion transcripts are shown in Supplementary Table 7.

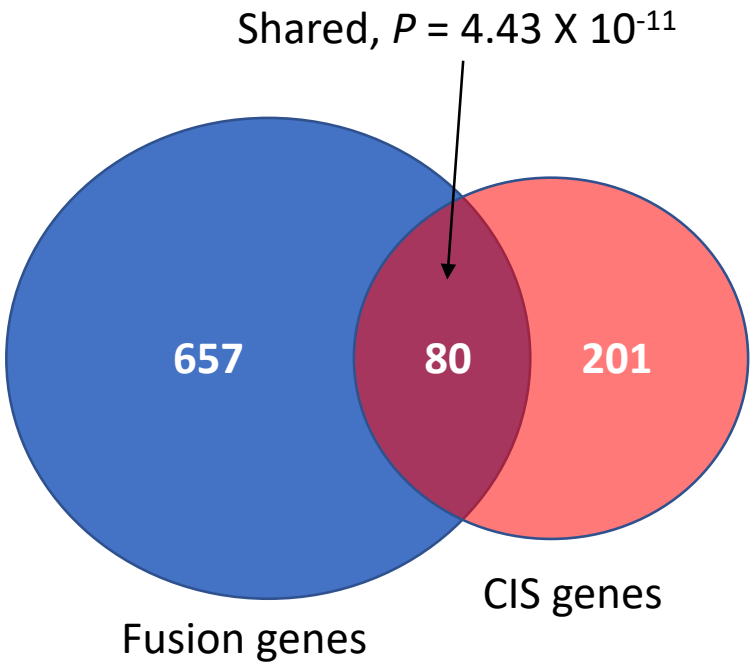


Fig 4.17. Of all genes with fusion transcripts, 80 genes overlapped with CIS genes identified by QI-seq. *P* value was calculated using a two-sided Fisher's exact test. All fusion transcripts detected with the carp-beta-actin splice acceptor (CbASA) and splice donor (SD) contained *PiggyBac* in the sense orientation, and all those with Engrailed-2 exon-2 splice acceptor (En2SA) contained *PiggyBac* in the antisense orientation, suggesting the transposon insertion had functional consequences in all cases.

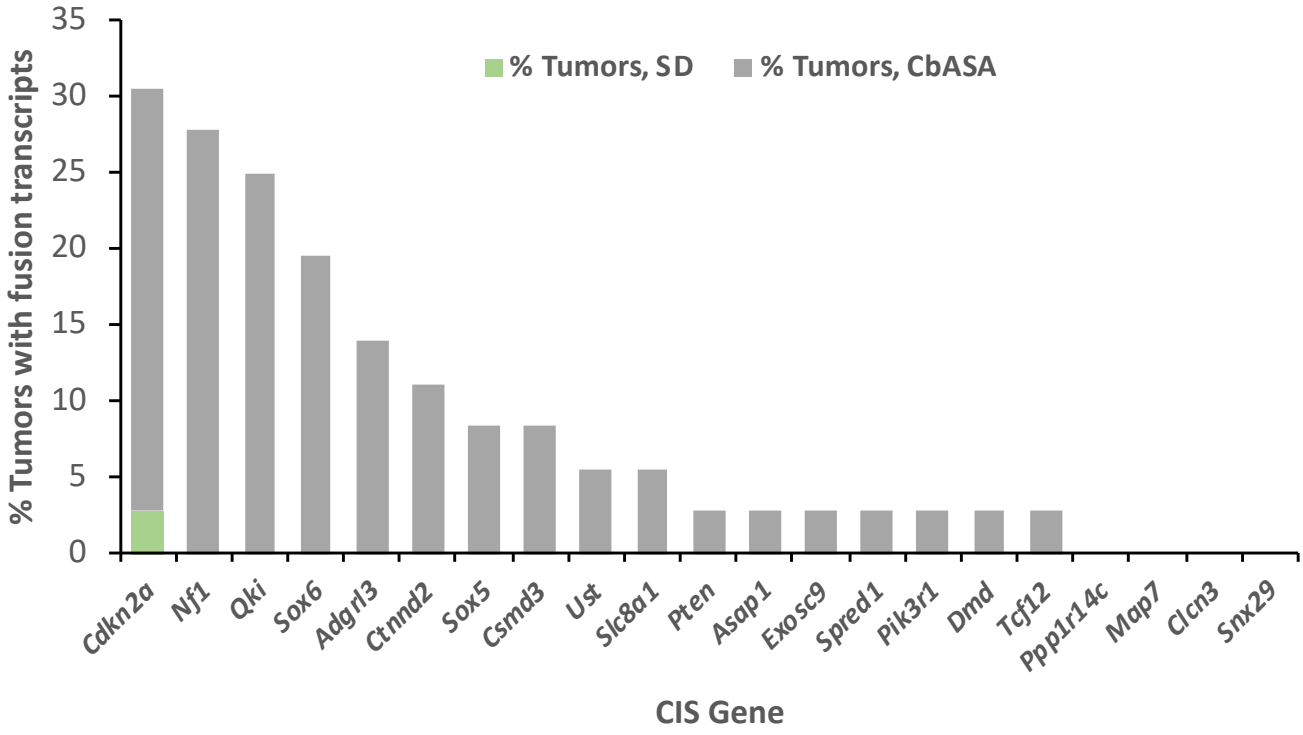


Fig 4.18. Bar plot showing percentage of gliomas with fusion transcripts amongst top 20 CIS genes (*Qki* is also included here).

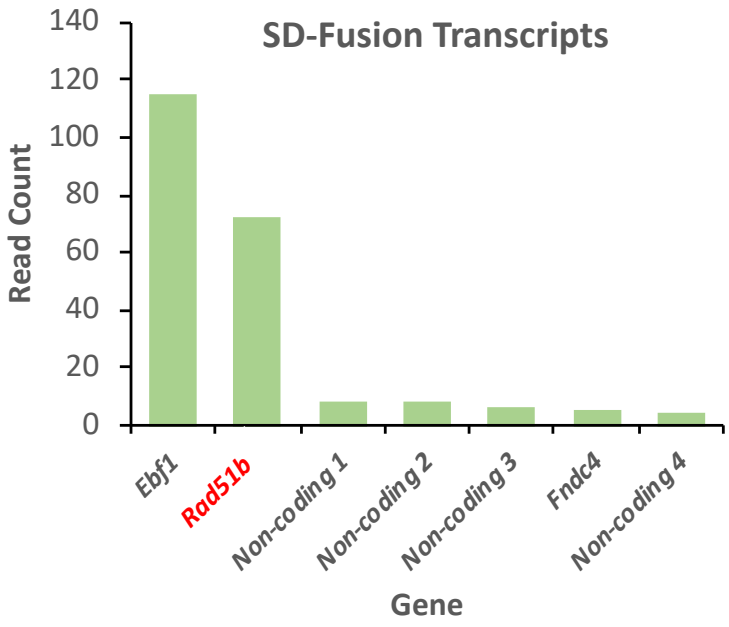


Fig 4.19. Bar plot showing the top fusion transcripts containing the PB splice donor ranked by read count; among them, only *Rad51b* was also identified as a CIS gene.

Determining the effects of *Pten* loss on *EGFRvIII* gliomagenesis in mice

EGFR and *PTEN* in Human GBMs

Mutations in *PTEN* were discovered relatively early in glioma genetics, and since then many studies have reported use of *PTEN* alterations as potential prognostic markers in these patients either alone or in combination with other genes such as *EGFR* [206, 225-228]. Within the TCGA 2016 dataset for GBMs, the somatic mutation rate for *PTEN* is 31.6% (23 missense and 21 truncating mutations), making this gene one of the most commonly altered in this cancer. The majority of these mutations were in the functional protein domains (the dual specificity phosphatase catalytic domain and the C2 domain). Deletions in *PTEN* are also frequent in both LGGs and GBMs, and correlate with reduced PTEN expression [93]. PTEN has been found to be mutated in many cancers, including of the brain, breast and prostate [229]. As such, several groups have attempted to model the effects of *Pten* loss in mice, and it has been consistently reported that this leads to the effect of accelerating tumorigenesis in different backgrounds, such as in combination with *Trp53* and *Pten* loss [94, 230, 231]. PTEN itself is a critical negative regulator of the PI3K pathway; PTEN dephosphorylates the lipid signalling intermediate PIP_3, thus suppressing PI3K and its effects on cell proliferation and growth [232]. *Pten* loss has been found to cooperate with *Egfr* in driving brain tumors in mice, however in the context of predisposing *Cdkn2a* deletions. Given that we have shown here that *EGFRvIII* alone can initiate gliomagenesis, we proceeded to determine if *Pten* loss in this context would accelerate tumorigenesis and particularly whether it would do so in the spinal cord, where *Pten* loss has not previously been shown to drive glioma growth. *Pten* was a CIS in both brain and spinal gliomas, Fig 4.20.

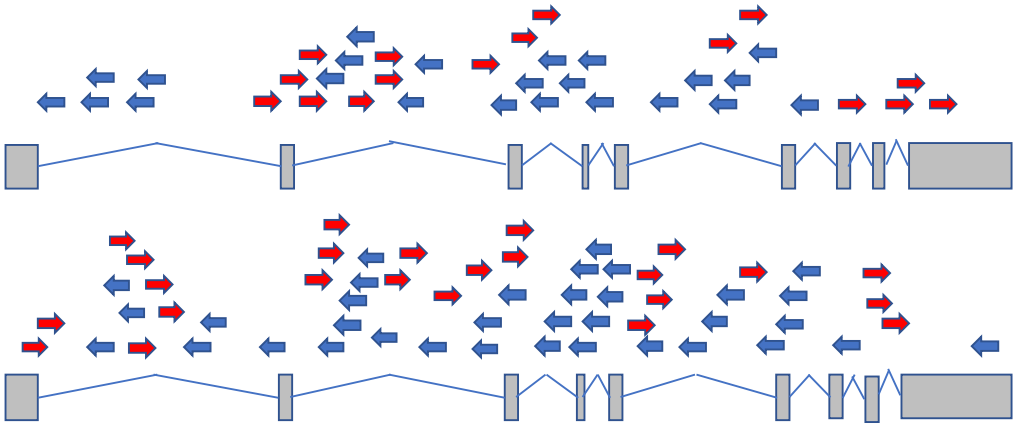


Fig 4.20. All *Pten PiggyBac* insertions from brain gliomas and spinal cord gliomas are plotted across the structure of the gene, with the pattern implying disruption; note the higher density of insertions in this gene in spinal cord tumors.

***EGFRvIII/+ ; Pten*$^{+/-}$ Mice**

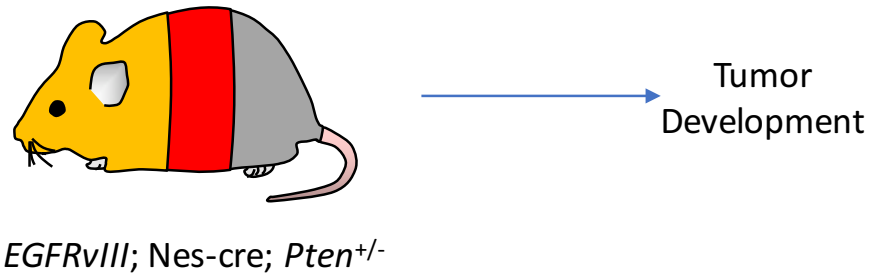


Fig 4.21. Conditional mice with both *EGFRvIII* and *Pten* heterozygous loss (exons 4 and 5 deleted with cre[42]) were generated, and monitored for brain and spinal tumor development.

To explore the role of *Pten* inactivation on brain compared with spinal gliomagenesis, we generated triple transgenic mice carrying the conditional allele of *EGFRvIII*, *nes-cre* and a conditional knockout *Pten* allele [42]*, Pten*^{Loxp}*/+* (n = 11; Fig 4.21). *EGFRvIII*/+ ; *Pten*$^{+/-}$; Nes-cre/+ mice started developing neurological signs from around 8 weeks, including macrocephaly, abnormal gait and limb weakness, which gradually progressed in severity until culling was necessary. There was a predominance of spinal signs (limb weakness and gait anomalies) in this cohort. These mice showed a reduction in survival time compared with mice just carrying the *EGFRvIII* and *nes-cre* alleles (median age 13.0 vs 36.2 weeks, $p < 0.001$, log-rank test; Fig 4.22). Histological examination of *EGFRvIII* ; *nes*-cre; *Pten*$^{+/-}$ mice identified extensive grade II gliomas surrounding the spinal cord at all levels with widespread leptomeningeal and nerve root invasion (from 9/9 mice histologically examined) (Fig 4.23). Of lesser clinical significance, microneoplasias in the SVZ and base of brain were observed.

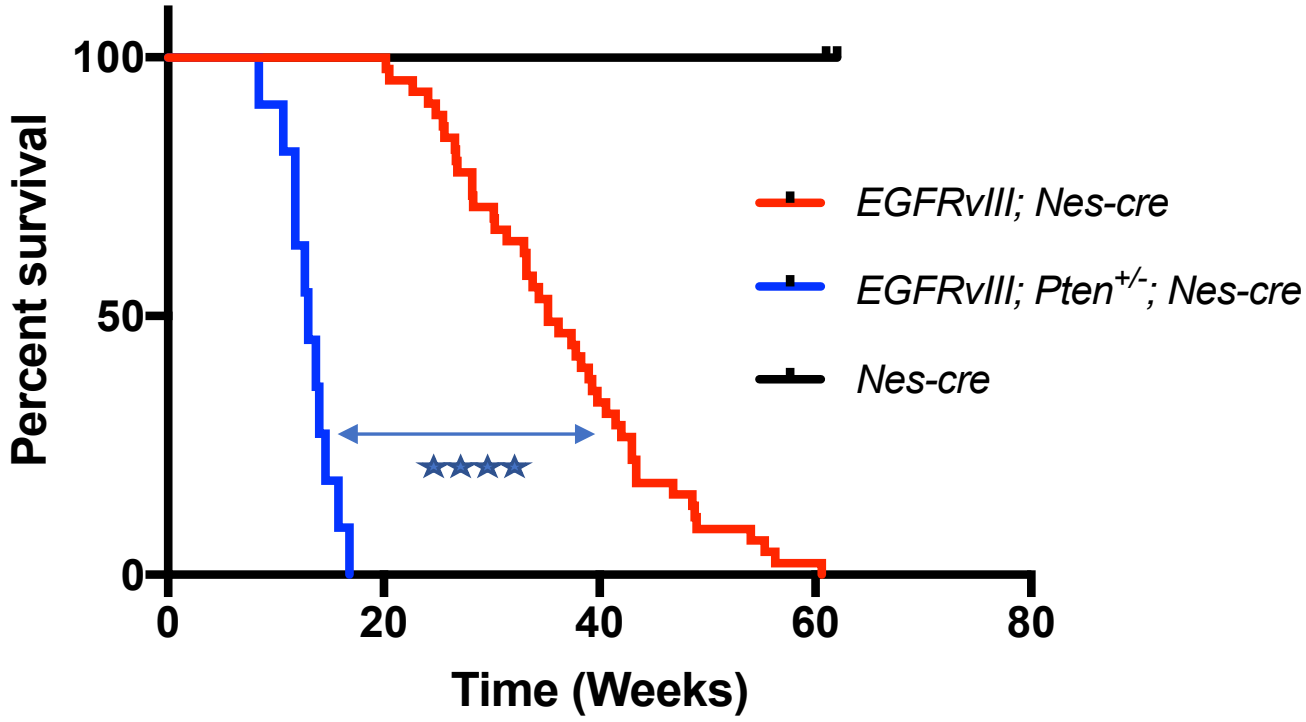


Fig 4.22. Kaplan Meier progression-free survival curves for *EGFRvIII Pten*+/+ mice (red line, n=31) compared with *EGFRvIII Pten*+/- mice (blue line, n=11). There is a significantly shorter progression-free survival when there is loss of one PTEN allele because of signs of spinal cord compression due to spinal tumor development. **** denotes $p < 0.0001$, log-rank test.

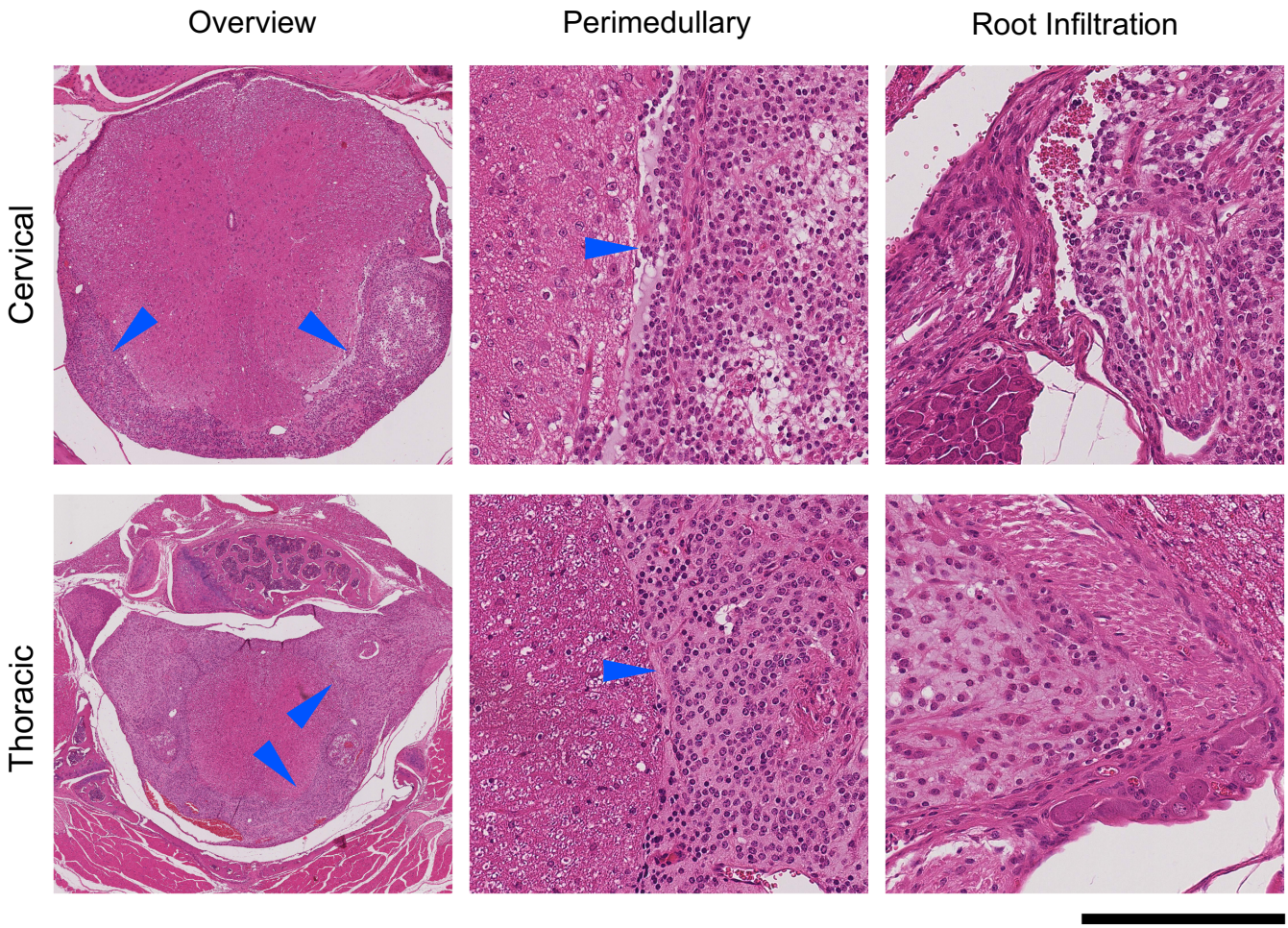


Fig 4.23. *EGFRvIII* ; *nes*-cre; *Pten*$^{+/-}$ spinal tumor growth and nerve root invasion. Left panels show cervical and thoracic spinal cord with encasement by tumor cells growing within the subarachnoid space. Middle panels, detailed view of the spinal cord and tumor cells. Right panels, tumor cells invading root structures. Scale bar corresponds to 0.8 mm for left upper panel and 1.6 mm for left lower panel, and 100 µm for all other panels.

Discussion

Previous studies suggest that constitutive activation of *EGFR* pathways has a strong oncogenic effect that triggers cellular defence mechanisms such as apoptosis or senescence, and therefore that mutations that lead to mitigation of cell death mechanisms are needed for *EGFR* to induce glioma formation. Our work in mice has shown that *EGFRvIII* is capable of initiating gliomas in mice, but given the long latencies for tumor formation it must be that additional genetic events are needed to cooperate with EGFR activation to accelerate tumorigenesis. The top PB transposon CIS we identified was *Cdkn2a* – a commonly deleted tumor suppressor in human gliomas whose alteration frequently co-occurs with *EGFR* amplification [233]. Loss of the protein product of this gene leads to loss of the Rb pathway needed for cell cycle arrest, thus overcoming a critical cell defence mechanism in the face of oncogenic signals to proliferate from constitutive EGFR activation. Among the CIS were other genes related to DNA damage repair mechanisms related to p53 (which can trigger apoptosis), such as *Rad51b* and *Nbn*. This is therefore very much in keeping with earlier studies suggesting that EGFR activation needs further genetic events to disrupt apoptosis or cell cycle arrest pathways, and moreover our forward genetic screen identifies previously unknown genes in these pathways that can cooperate with EGFR in gliomagenesis.

EGFRvIII activation leads to selective constitutive activation of the PI3K-Akt pathway with lesser activation of the Ras pathway, unlike the wild-type *EGFR* which strongly turns on both of these pathways [234]. These two pathways however are thought to cooperate with each other in glioma formation, as shown clearly in a study in mice by Holland et al [29]. It is not surprising therefore that *EGFRvIII* activation benefits from genetic alterations that also switch on the Ras pathway. The second highest ranking CIS in our study was *Nf1* – a known tumor suppressor in many human cancers including gliomas, whose loss triggers over-activation of the Ras pathway. In addition to this, we identified *Spred1* to be another of the highest-ranking CIS; and this gene has only recently been characterised to have a very similar role to *Nf1* as a negative regulator of

Ras but rather intriguingly its germline mutation in the Legius syndrome has not so far been associated with tumors making its role as tumor formation unclear. Our work implicates *Spred1* as a tumor suppressor whose loss can cooperate with *EGFRvIII* in glioma formation, most likely due to Ras over-activation which is known to synergise with PI3K-Akt signalling in tumor formation. As previous studies have found that *Pten* cooperates with *Egfr* in gliomagenesis, our work corroborates this notion as *Pten* was a top CIS in the screen. *Pten* is a protein tyrosine phosphatase that negatively regulates the PI3K-Akt pathway to suppress cell cycle progression and proliferation; *EGFRvIII* primarily signals via the same pathway and therefore loss of *Pten* overcomes an important blockade for constitutive activation of the PI3K-Akt pathway and uncontrolled cell proliferation.

Cancer evolution principles state that there a core set of driver genes, 'truncal' events, that are key to tumor formation particularly in the early stages. At later stages, there is likely to be accumulation of many of more genetic events leading to branching evolution, which also explains the great deal of genetic inter- and intratumor heterogeneity seen in cancers. In our PB forward genetic screen, *EGFRvIII* is the initiating driver event and the PB transposon CIS demonstrates the other core truncal driver events – these are the highest-ranking CIS that have many insertions and are observed in a high proportion of tumor samples. These truncal events include not only the known human core drivers, including *Cdkn2a*, *Pten* and *Nf1*, but also novel genes such as *Spred1*, *Sox6* and *Ppp1r14c*.

In this mouse model of glioma, the tumors can be observed to derive from glioma precursors termed microneoplasias, which express neural stem cell markers suggesting that the ability to maintain stemness in these cells is contributing to their tumorigenic properties. *Tcf12* was one of the highest-ranked CIS genes in our *piggyBac* screen, suggesting this gene may be 'hit' early on in gliomagenesis and supports tumor propagation. The function of this gene is thought to be initiation of neural differentiation [210]; loss of function of this gene through transposon insertions may therefore support tumor formation in these mice by enhancing the ability of early

tumor cells (such as in microneoplasias) to maintain their stemness and avoid differentiation into terminal neurons or glia. This is also consistent with other studies suggesting that gliomas with large proportions of cancer stem cells tend to be more aggressive with shorter patient survival [19]. Based on the disruptive pattern of transposon insertion sites, our data suggest that *Tcf-12* is a likely tumor suppressor in this cancer; although one may expect tumor suppressor genes to be mostly downregulated in cancers, the finding that *Tcf-12* is strongly upregulated from the RNA-seq data may imply that tumor precursors need to upregulate *Tcf-12* to stimulate differentiation and avoid prolonged stemness in order to guard the cells against cancer formation. Loss of function of this gene would therefore overcome this line of defence for the cell and lead to excessive stemness, promoting tumorigenesis. However, further mechanistic studies are required to explore this hypothesis and determine how *Tcf12* supports tumorigenesis.

Continuing with the theme of neurodevelopmental factors, other top CIS amongst all gliomas combined were *Sox6* and *Sox5*: these genes are believed to trigger neuronal differentiation during brain development. *Sox5* and *Sox6* tend to be expressed in a mutually exclusive pattern during brain development, driving differentiation into distinct neuronal subtypes: loss of *Sox6* reduced cortical progenitor differentiation and interneuron diversity suggesting it is critical for these processes in mice, a similar role to *Sox5* for cortical projection neuron development [209, 235]. Given that this group of genes are highly-ranked transposon CIS, it suggests that there is a strong selection for processes that disrupt neural differentiation as they are likely to cooperate with *EGFRvIII* by allowing for uncontrolled cellular proliferation that *EGFRvIII* stimulates. This is also in keeping with our RNA-seq data, which demonstrated enrichment for cell differentiation pathways in *EGFRvIII*-driven gliomas.

Amongst the other top CIS in the combined tumor cohort, *ASAP1* has been linked to increased metastasis in prostate cancer, as there is higher expression of it in metastatic samples compared with primary ones and knockdown of this gene in prostate cancer cell lines reduced invasion *in vitro* [236].

More detailed analysis has resolved some of the genetic heterogeneity within a single tumor. Two tumors were examined which revealed a clonal set of 4 or 5 mutations and distinct sub-clones with non-overlapping mutations. While the biological plausibility of some of the less frequently mutated genes cannot be adequately assessed from this small sample set, this result illustrates the clonal heterogeneity of this disease in mice (correlating with the striking heterogeneity observed in human patients), the need to further understand the underlying genetic architecture in the development and application of improved therapeutic strategies.

Relations to Previous Transposon Screens

There are a few reports of previous transposon screens for gliomas in mice, all of which employed *sleeping beauty* (SB). An early report by Collier et al generated a small number of gliomas from constitutively expressed SB, some of which also had RB-knockout; the CIS from these tumors identified *Csf1* as a putative driver [65]. It is interesting to note that the common human glioma drivers, such as *Egfr*, *Pdgfr*α and *Nf1*, were not identified as CIS in this SB screen; this may be due to inherent insertion biases from SB, the small number of tumors generated, or the genetic background of these tumors. Koso et al (2012) used two rounds of SB-insertional mutagenesis to generate gliomas: neural stem cells carrying the *Trp53*^{R172H} mutation were immortalised *in vitro* from SB insertions and the CIS from these suggested 'immortalising' drivers, and these cells were then transplanted subcutaneously into SCID mice to form tumors with further SB insertions, suggesting tumor initiation drivers [66]. This SB screen from Koso et al yielded some CIS genes in line with human glioma drivers, notably *Nf1*, *Pten* and *Crebbp*. Our CIS list contains genes that are known to cooperate with *EGFR* in human gliomas, including *Cdkn2a*, suggesting that the predisposing mutation is important in transposon screens because the CIS are likely to be specifically cooperating with this predisposition. Moreover, the majority of top CIS in our PB screen differ from the top CIS in the SB screen from Koso et al, suggesting that PB and SB screens can provide complementary information as well as highlighting the importance of using different

predisposing mutations. An advantage of our study is the use of a conditional system to express PB only in the CNS, allowing tumors to develop in the CNS with a competent immune system, since this is much more reflective of the microenvironment in which human gliomas arise and this may impact the genetics and biology of these tumors.

Differences in CIS Genes Between Brain and Spine Gliomas

As described previously, the genetics of spinal astrocytomas is poorly understood, despite this being the commonest intramedullary spinal cord tumor in children and adolescents with significant morbidity and mortality. In one study of pilocytic astrocytomas in the brain and spinal cord, a subgroup analysis of a small cohort of less than 20 cases of midbrain/brainstem/spinal cord astrocytomas was performed, revealing *CDKN2A* had a homozygous deletion in 20% of cases, and loss of heterozyogosity (LOH) was found in 10q23 (containing *PTEN*) in 50% of cases [193]. However, whether *PTEN* is a driver in these spinal astrocytomas is unclear from previous studies, and more work is needed also to establish the true prevalence of these and other genetic aberrations in this disease.

It is interesting to note that the some of the CIS genes of brain and spine gliomas are the same but many differ. A few of the top CIS genes are the same in both groups, including established tumor suppressor genes for example *Cdkn2a* and *Nf1* as well as novel putative drivers such as *Spred1* and *Map7*. The majority of genes differ however, suggesting that there is a core set of the same driver genes for brain and spine gliomas but that these otherwise have tumor-specific cancer genes which arise perhaps later on in tumor evolution. For instance, *Pik3r1* and *Sox6* were frequently mutated (through *piggyBac*) in brain gliomas but not in the corresponding spinal tumors, whereas *Ppp1r14cC* and *Pten* were more frequently mutated in spinal tumors. This is not to say however that *Pten* does not cooperate with *EGFR* for brain tumors; indeed, many studies have suggested that there is this cooperation specifically in these tumors, and our study also

reveals that there are many *Pten* insertions in brain gliomas so that *Pten* is a top CIS in the combined brain and spine tumor cohort. These data imply that *PTEN* may be an even stronger driver in spinal gliomas in this *EGFRvIII*-context than in brain gliomas. To explore the role of *PTEN* more completely, we generated mice with conditional *EGFRvIII* activation and *Pten* loss-of-function mutations. These mice exhibited accelerated development of spinal tumors, confirming a key role of *Pten* in spinal gliomagenesis. These data are consistent with genomic studies in human patients with spinal gliomas reporting loss of heterozygosity at 10q23 (containing *PTEN*) in up to 50% of tumors[193].

Study Limitations

The recurrent nature of integrations in 281 genes across 96 gliomas provides strong statistical support for their selection in gliomagenesis. However, functional validation of individual genes is needed to fully confirm their role as drivers and also to understand how they may promote tumor growth. In order to functionally validate the CIS genes from our *piggyBac* screen as cooperative drivers with *EGFRvIII*, it would be useful to individually disrupt or activate the most promising of these genes in the context of *EGFRvIII* expression. A simple way of achieving this would be through siRNA knockdown of each gene or CRISPR-cas9 mediated knockout of it in glioma cell lines that express *EGFRvIII*. The cell lines would then be subjected to phenotypic analyses to determine if this genetic alteration impacts cancer-related phenotypes, such as cellular proliferation and / or invasion. To confirm these findings *in vivo*, one could cross the *EGFRvIII* ; *nes*-cre mice with a conditional cas9-expressing mouse and subsequently inject single-guide RNAs (sgRNAs, cas9 targeting sequences) into the brain of the *EGFRvIII*/cas9-expressing mice. This would have the effect of knocking-out the gene(s) of interest in the brain of mice that also express *EGFRvIII*, and one could then examine if there is an acceleration of tumor formation and whether there are any phenotypic differences in the resulting brain tumors. A high-throughput way of validating the genes from our screen would be to inject a mini-library of sgRNAs, perhaps targeting the top 40 – 100 CIS genes, and sequencing the tumors to analyse for the most enriched sgRNAs as a way of demonstrating which of the CIS genes are the strongest cooperative drivers. A difficulty with interpreting transposon integrations is determining whether there is gene activation or inactivation when there are only a few (yet still significant) insertions. Therefore, for the CIS genes with fewer integrations, functional validation is particularly important to determine whether these genes are tumor suppressors or oncogenes in gliomas.

Another limitation of the study is that fusion transcripts were not detected for all 281 CIS genes by RNA-sequencing. Potential reasons for this include a smaller sampling size with

transcriptomics (36 tumors were subjected to RNA-sequencing but 96 tumors were subjected to QI-seq), and intratumor heterogeneity leading to detection of some but not all CIS gene fusion transcripts depending on the site of tumor sampling.

Further work should also entail performing RNA-sequencing on the *EGFRvIII* / *Pten*$^{+/-}$ spinal tumors generated in this study to determine if the enriched oncogenic pathways differ in the presence of *Pten* loss. This would provide an indication of the mechanisms by which *Pten* loss accelerates spinal tumorigenesis, which could then be subjected to further dissection.

Conclusions

In this study, we have identified a cohort of 281 CIS genes, including both known and novel putative drivers, that cooperate with *EGFRvIII* in driving glioma progression *in vivo*. Since we employed a conditional mouse model in which all gliomas were shown to be initiated by *EGFRvIII*, the implication is that all other driver genetic events were acquired after this initiating mutation. These genes included those that induce senescence (eg *Cdkn2a*) in the presence of oncogenic signalling, and whose loss enables cell survival and continued proliferation. Amongst the top genes in the list were those that induce neural differentiation during brain development, such as *Sox6* and *Tcf12*; consistent with our data presented previously demonstrating the transcriptome of *EGFRvIII* gliomas is enriched in pathways for cellular differentiation and neurogenesis. The *piggyBac* insertional mutagenesis screen also highlights a key role for known and novel players in the PI3K and MAPK pathways collaborating with mutant-*EGFR* to drive gliomagenesis *in vivo*.

Chapter Five: A *piggyBac* Transposon Screen *In vivo* for Genetic Cooperative Partners of *Trp53* in Gliomagenesis

Abstract

TP53 is amongst the most commonly mutated genes in low grade and high-grade gliomas, suggesting it occurs early in the evolution of these tumors. Large-scale sequencing studies of gliomas from patients have a provided a wealth of information on the genetic and epigenetic landscapes of these tumors. However, there is a need for functional studies to identify the precise roles of the many genes that are altered in these tumors, including which and how genes may cooperate with *TP53* alterations in driving tumorigenesis. Here, we have generated cohorts of mice with a conditional *Trp53* mutation and *piggyBac* transposition expressed under the control of *nestin*-cre in the central nervous system. Preliminary data showed high-grade gliomas were generated in this context, albeit with a long latency. In order to increase the incidence and reduce the time for tumor formation we are generating similar mice with additional loss of *Pten*. Given the complexity of the breeding strategies with the extensive time necessarily required to achieve these, this study is ongoing at the time of PhD submission. I describe the results thus far and discuss future experiments in this Chapter.

Introduction and Aims

TP53 is activated in response to cellular stresses, and it binds to DNA to activate transcriptional programmes which allow for control of the cell cycle and have the potential to activate cell death. This function is important for tumor suppression, and *TP53* was one of the earliest cancer genes to be described, particularly in the context of patients with germline mutations in *TP53*. These Li-Fraumeni syndrome patients are prone to many cancers including those of the breast, lung, brain and oesophagus [237-240]. Given the critical role of p53 protein in sensing DNA damage, it is not surprising that *TP53* itself or its pathway is mutated in the vast majority of cancers, including gliomas. This includes somatic mutations in cancers as diverse as melanomas and colorectal cancer [241, 242]. Analysis of TCGA data on GBMs and LGGs demonstrated *TP53* is commonly genetically altered in these tumors: 33.1% of GBMs and 51.6% of LGGs have an alteration in this gene. Missense mutations of *TP53* are common - all GBM *TP53* mutations were found in the DNA-binding domain (with the exception of one mutation which was in the tetramerisation motif), suggesting these are functional mutations [27, 28], Fig 5.1 and 5.2. Collectively, these observations suggest there is clonal selection for loss-of-function alterations of *TP53* in GBM.

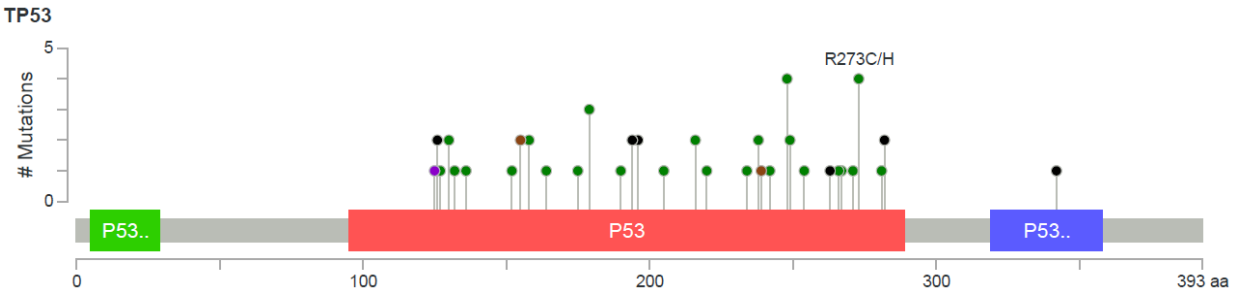


Figure 5.1. This plot demonstrates the number of mutations present in GBMs and where they are located along the amino acid sequence of the p53 protein; each point represents a mutation type and how frequent it is. All mutations are protein coding regions, in particular in the DNA-binding domains of *TP53*, likely abrogating its function as a transcription factor. Green box = *TP53* transactivation domain; red box = DNA binding domain; blue box = *TP53* tetramerisation motif. This plot was generated using CBioportal software, see Materials and Methods.

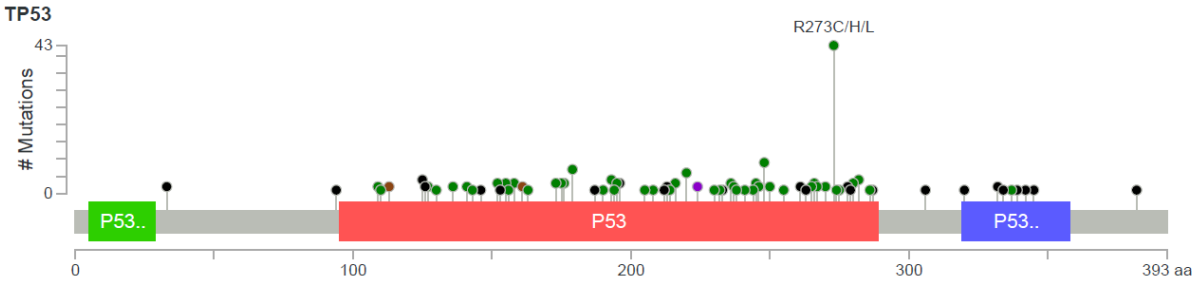


Figure 5.2. Mutations in *TP53* in human LGGs, TCGA dataset. The somatic mutation rate is 51.6%, with 152 missense mutations, 35 truncating, 2 in-frame, and 1 other type of mutation. See previously for descriptions of the protein domain annotations.

The classical view of the role of p53 is that following cellular stress signals such as DNA damage and cellular hyperproliferative signals, MDM2 and MDM4 (negative regulators of p53) are displaced from p53, allowing it to become activated and increase transcription of genes responsible for cell cycle arrest and apoptosis [243, 244]. P53-induced senescence also adds to its tumor suppressive role, for example mice with *Pten*$^{-/-}$ do not develop prostate tumors as quickly as *Pten*$^{-/-}$ *Trp53*$^{-/-}$ mice because p53 induces cellular senescence in the former [245].

Although loss of p53 function itself has tumor promoting effects, there is also evidence that mutant p53 protein has additional gain-of-function oncogene-like properties, for instance disruption of wild-type p53-independent apoptosis [246]. A strong example of this is that genetically-engineered mice which express mutant *Trp53* (*Trp53*^{R172H}) develop a broader spectrum of tumors compared with mice that are heterozygous or null for the wild-type *Trp53* allele: they develop a higher incidence of carcinomas and sarcomas, and have a greater propensity for metastasis and genomic instability, although this type of mutation has been proposed to be a dominant-negative mutation [43, 247, 248]. Oncogenic mutations of *TP53* have been detected in human cancers [249]. Moreover, such oncogenic mutations of *TP53* are also highly expressed [250]. Mutant p53 is thought to have oncogenic properties due to various mechanisms, such as loss of ability to regulate topoisomerase I which normally regulates DNA folding. The G2-M checkpoint is faulty in *TP53*^{R248W} cells, whereas it is normal in *TP53*$^{-/-}$ cells,

further suggesting oncogenic functions of mutant *TP53*. One mechanism proposed to account for these features of p53 mutants is that they interact with the Mre11 nuclease, inhibiting binding of the Mre11-Rad50-NBS1 complex to DNA double strand breaks. This leads to defective Ataxia-telangiectasia mutated (ATM) activation, thus overcoming an important DNA damage response mechanism and promoting carcinogenesis [251]. Over 80% of *TP53* mutations occur in its DNA-binding domain, implying that its role as a transcription factor are critical for its function in tumor suppression [252].

More recently, other roles have been proposed for p53 in addition to its canonical functions. In particular, it is believed that p53 may inhibit invasion and metastasis, enhances communication in the tumor microenvironment, blocks stem cell self-renewal, and inhibits reprogramming of differentiated cells into stem cells [253].

TP53 is one of the most commonly mutated genes in gliomas, found in 20-30% of these tumors. Earlier studies suggested *TP53* is mutated mainly in low-grade gliomas, but more recent work has identified this as a common abnormality in glioblastomas (primary and secondary) too, present in 25-35% of cases [94]. Moreover, *TP53* mutations have been found to occur with *PTEN* loss / mutations in some human GBMs, although mutations in these genes tend to be mutually exclusive in low-grade gliomas, suggesting the combination of mutations in these genes may drive malignant progression [27, 28].

An important study modelling GBMs in mice used GFAP-cre to drive conditional loss of *Trp53* allele in combination with homozygous *Pten* loss; in this model, the cre is expressed throughout the central nervous system, including neural stem cells and mature neurons and astrocytes. Given our success in generating gliomas with *EGFRvIII* expression under nestin-cre control, we decided to use the same approach here by crossing the $Trp53^{R172H}$ allele with nestin-cre. This gives strong expression of the conditional allele in the central nervous system (and eye). A study that also expressed a *Trp53* mutant allele under control of nestin-cre employed the *Sleeping Beauty* transposon as a forward genetic screening approach to identify genetic drivers of glioma, however the tumors generated were from *in vitro* rather than *in vivo* transformation of neural stem cells [66]. This study was discussed in the previous Chapter of this Thesis.

TP53 Mutations in Medulloblastomas

TP53 mutations are also present in medulloblastomas, albeit at a lower frequency than gliomas [254, 255]. Medulloblastoma (MB) mainly occurs in infants and children, representing the commonest paediatric malignant brain tumor and accounting for 20% of all childhood brain cancers. MB typically occurs in the posterior fossa and has a propensity for metastasis through the cerebrospinal fluid to other sites within the CNS (including brain and spinal cord). It is a high-grade tumor with a typically high proliferative rate. The five-year survival rate with treatment (usually a combination of surgical resection, chemotherapy and cranio-spinal irradiation) is approximately 60%. However, the long-term consequences of treatment are significant, and include neuro-cognitive deficits and neuroendocrine dysfunction [256].

Molecular analyses have demonstrated that there are four major subtypes of this cancer: Wnt-driven tumors (with upregulation of canonical Wnt signalling), Sonic hedgehog (SHh) driven tumors, grade 3 and grade 4 medulloblastomas. This consensus was derived from a multitude of studies, one of the most important being an international study in 2006 demonstrating that MBs can be divided into subgroups according to their transcriptomic profiles, and these subgroups showed intra-group similarities in chromosomal aberrations, mutational profiles, tumor histology and prognosis[257, 258]. The best prognosis, with a 5-year overall survival rate of over 90%, is conferred by the Wnt subtype, whereas the poorest prognosis is conferred by the Group 3 tumors which have a 5-year survival rate of only 40-60%. Group 4 and SHH MBs have an intermediate survival rate of 75%.

The driver genes in each MB subgroup (as determined by significant recurrent mutations and / or copy number aberrations) appear distinct, and are reported as follows [259, 260]:

- Wnt subtype – *CTNNB1* (91%), *DDX3X* (50%), *SMARCA4* (26%), *TP53* (14%), and *KMT2D* (12%).
- SHH subtype – *PTCH1* (28%), *TP53* (14%), *KMT2D* (13%), *DDX3X* (12%), *MYCN* amplification (8%), *BCOR* (8%), *LDB1* (7%), *TCF4* (6%), and *GLI2* amplification (5%).

- Group 3 subtype – *MYC* amplification (17%), *PVT1* amplification (12%), *SMARCA4* (11%), *OTX2* amplification (8%), *CTNDEP1* (5%), *LRP1B* (5%), and *KMT2D* (4%).
- Group 4 subtype – *KDM6A* (13%), *SNCAIP* gain (10%), *MYCN* amplification (6%), *KMT2C* (5%), *CDK6* amplification (5%), and *ZMYM3* (4%).

TP53 mutations can occur in any of these subtypes, but in SHH-altered MBs in particular *TP53* mutations may portend a poorer prognosis and treatment failure [261]. A previous *Sleeping Beauty* transposon screen in mice which employed a *Trp53* mutation as a predisposition allele demonstrated that *Sleeping Beauty* transposition both accelerated and increased the incidence of tumor formation. This was also true in a screen in mice which had a *Ptch1* mutation as a predisposition for a *Sleeping Beauty* transposon screen [56]. Importantly, this study demonstrated that there was a different pattern of common insertion sites (putative genetic drivers) in the metastases compared with the primary medulloblastoma. The authors validated some of these findings through whole-exome sequencing of paired primary medulloblastomas and metastases from a small number of human patients.

Aims

In this set of experiments, we set out to determine the driver genes cooperative with *Trp53* that are necessary for glioma tumor formation and progression *in vivo*. In order to do this, we generated mice containing *Trp53*^{R172H} [43] and *piggyBac* transposition alleles (in addition to control cohorts of mice), and collected the resulting tumors from the brain and spine for histopathological analysis and analysis of the transposon insertion sites. We anticipated that the resulting tumors would comprise both gliomas and medulloblastomas, given *TP53* is mutated in both types of tumor in humans and that cre is expressed in the cerebrum and cerebellum under the nestin-promoter. Given the long times needed for tumor formation in this genetic context, the study is ongoing at the time of PhD thesis submission; the main results thus far will be discussed.

Results:

Clinical Phenotypes of Mice

Three cohorts of transgenic mice were generated for this study, as described in Materials and Methods – *Trp53*^{R172H}/+ ; ATP1S2/+ ; TSPB/+ ; nes-cre/+ (*Trp53*^{R172H}-PB mice), *Trp53*^{R172H}/+; ATP1S2/+; nes-cre/+ (*Trp53*^{R172H}-only mice), and ATP1S2/+ ; TSPB/+ ; nes-cre/+ (PB-only mice). The sizes of the cohorts are stated in the Methods. Several mice carrying *Trp53*^{R172H} and *nes*-cre alleles developed abdominal (determined to be hepatosplenomegaly) and lymph node masses, requiring culling (see Tables 5.1 and 5.2). The pathological diagnoses for these individual lesions were not characterised for this study, but are due to loss of one copy of the wild-type *Trp53* allele in tissues leading to oncogenesis, for example lymphomas and sarcomas as previously described [43]. In these cases nevertheless, the brains and spinal cords were still processed for histological assessment of these organs. Some mice carrying *Trp53*^{R172H} and *nes*-cre alleles (particularly in the *Trp53*^{R172H} ; PB cohort, but also one mouse in the *Trp53*^{R172H}-only cohort) however did indeed develop neurological signs, in particular seizures, paralysis of one or more limbs, and macrocephaly. The age at which mice started developing these signs was from 36 weeks onwards. All mice, their clinical signs and histology are outlines in Tables 5.1 and 5.2.

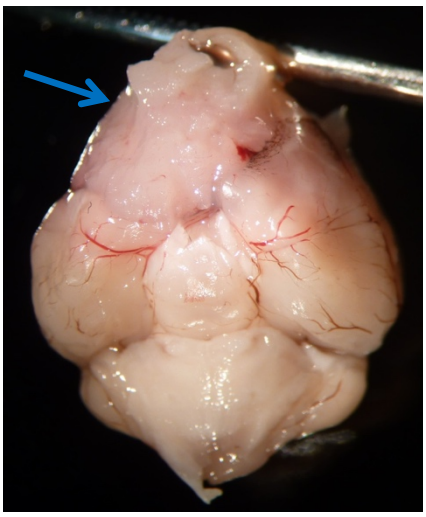


Figure 5.3. Gross inspection of the forebrain of this mouse showed a large tumor, confirmed to be a glioblastoma by histopathological analysis.

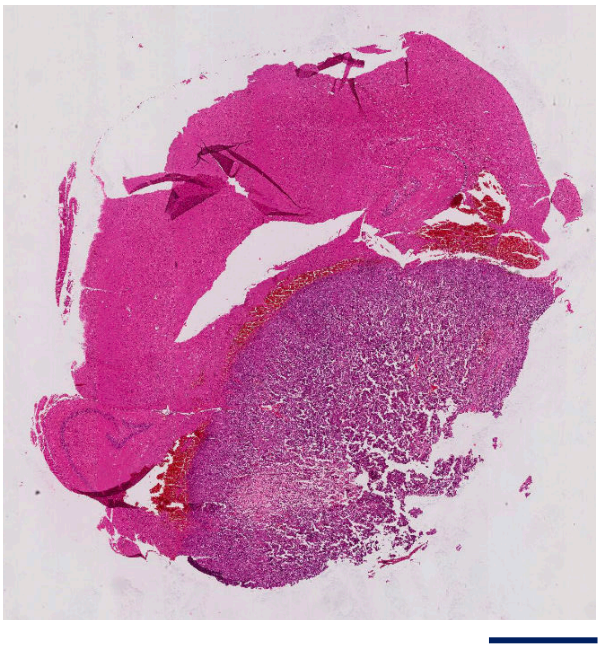


Fig 5.4. Low power image of glioblastoma at base of brain in a *Trp53*^{R172H} – PB mouse. The dark stained area in the lower half of the image represents the tumor. Scale bar represents 1mm.

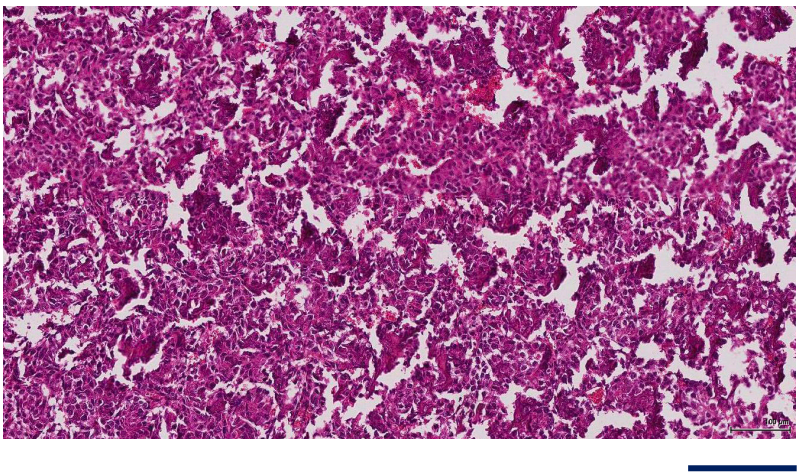


Fig 5.5. Higher magnification image of the glioblastoma from Figure 20. This H&E stained section shows a high cellular density, typical of a high-grade glioma. Scale bar represents 100μm.

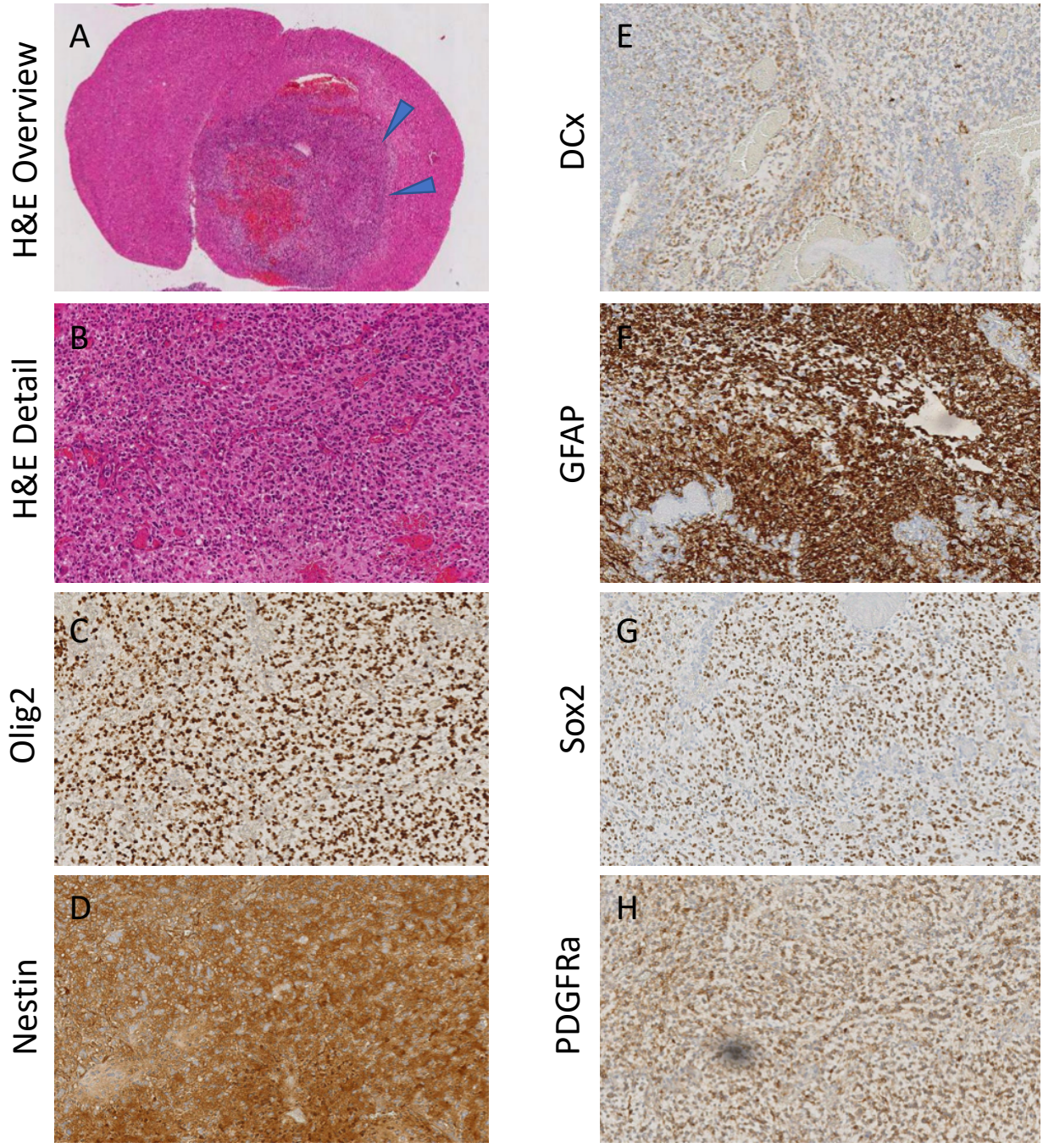


Fig 5.6. Typical example of a glioblastoma from a Trp53^{R172H}-PB with immunohistochemical staining for glioma markers. H&E stains are shown in (a) and (b), Olig2 in (c), nestin in (d), double-cortin in (e), GFAP in (f), Sox2 in (g) and PDGFRa in (h). Scale bar represents 1mm and 100μm. See text for further details, and see Materials and Methods for histological criteria for glioma grading. All histopathology in this Chapter and this Thesis was reviewed by Consultant Neuropathologist, Professor Sebastian Brandner.

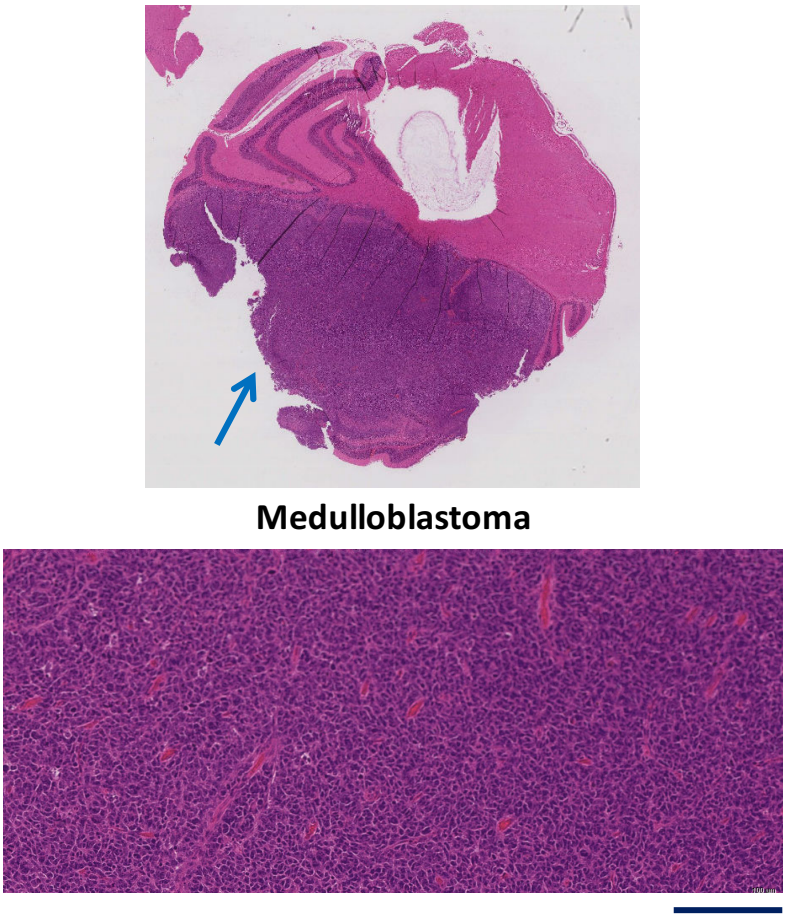


Fig 5.7. Medulloblastoma observed in the cerebellum of a 24.5 week old *Trp53*^{R172H} – PB mouse. Higher power view is seen in the lower panel. This H&E section shows a highly cellular tumor that is poorly differentiated, and the cells contain little cytoplasm; new small blood vessels can be observed in the tumor. Scale bar represents 1.5mm for the upper panel and 100μm for the lower panel.

INPF	Age	Signs	Histology (SB Comments)	Spine Pathology
30.1f	24.5	Moribund	Medulloblastoma	No tumor
18.1A	35.4	Swollen abdomen, hunched	No tumor	No tumor
24.1a	32.5	Sunken abdomen, hunched	No obvious pathology	No tumor
19.1c	40.2	Swollen abdomen, hunched. Necropsy - enlarged liver and spleen.	No tumor	No tumor
19.3f	36.2	Hunched, rapid breathing	No tumor	No tumor
26.1f	39	Swollen abdomen; hunched	Grade 4, giant tumor (glioma)	No tumor
18.1c	46.2	Reduced and uncoordinated movement, piloerection	No tumor	No tumor
18.3g	42.4	Paralysed	No tumor	Malignant soft tissue tumor in spine, not related to spinal cord
18.2d	50.2	Hunched, rapid breathing, piloerection, inactive	No tumor	Epidermoid tumor in spinal cord
18.2a	51.6	Culled for urogenital mass	No tumor	No tumor
29.1b	50	Macrocephaly, hunched, piloerection	Expansive, demarcated high grade glioma, haemorrhage (grade IV)	No tumor
19.2b	54.3	Immobile, piloerection	No tumor	No tumor
30.2f	47.6	Moribund, macrocephaly	Large expansive glioblastoma in forebrain (grade IV)	No tumor
29.3b	44.4	Hyperventilation	Isolated tumor islands at the base of brain, pituitary gland? (grade 1/2)	No tumor
25.1e	52.8	Abdominal mass	No tumor	No tumor
25.1g	52.8	Paralysed	No tumor	Astrocytoma, malignant solid tumor
30.1d	51.7	Abnormal behaviour and posture	No tumor	No tumor
14.1c	63.2	Mass under skin	No tumor	Bone tumor invading spinal cord
18.1e	58.2	Abdominal mass, hunched	Intraparenchymal tumor nest, multifocal (grade IV glioma)	No tumor
19.3a	55.3	Hunched, piloerection	No tumor	No tumor
26.2g	65.8	Paralysis	Small island of malignant primitive tumor, frontal basal	Osteoid tumor in thoracic spine

Table 5.1. Table of culled experimental mice, showing clinical details and pathology from the brain and spine. These mice are all heterozygous for *Trp53*^{R172H}, nestin-cre, ATP1-S2 and TSPB (*Trp53*^{R172H} – PB cohort). INPF is the prefix for this mouse colony. SB comments reflects comments from our neuropathologist, Professor Sebastian Brandner.

Thus far in this study, there were 5 gliomas (including one astrocytoma in the spinal cord), one epidermoid tumor, one bone tumor, a medulloblastoma, and a possible pituitary tumor, as confirmed on histopathological analysis of H&E stained sections by a Consultant Neuropathologist, Fig 5.3, 5.4 and 5.5. The youngest age for a mouse in this cohort with a glioma was 39 weeks, and therefore there is a relatively long latency for onset of these tumors in this genetic context. The mouse with a medulloblastoma was only 24.5 weeks old at death.

To confirm the histological diagnosis of GBM in the relevant samples from *Trp53*^{R172H} ; PB mice, we performed immunohistochemical staining on 3 tumors for a panel of protein markers that are relevant to human gliomas, Fig 5.6. The results were as follows:

- Olig2 – positive cytoplasmic staining in all tumor cells.
- PDGFRa – positive cytoplasmic / membrane (usually diffuse pattern) staining in regions of the tumors.
- Sox2 – positive nuclear staining in all tumor cells (moderately to strongly positive).
- Double-cortin – weakly positive cytoplasmic staining in most areas of the tumors.
- GFAP – cytoplasmic / cell processes are strongly positive in the majority of the tumor.
- Nestin – positive cytoplasmic / cell processes staining, with some tumor-associated vessels also positive.

This staining pattern supports the histological diagnosis of these being glial tumors (see Materials and Methods for the grading system we used).

INPF	Age	Signs	Histology (SB Comments)	Spine Pathology
11.2A	25.5	Uncoordinated (but likely general illness)	Multiple small SVZ clusters; subarachnoid spread of small round cells, possibly lymphoma	Negative
27.1c	35.8	Swollen toe.	Negative	Subdural tumor growth
11.3a	48.4	Leg mass (likely lymph node abnormality)	No tumor	No tumor
20.1f	54.2	Reduced movement	No tumor	No tumor
14.2d	57.6	Uncoordinated movements, hunched, lethargic	Primitive, well demarcated neuroectodermal tumor (grade IV)	No tumor
14.1d	44.7	Mass under skin	Normal	Subdural tumor growth
14.2h	45.2	Swollen abdomen	No tumor	No tumor
14.1e	75	Paralysis	No tumor	Osteoid tumor in lumbar spine
14.2a	49	Seizures	No tumor	No tumor

Table 5.2. Table of culled control mice, providing clinical details and pathology of the brain and spine. All of these mice are heterozygotes for: *Trp53*^{R172H} ; nestin-cre. They lack transposase and thus *piggyBac* transposon is not mobilised.

Relatively few control mice with *Trp53*^{R172H}-only (with no transposition) required culling at the time of submission of this thesis, despite the median age of this cohort being 48 weeks. As can be observed from the table, one mouse brain showed evidence of proliferation of SVZ and subarachnoid space round cells, although these were not tumors. Two mice had small subdural glioma-like growths. One mouse developed a high-grade brain tumor, which was a primitive neuroectodermal tumor (PNET) on histological analysis; this unusual type of tumor was not present in the *Trp53*^{R172H} - PB mice. In no case was a glioblastoma observed, unlike in the *Trp53*^{R172H}-PB cohort. However, meaningful Kaplan-Meier curves of survival times of mice from both cohorts cannot yet be drawn given the majority of mice are still alive. Therefore, additional observation time is required for the remaining mice to determine whether these observations will be reflected across all mice in this study.

Discussion

In this study, we generated cohorts of *Trp53*^{R172H} mice with and without transposition in order to determine the cooperative driver events that support *Trp53* mutations in gliomagenesis *in vivo*. *Trp53* mutant mice with transposition developed tumors with relatively long latency and low incidence (only 8 confirmed intrinsic CNS tumors after aging mice for one year). This observation strongly argues for the need for additional genetic drivers on a *Trp53* mutant background in order to produce brain and spinal tumors. In the *Trp53*^{R172H} – PB cohort of mice, there were 5 high-grade gliomas confirmed after the mice were aged for one year; in comparison, no high-grade gliomas were seen in the *Trp53*^{R172H} – only cohort, with the exception of one primitive neuroectodermal tumor that was considered to be high-grade although this type of tumor is a distinct entity compared with glioblastoma [262]. Although it remains to be seen whether these observations will be seen in the remaining mice of this study, these findings suggest that *piggyBac* transposition, through altering the appropriate cancer genes, is enabling progression of *Trp53* mutant cells towards a glioblastoma phenotype. It will be interesting to determine if there are significant differences in the pathological spectrum of CNS tumors generated in *Trp53*^{R172H} – PB mice compared with *Trp53*^{R172H} – only mice, and if so to determine the set of genes responsible for pushing *Trp53* mutant cells into a GBM phenotype through mapping of the transposon common integration sites. Moreover, it remains to be seen whether this model with transposition leads to a substantial proportion of medulloblastomas (so far, one has been generated), as this will enable comparison of *Trp53*-cooperative driver genes in gliomas with medulloblastomas.

The idea that certain transposon integrations will favour different tumor types is established, in line with the notion of context-dependent cancer genes. For example, the *piggyBac* screen for pancreatic cancer performed by Rad and colleagues helped elucidate the genetic basis of distinct subtypes of pancreatic cancer. They found that hepatoid pancreatic cancers (a rare subtype) in their mice were enriched with *Fidgetin* (*Fign* – a member of the AAA-ATPase superfamily) *piggyBac* integrations, suggesting this gene may help drive this subtype of pancreatic cancer, although further functional proof is awaited.

In order to ensure a sufficiently large number of tumors is generated to have reliable transposon integration sites, we have also started generating mice with *Trp53*^{R172H} and *Pten* loss in combination with transposition (all with cre driven by the nestin promoter). We anticipate these mice will produce gliomas with a higher frequency than *Trp53*^{R172H} and transposition alone, given that *Trp53* and *Pten* have previously been shown to synergise in driving glioma formation in mice [94]. Having mice with this combination of alleles will also help elucidate the influence of different initiating mutations on subsequent genetic evolution of tumors, particularly in comparison with an *EGFRvIII* initiating mutation for which we have data as discussed in previous Chapters.

Although the genetic events that cooperate with *TP53* and *PTEN* loss to drive gliomagenesis are poorly understood, recent work has pointed towards the *QKI* ('quaking') gene as having an important role in this as a putative tumor suppressor. Deletion of *Qki* in mice accelerated glioma formation in a *Trp53* and *Pten* null background, and it was suggested that this was due to enrichment of receptors needed for maintenance of cell self-renewal and therefore stemness, an important aspect of tumorigenesis [263]. Interestingly, *Qki* was a common integration site for *piggyBac* in our *EGFR*-PB screen for gliomas, with a disruptive integration pattern, suggesting deletion of this gene contributes to gliomagenesis in an *EGFRvIII*-driven background as well. It remains to be determined which cooperative partners are specific to having a *Trp53 / Pten* initiating mutation compared with *EGFR*, and which partners are common across all mutational backgrounds. We hope further work from sequencing tumors generated in this study will help provide the answers to these questions.

Study Limitations

The main limitation of this study is the long times for tumor generation, partly associated with the complex breeding strategies involving multiple breeding steps but also because the latency itself for tumor formation in the experimental mice is rather long. As such further work is required to complete this study, in particular the histopathological analysis needs to be extended to all samples, and there needs to be sequencing to map the *piggyBac* integration sites in the tumors available.

Another limitation is that many mice required culling because they developed tumors outside of the central nervous system. This is because of the known effects of the *Trp53*^{R172H} allele, in that all cells in the mouse will only have one functional copy of *Trp53* (and cells in which recombination occurs because of *nestin*-cre will also express the mutant allele). As a result, there is a known predisposition in these mice to many cancers, particularly lymphomas and sarcomas [43]. Therefore, this precluded a substantial fraction of mice developing the intended brain tumors in our model. A possible way of circumventing this problem and potentially increasing the proportion of intrinsic brain tumors would be to use a conditional *Trp53* knockout allele, in which all somatic cells carry both copies of the functional *Trp53* except cells in which recombination takes place (which will carry only one functional copy). This approach may therefore reduce the fraction of mice developing tumors outside of the nervous system when used with the *nestin*-cre driver.

To establish the cells in which recombination of the *Trp53*^{R172H} allele has occurred in our model, a further important step would be to perform immunohistochemical staining for the mutant protein, using an appropriate antibody that recognises mutant forms of the protein but not wild-type. This experiment is necessary to confirm that gliomas (and medulloblastomas) generated in the model used here are indeed, at least partially, driven by this mutant allele.

Conclusions

Given that most of the mice remain under observation at the time of submission of this Thesis, with only a fraction of mice having been sacrificed and histologically assessed thus far, strong conclusions cannot yet be drawn. However, with further experiments as described above that are needed to comprehensively complete this study, we anticipate this work will be complementary to our work *EGFRvIII*-induced gliomas in helping elucidate the genetic evolution of these tumors.

Chapter Six: General Discussion

In this Chapter, I summarise the main findings from this Thesis, discuss potential future directions of transposon mutagenesis screening particularly in the context of gliomas, and also the challenges faced in developing treatments for these tumors.

In this thesis, I have focused on the role of *EGFR*, particularly its activating mutation *EGFRvIII*, in gliomagenesis and how it requires cooperative genetic partners for cancer progression. Note the cell of origin for gliomas was not the focus of my work and so will not be discussed further in this Chapter. The main advance of this body of work is showing that *EGFRvIII* can act as an initiating event in brain tumorigenesis without the prior introduction of tumor suppressor losses. There is a relatively long latency for development of fully-formed gliomas and a low incidence of high-grade tumors, implying a requirement for subsequent additional driving genetic events. Possible reasons why this was not observed in previous studies are discussed in Chapter 3, but include the longer observation times in our study, and the use of the *nestin*-cre driver to for early expression of *EGFRvIII*. We also demonstrated for the first time that *EGFRvIII* can cause spinal glioma formation. Further work should include investigating the cell of origin for these tumors in this mouse model, as previously discussed. Through whole-exome sequencing, we identified these tumors somatically acquired recurrent mutations in *Trp53*, *Tead2* and *Sub1* (all of which have recurrent alterations in human gliomas), as well as deletions in *Cdkn2a* and *Nf1*, and amplification of *EGFRvIII*. RNA-sequencing of these tumors showed aberrant expression of homeobox genes and enrichment of pathways for regulating cell differentiation, as well as known oncogenic pathways including MAPK, p53 signalling and Jnk pathways. We next performed an *in vivo* genome-wide forward genetic screen for *EGFRvIII*-cooperative drivers using a conditional *piggyBac* transposon insertional mutagenesis system. Sequencing and analysis of the *piggyBac* integration sites in 96 gliomas identified a panel of 281 genes which were common integration sites (CIS).

Analysis of the CIS in the PB-cohort provided strong evidence of a number of known and unknown putative genetic drivers collaborating with *EGFRvIII*. Although functional validation of individual genes are needed to definitively support them as drivers, there are multiple lines of evidence which support our conclusion. First the observation of integration sites in the same genes in a significant fraction of tumors provides strong statistical support for selection of these mutations in gliomagenesis. Second, the position of these integrations with respect to the gene body and consequence on expression, consistently disrupting or activating gene expression, such as disruption of *Nf1* and another Ras-inhibitor *Spred1*. Third, data from RNA-seq support the integration pattern because the transposon is designed to promote the expression of the gene, such as those seen with transcripts emanating from the transposon which splice into *Rad51b* or cases where transcripts from the gene splice into the acceptor sites encoded by the transposon thereby disrupting gene expression such as *Cdkn2a*, *Nf1*, *Pten*, *Sox6*, *Sox5*, *Spred1*, *Qki* and *Ust*. Fourth, the overlap of genes identified with mutations / focal deletions by exome sequencing and mutated by *piggyBac* cross-validates their biological selection – including *Cdkn2a*, *Cacul1*, *Esr1*, and *Myo10* (focal deletions); *Nf1*, *Prex2* and *Dgkb* (recurrent mutations); *Cdkn2a*, *Nbn*, *Enc1* and *Spag17* (single mutations). Finally, the correlation with human genetic data is compelling, not only for the previously described genes but also for genes like *SPRED1*, *TCF12* and *SOX6* which are deleted in 27%, 23% and 18% of GBMs, respectively. Interestingly, *piggyBac* identified multiple tumor suppressors co-deleted in large regions in human tumors including *SPRED1* and *TCF12*, and *QKI* and *UST*. The conserved role of these genes in both species validates the similarity and therefore relevance of the mouse model to the human disease.

Comparison of CIS between brain and spinal gliomas revealed that these two types of tumor share many common core drivers such as *Cdkn2a* and *Nf1*, but otherwise each have a some unique putative driver genes acquired later in tumor evolution (although these require further functional validation). We validated *Pten* as a novel cooperative driving event with *EGFRvIII* in spinal tumors, whereas previously this role for *Pten* was only proposed for brain gliomas. The

putative driver genetic events in this work will also provide a comprehensive gene list for further mechanistic work into how genetic alterations support glioma progression.

In order to provide conclusive evidence of these novel genes as drivers in glioma, it would be worthwhile generating conditional knock-out mice in genes of the most convincing CIS. These mice could then be crossed with *EGFRvIII* and *nestin*-cre to demonstrate whether tumorigenesis is accelerated as would be expected; we showed this for *Pten* which accelerated spinal gliomagenesis in particular with *EGFRvIII*. More mechanistic studies could also be done using these models, for example RNA-sequencing analysis may demonstrate different or additional pathways are activated in these tumors compared with those initiated with *EGFRvIII* alone. Producing such mice may be suitable only for relatively small numbers of candidate genes, as it is expensive and time-consuming to produce larger numbers of conditional knock-out mice. An alternative method for potentially validating more candidate genes more efficiently is to use somatic genome engineering with CRISPR-cas9. This method has been applied to producing glioblastomas *in vivo* in mice through targeting known drivers (*Trp53*, *Pten* and *Nf1*) for knock-out as a proof-of-principle; the same study also generated medulloblastomas in mice through CRISPR-mediated somatic disruption of *Ptch1* [264]. Even more recently, Chen and colleagues have used a pooled CRISPR library to screen for driver genes of GBM by stereotaxic injections into the brain of mice; the pool contained sgRNAs for pan-cancer tumor-suppressor genes from TCGA but excluded oncogenes given that this method is for gene disruption rather than activation [265]. The results were able to profile which of the pan-cancer genes are most relevant for GBM. It is conceivable that we could apply this method of CRISPR pooled libraries to validating our list of 281 glioma CIS genes, either as a complete set or for subsets of these genes; given our list also has known and putative oncogenes such as *Pdgfra*, it may be worth considering a separate oncogene screen using activating versions of cas9 that have been shown to have efficacy in conducting functional screens [83, 266].

Spontaneous Mutations versus Transposons in Cancer Gene Discovery

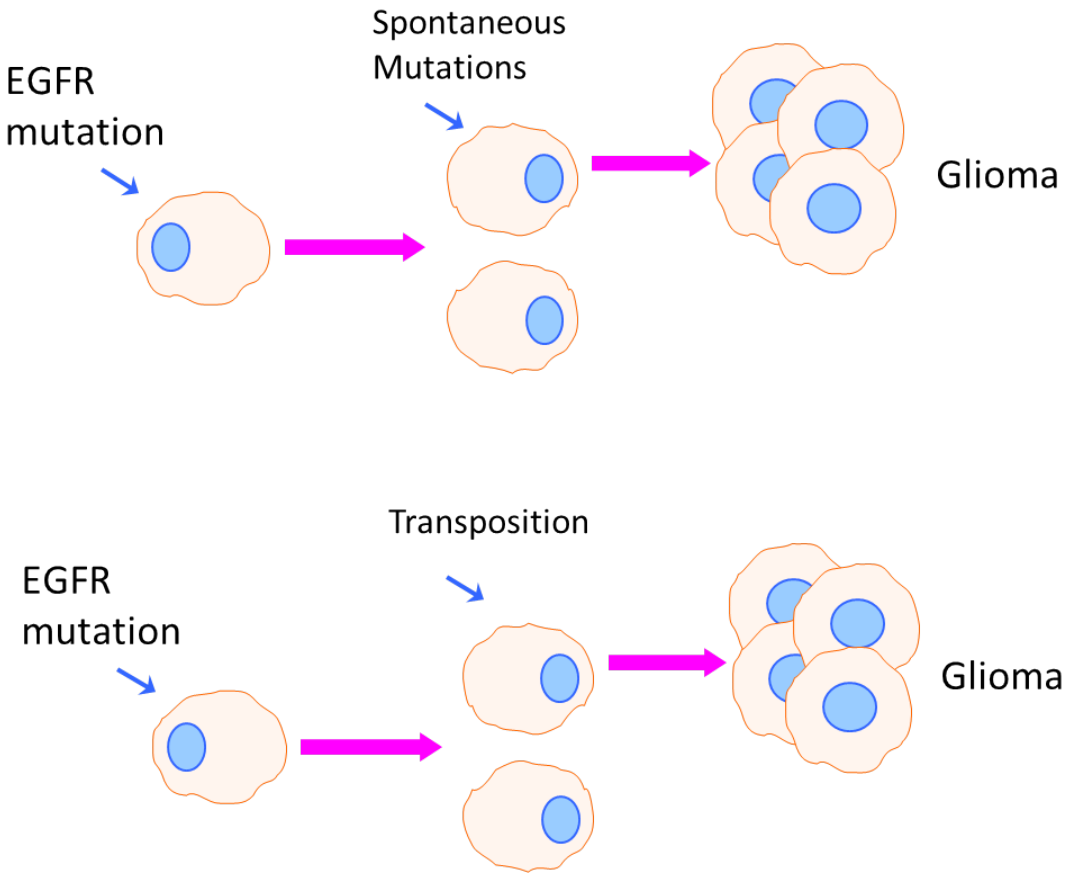


Figure 1. Comparing methods of tumor evolution in model organisms. After acquiring an *EGFR* mutation, tumor precursors will acquire genetic alterations that can be selected for through Darwinian natural selection principles; alternatively, transposition can accelerate these alterations. Tumors were sequenced to identify the genes driving cancer in both cases.

In this PhD, after discovering *EGFRvIII* was sufficient to initiate gliomagenesis in mice, I also used whole-exome sequencing to determine the additional genetic alterations that are acquired during tumor evolution whilst in parallel conducting a forward genetic screen with transposon mutagenesis to determine cooperative genetic drivers with *EGFRvIII*. We demonstrated that in the presence of transposon mutagenesis, there were significantly fewer spontaneous genetic alterations in resulting tumors, likely because transposon insertions were being selected for in

tumorigenesis over spontaneous alterations. These common integration sites revealed the known glioma genes in addition to novel candidate drivers; nevertheless, spontaneous genetic alterations revealed by whole-exome sequencing also occurred in cancer genes. Amongst the top mutated genes in these tumors was *Trp53*, which occurred less frequently in tumors with transposition. This difference in spectrum of mutations between the two tumor cohorts is likely to reflect the increased selection for transposon insertions in cancer genes in the EGFR-PB cohort. Moreover, the other top mutated genes in *EGFRvIII*-only tumors, such as *Tead2* and *Nt5c2* (also frequently aberrated in human gliomas) were not common integration sites from transposition. These findings highlight that whole-exome sequencing from tumors in mice and transposition-induction of tumors are potentially complementary methods of cancer gene discovery, which together are powerful tools for cancer gene discovery.

Although it still remains a major challenge to infer cooperation cancer genes from human genomic studies alone, such as between *EGFR* and other drivers, there are alternative approaches to answer this question compared with our approach here. A recent elegant study by Blakely and colleagues analysed genomes from 1,122 *EGFR*-mutant lung cancers from human patients and found that in addition to *EGFR*, the majority of tumors carried co-occurring genetic alterations in other driver genes such as *CTNNB1*, *PIK3CA*, *RAF*, *MET* and *MYC* [267]. These findings led the authors to conclude that such lung cancers are not single-driver gene entities, but rather have co-occurring driver events. The strength of this work comes from analysing a large number of patient tumors in order to determine significant co-occurring alterations. If a similar study were to be performed for *EGFR*-mutant gliomas, it would potentially provide a strong way of cross-validating our results from this study in patients. However, it must be borne in mind that there are challenges with interpreting human cancer genomes, in particular identifying driver genes from passenger genes, and identifying rare cancer genes amongst large genomic amplifications or deletions. Therefore, studies in mice provide complementary tools for identifying driver genes.

Novel developments in transposon mutagenesis screening

There are a number of recent developments in transposon mutagenesis screening that advance its utility for cancer gene discovery. An elegant study employed a single-copy of the *Sleeping Beauty* transposon per cell as part of a whole-body mouse cancer screen for genes cooperating with *Pten* in driving prostate, skin and breast cancers [67]. This model had several advantages, not least that having only one copy of the transposon per cell reduced the number of passenger mutations, helping prioritisation of the strongest candidate cancer genes for further functional validation. Another advantage was that the transposon was coupled to *Pten* inactivation in the same genome, which may increase the sensitivity of the screen for identifying *Pten*-cooperative cancer genes. Moreover, the transposon lacked a strong promoter for driving endogenous gene expression, and thus was an inactivating-only transposon. Although this meant the screen was not designed for finding putative oncogenes, it greatly simplified downstream analysis and interpretation of common integration sites, which all reflected putative tumor suppressor genes. Further exploration of this model is warranted to confirm reduced passenger mutations induced by the transposon and to compare the sensitivity of this screen for discovering cancer genes compared with models with multiple transposon copies (such as the screen I have presented in this Thesis).

Another recent advance, of particular importance for the brain cancer field, was the use of transposon mutagenesis screening for identifying drivers of medulloblastoma at recurrence after treatment [268]. The investigators used a *Sleeping Beauty* model to produce medulloblastomas in mice; they microsurgically removed these tumors and treated the mice with radiotherapy, reflecting the standard of care in human patients with this disease. As expected, medulloblastomas recurred after this treatment. Genetic sequencing revealed different common integration sites between primary and relapsed medulloblastomas. In keeping with this, genomic sequencing of human primary and relapsed medulloblastomas revealed different mutations. These data suggests distinct genetic drivers are inducing a primary as opposed to a relapsed medulloblastoma. Moreover, they found that the dominant clone of relapsed medulloblastomas arose partly through clonal selection (imposed presumably by surgery and radiotherapy) of a

minor subclone present in the primary tumor. Implications of these findings are that treatments aimed at truncal mutations in the primary tumor are unlikely to provide cures if they are not present in the relapsed tumors, advocating the need for repeated tumor biopsy at tumor recurrence. Future studies using transposon mutagenesis to identify the molecular players driving tumor recurrence in other contexts, such as different treatments and different cancers, are warranted.

Given that *piggyBac*, like other transposons, continues to mobilize around the genome in the presence of transposase, it is also useful as a system for determining resistance mechanisms to chemotherapeutic agents. A recent study demonstrated this for *Trp53-Mdm2* resistance mechanisms in an $Arf^{-/-}$ model, in which *piggyBac* common insertions were found in *Trp53* and *Bcl-xl*, the latter of which were activating insertions [269].

Despite the wealth of useful data provided *in vivo* transposon-based cancer screens in mice, these studies are typically expensive, time-consuming and resource-heavy, given that multiple mouse crosses are required demanding relatively productions of relatively large numbers of mice. For these reasons, there is increased demand for reliable, *in vitro* transposon mutagenesis models for performing cancer screens. Useful advances on this front have been made recently. For example, Fan et al have reported a *piggyBac* screen with an *EGFR*-mutant lung cancer cell line in the presence of an EGFR inhibitor; sequencing and analysis of the transposon integration sites in this cell line identified *MET* activation (known to drive resistance to EGFR inhibitors) as well as a novel player, *YES1* (a Src family kinase) [270]. The investigators then processed human clinical datasets of lung cancer patients treated with EGFR inhibitors, and identified the presence of *YES1* amplification in a subset of these patients. Treatment of an EGFR-mutant lung cancer cell line containing activating *YES1* insertions with a YES1 inhibitor or *YES1* siRNA knockdown sensitised the cells to treatment with EGFR inhibitors, supporting the role of YES1 in driving drug resistance. This screen is a clear demonstration of the utility of *in vitro* transposon screens for identifying genetic drivers of treatment resistance. There are several such reports of *in vitro* transposon-based cancer screens [271-274], supporting the promise of these tools for cancer screening.

EGFR inhibitors have proved to be successful in some cancers that carry *EGFR* mutations, such as lung and colorectal cancers, but have not shown to improve survival in GBM. In the latter case, there must be mechanisms for tumor cells to resist growth inhibition by these drugs, although these mechanisms are poorly understood. In this PhD, I established multiple primary cultures from *EGFRvIII*-PB tumors; these can be expanded *in vitro* and then injected into mice that can be treated with EGFR inhibitors (or the cells can be directly treated with these drugs for *in vitro* screening). As tumors develop *in vivo* in the presence of continued drug treatment, they will develop genetic alterations driven by *piggyBac* insertions that will enable them to escape growth inhibition by EGFR inhibitors. Sequencing for *piggyBac* common integration sites in the resulting tumors will help identify these genetic drivers of drug resistance. Understanding these genetic alterations may help enable design of rational combinatorial therapies involving EGFR inhibitors for treating GBM patients.

Another important aspect of our work worth exploring in future is how the nature of the initiating driver mutation in gliomas affects the timing and nature of subsequent genetic drivers. It is clear that many driver genes are only acting as such in particular contexts, such as in cooperation with other genes like *EGFR*. Therefore, depending what the initiating cancer mutation is, a cancer is likely to be predisposed to evolving in a particular way with clonal selection for certain mutations over others. This hypothesis is challenging to explore in end-stage tumors from patients using statistical methods with sequencing data alone. Rather, modelling in mice carrying these sensitizing mutations is an orthogonal and potentially clear-cut way for tackling this challenge. Given the time-constraints of completing a PhD and the long period of time needed for crossing mice and generating tumors, it was not possible to complete the *Trp53*-transposon screen for this thesis. This project will hence be on-going and once the results of this genome-wide forward genetic screen are known, it would be interesting to compare the CIS from *Trp53*-induced gliomas with those of *EGFRvIII* gliomas. A *Sleeping Beauty* screen for intestinal cancer drivers discovered that there were different patterns of CIS genes depending on which sensitizing mutations were

carried by the mice (*Apc*, *Smad4*, *Trp53* or *Kras* mutations), consistent with the notion that the founding mutation influences the genetic evolution of a tumor [53].

EGFR as a therapeutic target in gliomas

Given that *EGFR* was the first oncogene to be associated with glioblastoma (GBM), it is worth considering the therapies directed against EGFR that have been and are being developed for treatment of this disease. In the period when the first oncogenes in cancer were being described, it was discovered that the protein sequence of EGFR was similar to the viral oncogene, v-erb B, suggesting EGFR itself may have oncogenic activity[91]. Since then amplifications and various mutations, particularly truncating mutations that cause constitutive activation of the receptor, were described in up to 60% of GBMs. The *EGFRvIII* variant attracted particular interest, given that deletion of exons 2 -7 in this mutant leads to a novel antigenic epitope that is specific to this cancer and not expressed in normal tissues, forming a GBM 'signature molecule'. Various methods of targeting *EGFR* amplification and / or *EGFRvIII* have been developed, including small-molecule tyrosine kinase inhibitors, monoclonal antibodies, conjugated antibodies, CAR-T cells and vaccines. I will discuss the key agents, the challenges faced with these therapies, and potential future directions for EGFR-based therapies in glioblastoma.

EGFR as small molecule inhibitor target

A number of small molecule tyrosine kinase inhibitors (TKIs) are available and approved for a variety of cancers, including colon, pancreas and lung, although none are thus far approved for the treatment of glioblastoma due to disappointing results in clinical trials to date. The main such agents include erlotinib, gefitinib, afatinib and lapatinib. Erlotinib alone demonstrated no clinical efficacy in a trial in newly diagnosed GBM patients [143], and gefitinib did not improve overall survival in a phase II trial[275]. Afatinib and lapatinib showed very limited efficacy as single agents in early clinical trials in recurrent GBMs [276, 277].

A major problem with these TKIs for treating GBMs is poor brain (and more specifically tumor) penetrance in human patients. This itself is also difficult to measure, save for novel mass spectrophotometric methods that can be applied in animal models to more accurately measure drug-tissue concentrations. In clinical trials including tissue measurements of erlotinib and gefitinib, available from recurrent tumor resections, the concentration of erlotinib in GBM was only 5-7% that of the plasma, which may at least partly explain its poor results, although the concentration of gefitinib in GBM tissue was better (2 – 3 times the plasma concentration)[278]. Another important challenge is that the fact that these cancers display an 'adaptive' capacity: GBM cells activate many redundant pathways (and also genes in the same pathway, such as *Nf1* and *Spred1* in the Ras pathway as we found in our mouse gliomas), so they can overcome inhibition of a single molecule within one of these pathways.

EGFR as an immunotherapy target

Monoclonal antibodies can be developed in different ways to produce different effects on their target, such as blocking a receptor (in this case EGFR) and preventing ligand binding, causing internalisation and degradation of the receptor, binding the target and activating antibody-dependent cell-mediated cytotoxicity (ADCC), or binding the target and causing cell damage through a conjugated toxin.

Cetuximab is a monoclonal antibody used in colorectal cancer and has been trialled in GBM; it is a blocking antibody for EGFR. In orthotopic xenograft models of GBM, cetuximab in combination with VEGF inhibition led to reduced tumor migration and invasion[279]. However, in clinical trials cetuximab did not improve outcomes in recurrent GBM either as a single agent or with other agents [280, 281]. A recognized difficulty in using cetuximab for GBM is also related to tumor penetrance in the central nervous system (including overcoming the blood-brain barrier). Therefore, current developments underway are aimed at improving tumor tissue concentrations of cetuximab, including direct intracranial infusion of the antibody, intracranial injection of adenovirus containing the antibody gene so that transduced cells produce the antibody to

increase local concentrations, and selective osmotic blood-brain barrier opening with intra-arterial mannitol infusion and then cetuximab infusion. It remains to be determined whether these methods will improve clinical efficacy of the antibody.

Nimotuzumab is another EGFR blocking antibody, which differs from cetuximab in having a lower binding affinity for EGFR and is therefore more selective for targeting EGFR-overexpressing cells (as in GBM) compared with normal cells that also express EGFR[282]. It showed potential efficacy in a phase II trial and also in a randomised phase III trials using nimotuzumab in addition to standard therapy (radiotherapy and temozolomide) for GBM[283, 284]. It is currently being explored further in subgroups of patients, including paediatric diffuse intrinsic pontine glioma.

A promising avenue for therapeutic based on EGFR as a target in GBM is the engineering of T-cells to express a chimeric antigen receptor to recognise a target independently of MHC-mediated antigen presentation, named chimeric antigen receptor T-cells (CAR-T-cells). This has shown efficacy in certain cancers such as leukaemia[285]. EGFRvIII in GBM is a particularly attractive target for this approach given that it is a unique antigen that is specific for the cancer and not expressed on normal tissues. CAR-T-cells against EGFRvIII are in development and some are in early phase clinical studies[286].

As mentioned, EGFRvIII contains a unique epitope that does not occur in normal tissues; as such, a vaccine, rindopepimut, has been developed containing the unique amino-acid sequence of EGFRvIII[145]. When this peptide is injected intradermally, it has been shown that an immunologic response against the peptide is mounted, which can lead to immune-mediated destruction of EGFRvIII-positive GBM cells[287]. Although rindopepimut showed good results in early phase trials, the recent phase III trial did not show improvements in overall patient survival with this vaccine[288].

The reasons for lack of success of these various agents targeting EGFR are complex. Aside from the issue of drug delivery into GBMs (requiring passage through the blood-brain barrier), another major challenge is the intratumoral heterogeneity in *EGFR* expression. This heterogeneity has been observed for EGFRvIII expression, in that although *EGFRvIII* is common amongst GBM

patients, it is not expressed in all tumor cells[289], so therapies such as the EGFRvIII vaccine do not lead to destruction of all GBM cells. More recent sequencing studies have identified other activating *EGFR* mutations in GBM, apart from *EGFRvIII*, including exon 12-13 deletion, exon 14-15 deletion, and C-terminal deletion of exon 25-27, as well as point mutations and gene fusions involving *EGFR* [27, 290]. These various *EGFR* mutations can co-exist in one GBM – single cell sequencing analyses have found that up to 32 different tumor subpopulations can be present in a single GBM biopsy with each one containing a different pattern of *EGFR* mutations[291]; this complexity presents potential mechanisms for resistance to single EGFR-targeting therapies. Another problem is the documented co-expression of multiple receptor tyrosine kinases in GBMs, including MET and PDGFRa in addition to EGFR[292]. Thus, combinations of small molecule inhibitors targeting multiple RTKs are likely to be more successful than single agents. Some evidence has also emerged that GBMs can develop resistance to EGFR inhibitors because these cells can carry the EGFR amplification on double minutes (extrachromosomal DNA); when these tumors are treated with EGFR inhibitors, the cells lose their double minutes containing the EGFR amplification, and when treatment is stopped these double minutes can re-appear[149].

A very recent study aimed to elucidate the pharmacogenomic landscape of patient-derived tumor cells (PDCs) from 385 tumors across 14 cancer types [293]. The study demonstrated these cells reflected the genetics and biology of the disease more accurately than cancer cell lines and patient-derived xenograft models. Subgroup analysis of *EGFR*-altered GBMs found that *EGFR* amplification, *EGFRvIII*, *EGFR* point mutations and fusions all predicted sensitivity to multiple EGFR inhibitors. Moreover, they found that in *EGFR*-altered GBM PDCs resistance to EGFR inhibitors could be overcome by the use of ibrutinib, a drug currently used in haematological malignancies that acts by inhibiting phosphorylation of Bruton's tyrosine kinase (BTK). *EGFR* amplification and *EGFRvIII* both conferred sensitivity to ibrutinib in GBM PDCs suggesting equivalent driving effects of these types of *EGFR* alteration. Given ibrutinib is able to cross the blood-brain barrier, this is a potential therapeutic approach worth exploring, although testing *in vivo* in genetically faithful models is required.

Future Challenges in Glioma Management

Over the last few years, the genetic landscape of gliomas has been the subject of intense investigation and it is very likely that the major drivers of these tumors will be defined more clearly over the next decade. The question arises then of what should we do with all of these data? Of course, it will be important to tailor what we know about these tumors as a population to individual patients, who will carry their own cluster of driver mutations in their tumor. As our knowledge and understanding of these tumors improves, so does our classification of the tumors into distinct subtypes based on the molecular data [34, 93, 173, 294]. Indeed, recent detailed molecular characterization of CNS primitive neuroectodermal tumors (PNETs) led to discovery of distinct new brain tumor identities [295]. Having highly specific classifications will enable us to make our management of patients more personalized, ideally with prognosis of patients being accurately reflected in the molecular subtype of glioma they have. Even more advanced than this would be to personalize therapy for patients based on the molecular profile of their tumor; at the most comprehensive level, this would involve whole-genomic profiling and transcriptomic sequencing of each patient's tumor, not only for the most accurate classification but also potentially to give drugs based on their unique genetic and epigenetic tumor profile. Undoubtedly, this will be complicated by issues such as rare subclonal driver mutations and intratumor heterogeneity, which is particularly marked for glioblastomas compared with other cancers. Drugs may require re-engineering in order to penetrate the blood-brain barrier. What does this mean in practice – will we require sequencing of tumors from multiple sites for every single patient in order to best select a therapy based on common genetic alterations across most sites? Taking the example of EGFR directed therapy, it has been noted that for tumors carrying the *EGFRvIII* mutation that not all cells in the tumor actually express the mutant protein and therefore giving these patients EGFR inhibitors is unlikely to lead to cure since not all cells will be inhibited by these drugs, leaving aside the complex resistance mechanisms cancers can develop after this and the difficulties with drug penetrance into the tumor. Other problems with approaches involving inhibition of oncogenes are that for oncogenes acquired early during carcinogenesis the tumors may no longer be dependent on these oncogenes for growth as they have acquired many more drivers, making early ones redundant. This would demand the need

for targeting of multiple independent cancer genes in order to have a durable suppression of tumor growth[144]. A further problem is that many of the altered genes found in gliomas, including in my work here with *piggyBac* mutagenesis and identification of mutations in mouse gliomas, are tumor suppressors rather than oncogenes. These genes are more difficult to target therapeutically, as they may require re-expression rather inhibition (which can be done using drugs in many cases). However, tumor suppressors (and their downstream pathways) are increasingly regarded as potentially powerful therapeutic targets [296], particularly if a definite structure such as a pocket can be identified, as exemplified by molecules blocking the interaction of p53 with MDM2 thus increasing wild-type p53[297]. Targeting the downstream activated pathway following loss of a tumor suppressor gene can also be an attractive approach, as exemplified by PI3K inhibitors which are being explored as therapeutic options in cancers with *PTEN* loss. Indeed, the confirmation that *Pten* loss accelerates leptomeningeal spinal LGG growth *in vivo* in my work suggests that PI3K inhibitors may be worth exploring as a potential therapeutic strategy for these tumors. However, the precise signalling pathways promoting tumorigenesis in this context need further exploration and the extent of Pten loss in human spinal tumors needs confirmation in larger studies [298].

Perhaps a complementary way of tackling gliomas is to determine how the proteins expressed on the cell surface of tumor cells differ from those of normal cells; this may allow us to define rationale targets on cancer cells for designing destructive therapeutic agents, such as antibodies or CAR-T cells, which would leave our normal cells alone and therefore potentially have few systemic side effects. In any case, it is almost certain that durable remissions of glioblastoma and other gliomas will only come about through multiple, complementary therapies that have been rationally based on molecular profiles of this cancer.

References

1. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
2. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome.* Nature, 2009. **458**(7239): p. 719-24.
3. Cheng, N., et al., *Transforming growth factor-beta signaling-deficient fibroblasts enhance hepatocyte growth factor signaling in mammary carcinoma cells to promote scattering and invasion.* Mol Cancer Res, 2008. **6**(10): p. 1521-33.
4. Amit, I., et al., *A module of negative feedback regulators defines growth factor signaling.* Nat Genet, 2007. **39**(4): p. 503-12.
5. Jiang, B.H. and L.Z. Liu, *PI3K/PTEN signaling in angiogenesis and tumorigenesis.* Adv Cancer Res, 2009. **102**: p. 19-65.
6. Collado, M. and M. Serrano, *Senescence in tumours: evidence from mice and humans.* Nat Rev Cancer, 2010. **10**(1): p. 51-7.
7. Ghebranious, N. and L.A. Donehower, *Mouse models in tumor suppression.* Oncogene, 1998. **17**(25): p. 3385-400.
8. Blasco, M.A., *Telomeres and human disease: ageing, cancer and beyond.* Nat Rev Genet, 2005. **6**(8): p. 611-22.
9. Ferrara, N., *Pathways mediating VEGF-independent tumor angiogenesis.* Cytokine Growth Factor Rev, 2010. **21**(1): p. 21-6.
10. Teng, M.W., et al., *Immune-mediated dormancy: an equilibrium with cancer.* J Leukoc Biol, 2008. **84**(4): p. 988-93.
11. Nelson, B.H., *The impact of T-cell immunity on ovarian cancer outcomes.* Immunol Rev, 2008. **222**: p. 101-16.
12. Sunshine, J. and J.M. Taube, *PD-1/PD-L1 inhibitors.* Curr Opin Pharmacol, 2015. **23**: p. 32-8.
13. Syn, N.L., et al., *De-novo and acquired resistance to immune checkpoint targeting.* Lancet Oncol, 2017. **18**(12): p. e731-e741.
14. Negrini, S., V.G. Gorgoulis, and T.D. Halazonetis, *Genomic instability--an evolving hallmark of cancer.* Nat Rev Mol Cell Biol, 2010. **11**(3): p. 220-8.
15. Korkola, J. and J.W. Gray, *Breast cancer genomes--form and function.* Curr Opin Genet Dev, 2010. **20**(1): p. 4-14.
16. Kinzler, K.W. and B. Vogelstein, *Cancer-susceptibility genes. Gatekeepers and caretakers.* Nature, 1997. **386**(6627): p. 761, 763.
17. Visvader, J.E. and G.J. Lindeman, *Cancer stem cells in solid tumours: accumulating evidence and unresolved questions.* Nat Rev Cancer, 2008. **8**(10): p. 755-68.
18. Caren, H., et al., *Glioblastoma Stem Cells Respond to Differentiation Cues but Fail to Undergo Commitment and Terminal Cell-Cycle Arrest.* Stem Cell Reports, 2015. **5**(5): p. 829-842.
19. Singh, A. and J. Settleman, *EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer.* Oncogene, 2010. **29**(34): p. 4741-51.
20. Kurzwelly, D., U. Herrlinger, and M. Simon, *Seizures in patients with low-grade gliomas--incidence, pathogenesis, surgical management, and pharmacotherapy.* Adv Tech Stand Neurosurg, 2010. **35**: p. 81-111.
21. Noorani, I. and N. Sanai, *Surgical Management of Incidental Gliomas.* Neurosurg Clin N Am, 2017. **28**(3): p. 397-406.

22. Mandonnet, E., et al., *Silent diffuse low-grade glioma: toward screening and preventive treatment?* Cancer, 2014. **120**(12): p. 1758-62.
23. Pallud, J., et al., *The silent phase of diffuse low-grade gliomas. Is it when we missed the action?* Acta Neurochir (Wien), 2013. **155**(12): p. 2237-42.
24. Stupp, R., et al., *Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma.* N Engl J Med, 2005. **352**(10): p. 987-96.
25. Perry, J.R., et al., *Short-Course Radiation plus Temozolomide in Elderly Patients with Glioblastoma.* N Engl J Med, 2017. **376**(11): p. 1027-1037.
26. Gilbert, M.R., et al., *A randomized trial of bevacizumab for newly diagnosed glioblastoma.* N Engl J Med, 2014. **370**(8): p. 699-708.
27. Brennan, C.W., et al., *The somatic genomic landscape of glioblastoma.* Cell, 2013. **155**(2): p. 462-77.
28. Cancer Genome Atlas Research, N., *Comprehensive genomic characterization defines human glioblastoma genes and core pathways.* Nature, 2008. **455**(7216): p. 1061-8.
29. Holland, E.C., et al., *Combined activation of Ras and Akt in neural progenitors induces glioblastoma formation in mice.* Nat Genet, 2000. **25**(1): p. 55-7.
30. Parsons, D.W., et al., *An integrated genomic analysis of human glioblastoma multiforme.* Science, 2008. **321**(5897): p. 1807-12.
31. Sasaki, M., et al., *IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics.* Nature, 2012. **488**(7413): p. 656-9.
32. Sasaki, M., et al., *D-2-hydroxyglutarate produced by mutant IDH1 perturbs collagen maturation and basement membrane function.* Genes Dev, 2012. **26**(18): p. 2038-49.
33. Bardella, C., et al., *Expression of Idh1(R132H) in the Murine Subventricular Zone Stem Cell Niche Recapitulates Features of Early Gliomagenesis.* Cancer Cell, 2016. **30**(4): p. 578-594.
34. Eckel-Passow, J.E., et al., *Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors.* N Engl J Med, 2015. **372**(26): p. 2499-508.
35. Tobin, M.K., et al., *Intramedullary spinal cord tumors: a review of current and future treatment strategies.* Neurosurg Focus, 2015. **39**(2): p. E14.
36. Jallo, G.I., D. Freed, and F. Epstein, *Intramedullary spinal cord tumors in children.* Childs Nerv Syst, 2003. **19**(9): p. 641-9.
37. Raco, A., et al., *Long-term follow-up of intramedullary spinal cord tumors: a series of 202 cases.* Neurosurgery, 2005. **56**(5): p. 972-81; discussion 972-81.
38. Chamberlain, M.C. and T.L. Tredway, *Adult primary intradural spinal cord tumors: a review.* Curr Neurol Neurosci Rep, 2011. **11**(3): p. 320-8.
39. Chamberlain, M.C., *Temozolomide for recurrent low-grade spinal cord gliomas in adults.* Cancer, 2008. **113**(5): p. 1019-24.
40. De Witt Hamer, P.C., et al., *The genomic profile of human malignant glioma is altered early in primary cell culture and preserved in spheroids.* Oncogene, 2008. **27**(14): p. 2091-6.
41. Mouse Genome Sequencing, C., et al., *Initial sequencing and comparative analysis of the mouse genome.* Nature, 2002. **420**(6915): p. 520-62.
42. Trotman, L.C., et al., *Pten dose dictates cancer progression in the prostate.* PLoS Biol, 2003. **1**(3): p. E59.
43. Olive, K.P., et al., *Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome.* Cell, 2004. **119**(6): p. 847-60.
44. Silver, L., *Mouse Genetics - Concepts and Applications. .* Oxford University Press, 1995.
45. Uren, A.G., et al., *Retroviral insertional mutagenesis: past, present and future.* Oncogene, 2005. **24**(52): p. 7656-72.

46. Mikkers, H., et al., *High-throughput retroviral tagging to identify components of specific signaling pathways in cancer.* Nat Genet, 2002. **32**(1): p. 153-9.
47. Mikkers, H., J. Allen, and A. Berns, *Proviral activation of the tumor suppressor E2a contributes to T cell lymphomagenesis in EmuMyc transgenic mice.* Oncogene, 2002. **21**(43): p. 6559-66.
48. Ivics, Z., et al., *Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells.* Cell, 1997. **91**(4): p. 501-10.
49. Cui, Z., et al., *Structure-function analysis of the inverted terminal repeats of the sleeping beauty transposon.* J Mol Biol, 2002. **318**(5): p. 1221-35.
50. Geurts, A.M., et al., *Gene transfer into genomes of human cells by the sleeping beauty transposon system.* Mol Ther, 2003. **8**(1): p. 108-17.
51. Mates, L., et al., *Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates.* Nat Genet, 2009. **41**(6): p. 753-61.
52. Collier, L.S., et al., *Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse.* Nature, 2005. **436**(7048): p. 272-6.
53. Takeda, H., et al., *Transposon mutagenesis identifies genes and evolutionary forces driving gastrointestinal tract tumor progression.* Nat Genet, 2015. **47**(2): p. 142-50.
54. Kas, S.M., et al., *Insertional mutagenesis identifies drivers of a novel oncogenic pathway in invasive lobular breast carcinoma.* Nat Genet, 2017. **49**(8): p. 1219-1230.
55. Bard-Chapeau, E.A., et al., *Transposon mutagenesis identifies genes driving hepatocellular carcinoma in a chronic hepatitis B mouse model.* Nat Genet, 2014. **46**(1): p. 24-32.
56. Wu, X., et al., *Clonal selection drives genetic divergence of metastatic medulloblastoma.* Nature, 2012. **482**(7386): p. 529-33.
57. Rahrmann, E.P., et al., *Forward genetic screen for malignant peripheral nerve sheath tumor formation identifies new genes and pathways driving tumorigenesis.* Nat Genet, 2013. **45**(7): p. 756-66.
58. Moriarity, B.S., et al., *A Sleeping Beauty forward genetic screen identifies new genes and pathways driving osteosarcoma development and metastasis.* Nat Genet, 2015. **47**(6): p. 615-24.
59. Chen, L., et al., *Transposon insertional mutagenesis in mice identifies human breast cancer susceptibility genes and signatures for stratification.* Proc Natl Acad Sci U S A, 2017. **114**(11): p. E2215-E2224.
60. Mann, M.B., et al., *Transposon mutagenesis identifies genetic drivers of Braf(V600E) melanoma.* Nat Genet, 2015. **47**(5): p. 486-95.
61. Dupuy, A.J., et al., *Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system.* Nature, 2005. **436**(7048): p. 221-6.
62. Mann, K.M., et al., *Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic adenocarcinoma.* Proc Natl Acad Sci U S A, 2012. **109**(16): p. 5934-41.
63. Rahrmann, E.P., et al., *Identification of PDE4D as a proliferation promoting factor in prostate cancer using a Sleeping Beauty transposon-based somatic mutagenesis screen.* Cancer Res, 2009. **69**(10): p. 4388-97.
64. Bender, A.M., et al., *Sleeping beauty-mediated somatic mutagenesis implicates CSF1 in the formation of high-grade astrocytomas.* Cancer Res, 2010. **70**(9): p. 3557-65.
65. Collier, L.S., et al., *Whole-body sleeping beauty mutagenesis can cause penetrant leukemia/lymphoma and rare high-grade glioma without associated embryonic lethality.* Cancer Res, 2009. **69**(21): p. 8429-37.
66. Koso, H., et al., *Transposon mutagenesis identifies genes that transform neural stem cells into glioma-initiating cells.* Proc Natl Acad Sci U S A, 2012. **109**(44): p. E2998-3007.

67. de la Rosa, J., et al., *A single-copy Sleeping Beauty transposon mutagenesis screen identifies new PTEN-cooperating tumor suppressor genes.* Nat Genet, 2017. **49**(5): p. 730-741.
68. Ding, S., et al., *Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice.* Cell, 2005. **122**(3): p. 473-83.
69. Rad, R., et al., *PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice.* Science, 2010. **330**(6007): p. 1104-7.
70. Rad, R., et al., *A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer.* Nat Genet, 2015. **47**(1): p. 47-56.
71. Urnov, F.D., et al., *Genome editing with engineered zinc finger nucleases.* Nat Rev Genet, 2010. **11**(9): p. 636-46.
72. Miller, J.C., et al., *A TALE nuclease architecture for efficient genome editing.* Nat Biotechnol, 2011. **29**(2): p. 143-8.
73. Ishino, Y., et al., *Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product.* J Bacteriol, 1987. **169**(12): p. 5429-33.
74. Mojica, F.J., et al., *Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria.* Mol Microbiol, 2000. **36**(1): p. 244-6.
75. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-12.
76. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.
77. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.
78. Fu, Y., et al., *High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.* Nat Biotechnol, 2013. **31**(9): p. 822-6.
79. Wang, T., et al., *Genetic screens in human cells using the CRISPR-Cas9 system.* Science, 2014. **343**(6166): p. 80-4.
80. Shalem, O., et al., *Genome-scale CRISPR-Cas9 knockout screening in human cells.* Science, 2014. **343**(6166): p. 84-87.
81. Koike-Yusa, H., et al., *Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library.* Nat Biotechnol, 2014. **32**(3): p. 267-73.
82. Gilbert, L.A., et al., *Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation.* Cell, 2014. **159**(3): p. 647-61.
83. Konermann, S., et al., *Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex.* Nature, 2015. **517**(7536): p. 583-8.
84. Golic, K.G. and S. Lindquist, *The FLP recombinase of yeast catalyzes site-specific recombination in the Drosophila genome.* Cell, 1989. **59**(3): p. 499-509.
85. Friedel, R.H., et al., *Generating conditional knockout mice.* Methods Mol Biol, 2011. **693**: p. 205-31.
86. Fisher, G.H., et al., *Development of a flexible and specific gene delivery system for production of murine tumor models.* Oncogene, 1999. **18**(38): p. 5253-60.
87. Tronche, F., et al., *Disruption of the glucocorticoid receptor gene in the nervous system results in reduced anxiety.* Nat Genet, 1999. **23**(1): p. 99-103.
88. Dubois, N.C., et al., *Nestin-Cre transgenic mouse line Nes-Cre1 mediates highly efficient Cre/loxP mediated recombination in the nervous system, kidney, and somite-derived tissues.* Genesis, 2006. **44**(8): p. 355-60.

89. Liang, H., S. Hippenmeyer, and H.T. Ghashghaei, *A Nestin-cre transgenic mouse is insufficient for recombination in early embryonic neural progenitors.* Biol Open, 2012. **1**(12): p. 1200-3.
90. Zhuo, L., et al., *hGFAP-cre transgenic mice for manipulation of glial and neuronal function in vivo.* Genesis, 2001. **31**(2): p. 85-94.
91. Downward, J., et al., *Close similarity of epidermal growth factor receptor and v-erb-B oncogene protein sequences.* Nature, 1984. **307**(5951): p. 521-7.
92. Holland, E.C., et al., *A constitutively active epidermal growth factor receptor cooperates with disruption of G1 cell-cycle arrest pathways to induce glioma-like lesions in mice.* Genes Dev, 1998. **12**(23): p. 3675-85.
93. Ceccarelli, M., et al., *Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma.* Cell, 2016. **164**(3): p. 550-63.
94. Zheng, H., et al., *p53 and Pten control neural and glioma stem/progenitor cell renewal and differentiation.* Nature, 2008. **455**(7216): p. 1129-33.
95. Alcantara Llaguno, S.R., et al., *Adult Lineage-Restricted CNS Progenitors Specify Distinct Glioblastoma Subtypes.* Cancer Cell, 2015. **28**(4): p. 429-440.
96. Zhu, H., et al., *Oncogenic EGFR signaling cooperates with loss of tumor suppressor gene functions in gliomagenesis.* Proc Natl Acad Sci U S A, 2009. **106**(8): p. 2712-6.
97. Jacques, T.S., et al., *Combinations of genetic mutations in the adult neural stem cell compartment determine brain tumour phenotypes.* EMBO J, 2010. **29**(1): p. 222-35.
98. Bachoo, R.M., et al., *Epidermal growth factor receptor and Ink4a/Arf: convergent mechanisms governing terminal differentiation and transformation along the neural stem cell to astrocyte axis.* Cancer Cell, 2002. **1**(3): p. 269-77.
99. Friedmann-Morvinski, D., et al., *Dedifferentiation of neurons and astrocytes by oncogenes can induce gliomas in mice.* Science, 2012. **338**(6110): p. 1080-4.
100. Doetsch, F., et al., *EGF converts transit-amplifying neurogenic precursors in the adult brain into multipotent stem cells.* Neuron, 2002. **36**(6): p. 1021-34.
101. Liu, C., et al., *Mosaic analysis with double markers reveals tumor cell of origin in glioma.* Cell, 2011. **146**(2): p. 209-21.
102. Lee, J.H., et al., *Human glioblastoma arises from subventricular zone cells with low-level driver mutations.* Nature, 2018.
103. Galvao, R.P., et al., *Transformation of quiescent adult oligodendrocyte precursor cells into malignant glioma through a multistep reactivation process.* Proc Natl Acad Sci U S A, 2014. **111**(40): p. E4214-23.
104. Hitoshi, Y., et al., *Spinal glioma: platelet-derived growth factor B-mediated oncogenesis in the spinal cord.* Cancer Res, 2008. **68**(20): p. 8507-15.
105. Bussolati, G., et al., *Formalin fixation at low temperature better preserves nucleic acid integrity.* PLoS One, 2011. **6**(6): p. e21043.
106. Jentsch, I., et al., *Karyotyping mouse chromosomes by multiplex-FISH (M-FISH).* Chromosome Res, 2001. **9**(3): p. 211-4.
107. Li, H. and R. Durbin, *Fast and accurate long-read alignment with Burrows-Wheeler transform.* Bioinformatics, 2010. **26**(5): p. 589-95.
108. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.
109. Cibulskis, K., et al., *Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples.* Nat Biotechnol, 2013. **31**(3): p. 213-9.

110. Cingolani, P., et al., *A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3.* Fly (Austin), 2012. **6**(2): p. 80-92.
111. Dees, N.D., et al., *MuSiC: identifying mutational significance in cancer genomes.* Genome Res, 2012. **22**(8): p. 1589-98.
112. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
113. Koboldt, D.C., D.E. Larson, and R.K. Wilson, *Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection.* Curr Protoc Bioinformatics, 2013. **44**: p. 15 4 1-17.
114. Olshen, A.B., et al., *Circular binary segmentation for the analysis of array-based DNA copy number data.* Biostatistics, 2004. **5**(4): p. 557-72.
115. Mermel, C.H., et al., *GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers.* Genome Biol, 2011. **12**(4): p. R41.
116. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
117. Aken, B.L., et al., *Ensembl 2017.* Nucleic Acids Res, 2017. **45**(D1): p. D635-D642.
118. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
119. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
120. Mootha, V.K., et al., *PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.* Nat Genet, 2003. **34**(3): p. 267-73.
121. de Ruiter, J.R., et al., *Identifying transposon insertions and their effects from RNA-sequencing data.* Nucleic Acids Res, 2017. **45**(12): p. 7064-7077.
122. Friedrich, M.J., et al., *Genome-wide transposon screening and quantitative insertion site sequencing for cancer gene discovery in mice.* Nat Protoc, 2017. **12**(2): p. 289-309.
123. de Ridder, J., et al., *Detecting statistically significant common insertion sites in retroviral insertional mutagenesis screens.* PLoS Comput Biol, 2006. **2**(12): p. e166.
124. Yates, A., et al., *Ensembl 2016.* Nucleic Acids Res, 2016. **44**(D1): p. D710-6.
125. Forbes, S.A., et al., *COSMIC: somatic cancer genetics at high-resolution.* Nucleic Acids Res, 2017. **45**(D1): p. D777-D783.
126. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.* Cancer Discov, 2012. **2**(5): p. 401-4.
127. Wong, A.J., et al., *Increased expression of the epidermal growth factor receptor gene in malignant gliomas is invariably associated with gene amplification.* Proc Natl Acad Sci U S A, 1987. **84**(19): p. 6899-903.
128. Klingler, S., et al., *Development of Resistance to EGFR-Targeted Therapy in Malignant Glioma Can Occur through EGFR-Dependent and -Independent Mechanisms.* Cancer Res, 2015. **75**(10): p. 2109-19.
129. Nikolaev, S., et al., *Extrachromosomal driver mutations in glioblastoma and low-grade glioma.* Nat Commun, 2014. **5**: p. 5690.
130. Choe, G., et al., *Analysis of the phosphatidylinositol 3'-kinase signaling pathway in glioblastoma patients in vivo.* Cancer Res, 2003. **63**(11): p. 2742-6.
131. Akhavan, D., T.F. Cloughesy, and P.S. Mischel, *mTOR signaling in glioblastoma: lessons learned from bench to bedside.* Neuro Oncol, 2010. **12**(8): p. 882-9.
132. Fan, Q.W., et al., *EGFR phosphorylates tumor-derived EGFRvIII driving STAT3/5 and progression in glioblastoma.* Cancer Cell, 2013. **24**(4): p. 438-49.

133. Huang, P.H., et al., *Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma.* Proc Natl Acad Sci U S A, 2007. **104**(31): p. 12867-72.
134. Al-Nedawi, K., et al., *Intercellular transfer of the oncogenic receptor EGFRvIII by microvesicles derived from tumour cells.* Nat Cell Biol, 2008. **10**(5): p. 619-24.
135. Nishikawa, R., et al., *Immunohistochemical analysis of the mutant epidermal growth factor, deltaEGFR, in glioblastoma.* Brain Tumor Pathol, 2004. **21**(2): p. 53-6.
136. Mukasa, A., et al., *Mutant EGFR is required for maintenance of glioma growth in vivo, and its ablation leads to escape from receptor dependence.* Proc Natl Acad Sci U S A, 2010. **107**(6): p. 2616-21.
137. Mitsudomi, T., et al., *Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial.* Lancet Oncol, 2010. **11**(2): p. 121-8.
138. Zhang, Z., et al., *Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer.* Nat Genet, 2012. **44**(8): p. 852-60.
139. Sos, M.L., et al., *PTEN loss contributes to erlotinib resistance in EGFR-mutant lung cancer by activation of Akt and EGFR.* Cancer Res, 2009. **69**(8): p. 3256-61.
140. Engelman, J.A., et al., *MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling.* Science, 2007. **316**(5827): p. 1039-43.
141. Pao, W., et al., *Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain.* PLoS Med, 2005. **2**(3): p. e73.
142. Mellinghoff, I.K., et al., *Molecular determinants of the response of glioblastomas to EGFR kinase inhibitors.* N Engl J Med, 2005. **353**(19): p. 2012-24.
143. Peereboom, D.M., et al., *Phase II trial of erlotinib with temozolomide and radiation in patients with newly diagnosed glioblastoma multiforme.* J Neurooncol, 2010. **98**(1): p. 93-9.
144. Stommel, J.M., et al., *Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies.* Science, 2007. **318**(5848): p. 287-90.
145. Gedeon, P.C., et al., *Rindopepimut: anti-EGFRvIII peptide vaccine, oncolytic.* Drugs Future, 2013. **38**(3): p. 147-155.
146. Mishima, K., et al., *Growth suppression of intracranial xenografted glioblastomas overexpressing mutant epidermal growth factor receptors by systemic administration of monoclonal antibody (mAb) 806, a novel monoclonal antibody directed to the receptor.* Cancer Res, 2001. **61**(14): p. 5349-54.
147. Mukherjee, B., et al., *EGFRvIII and DNA double-strand break repair: a molecular mechanism for radioresistance in glioblastoma.* Cancer Res, 2009. **69**(10): p. 4252-9.
148. Golding, S.E., et al., *Pro-survival AKT and ERK signaling from EGFR and mutant EGFRvIII enhances DNA double-strand break repair in human glioma cells.* Cancer Biol Ther, 2009. **8**(8): p. 730-8.
149. Nathanson, D.A., et al., *Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA.* Science, 2014. **343**(6166): p. 72-6.
150. Vogt, N., et al., *Molecular structure of double-minute chromosomes bearing amplified copies of the epidermal growth factor receptor gene in gliomas.* Proc Natl Acad Sci U S A, 2004. **101**(31): p. 11368-73.
151. Akhavan, D., et al., *De-repression of PDGFRbeta transcription promotes acquired resistance to EGFR tyrosine kinase inhibitors in glioblastoma patients.* Cancer Discov, 2013. **3**(5): p. 534-47.
152. Clark, P.A., et al., *Activation of multiple ERBB family receptors mediates glioblastoma cancer stem-like cell resistance to EGFR-targeted inhibition.* Neoplasia, 2012. **14**(5): p. 420-8.

153. Schulte, A., et al., *Erlotinib resistance in EGFR-amplified glioblastoma cells is associated with upregulation of EGFRvIII and PI3Kp110delta.* Neuro Oncol, 2013. **15**(10): p. 1289-301.
154. Liu, F., et al., *EGFR Mutation Promotes Glioblastoma through Epigenome and Transcription Factor Network Remodeling.* Mol Cell, 2015. **60**(2): p. 307-18.
155. Lal, A., et al., *Mutant epidermal growth factor receptor up-regulates molecular effectors of tumor invasion.* Cancer Res, 2002. **62**(12): p. 3335-9.
156. Bai, H., et al., *Integrated genomic characterization of IDH1-mutant glioma malignant progression.* Nat Genet, 2016. **48**(1): p. 59-66.
157. Capper, D., et al., *DNA methylation-based classification of central nervous system tumours.* Nature, 2018. **555**(7697): p. 469-474.
158. Krex, D., et al., *Long-term survival with glioblastoma multiforme.* Brain, 2007. **130**(Pt 10): p. 2596-606.
159. Stichel, D., et al., *Distribution of EGFR amplification, combined chromosome 7 gain and chromosome 10 loss, and TERT promoter mutation in brain tumors and their potential for the reclassification of IDHwt astrocytoma to glioblastoma.* Acta Neuropathol, 2018.
160. Tabori, U., et al., *Epidermal growth factor receptor gene amplification and expression in disseminated pediatric low-grade gliomas.* J Neurosurg, 2005. **103**(4 Suppl): p. 357-61.
161. Yan, C., et al., *Glioblastoma multiforme in conus medullaris with intracranial metastasis after postoperative adjuvant therapy.* Medicine (Baltimore), 2017. **96**(13): p. e6500.
162. Zadnik, P.L., et al., *Spinal cord tumours: advances in genetics and their implications for treatment.* Nat Rev Neurol, 2013. **9**(5): p. 257-66.
163. Jackson, E.L., et al., *PDGFR alpha-positive B cells are neural stem cells in the adult SVZ that form glioma-like growths in response to increased PDGF signaling.* Neuron, 2006. **51**(2): p. 187-99.
164. Chamberlain, M.C., *Leptomeningeal metastases: a review of evaluation and treatment.* J Neurooncol, 1998. **37**(3): p. 271-84.
165. Singh, S.K., et al., *Identification of a cancer stem cell in human brain tumors.* Cancer Res, 2003. **63**(18): p. 5821-8.
166. Shah, N. and S. Sukumar, *The Hox genes and their roles in oncogenesis.* Nat Rev Cancer, 2010. **10**(5): p. 361-71.
167. Verhaak, R.G., et al., *Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1.* Cancer Cell, 2010. **17**(1): p. 98-110.
168. Chakravarthi, B.V., et al., *MicroRNA-101 regulated transcriptional modulator SUB1 plays a role in prostate cancer.* Oncogene, 2016. **35**(49): p. 6330-6340.
169. Fults, D., et al., *p53 mutation and loss of heterozygosity on chromosomes 17 and 10 during human astrocytoma progression.* Cancer Res, 1992. **52**(3): p. 674-9.
170. Sorensen, S.A., J.J. Mulvihill, and A. Nielsen, *Long-term follow-up of von Recklinghausen neurofibromatosis. Survival and malignant neoplasms.* N Engl J Med, 1986. **314**(16): p. 1010-5.
171. Lamar, J.M., et al., *The Hippo pathway target, YAP, promotes metastasis through its TEAD-interaction domain.* Proc Natl Acad Sci U S A, 2012. **109**(37): p. E2441-50.
172. Lin, L., et al., *The Hippo effector YAP promotes resistance to RAF- and MEK-targeted cancer therapies.* Nat Genet, 2015. **47**(3): p. 250-6.
173. Cancer Genome Atlas Research, N., et al., *Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas.* N Engl J Med, 2015. **372**(26): p. 2481-98.
174. Ichimura, K., et al., *IDH1 mutations are present in the majority of common adult gliomas but rare in primary glioblastomas.* Neuro Oncol, 2009. **11**(4): p. 341-7.

175. Vassiliou, G.S., et al., *Mutant nucleophosmin and cooperating pathways drive leukemia initiation and progression in mice.* Nat Genet, 2011. **43**(5): p. 470-5.
176. Hills, S.A. and J.F. Diffley, *DNA replication and oncogene-induced replicative stress.* Curr Biol, 2014. **24**(10): p. R435-44.
177. Nitta, M., et al., *Targeting EGFR induced oxidative stress by PARP1 inhibition in glioblastoma therapy.* PLoS One, 2010. **5**(5): p. e10767.
178. Zhu, Y., et al., *Inactivation of NF1 in CNS causes increased glial progenitor proliferation and optic glioma formation.* Development, 2005. **132**(24): p. 5577-88.
179. Alcantara Llaguno, S., et al., *Malignant astrocytomas originate from neural stem/progenitor cells in a somatic tumor suppressor mouse model.* Cancer Cell, 2009. **15**(1): p. 45-56.
180. Tuttle, A.H., et al., *Comparing phenotypic variation between inbred and outbred mice.* Nat Methods, 2018. **15**(12): p. 994-996.
181. Bifari, F., et al., *Neurogenic Radial Glia-like Cells in Meninges Migrate and Differentiate into Functionally Integrated Neurons in the Neonatal Cortex.* Cell Stem Cell, 2017. **20**(3): p. 360-373 e7.
182. Francis, J.M., et al., *EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing.* Cancer Discov, 2014. **4**(8): p. 956-71.
183. Patel, D., et al., *Monoclonal antibody cetuximab binds to and down-regulates constitutively activated epidermal growth factor receptor vIII on the cell surface.* Anticancer Res, 2007. **27**(5A): p. 3355-66.
184. Stutz, M.A., et al., *LRIG1 negatively regulates the oncogenic EGF receptor mutant EGFRvIII.* Oncogene, 2008. **27**(43): p. 5741-52.
185. Emlet, D.R., et al., *Targeting a glioblastoma cancer stem-cell population defined by EGF receptor variant III.* Cancer Res, 2014. **74**(4): p. 1238-49.
186. Inda, M.M., et al., *Tumor heterogeneity is an active process maintained by a mutant EGFR-induced cytokine circuit in glioblastoma.* Genes Dev, 2010. **24**(16): p. 1731-45.
187. Felsberg, J., et al., *Epidermal Growth Factor Receptor Variant III (EGFRvIII) Positivity in EGFR-Amplified Glioblastomas: Prognostic Role and Comparison between Primary and Recurrent Tumors.* Clin Cancer Res, 2017. **23**(22): p. 6846-6855.
188. Del Vecchio, C.A., et al., *EGFRvIII gene rearrangement is an early event in glioblastoma tumorigenesis and expression defines a hierarchy modulated by epigenetic mechanisms.* Oncogene, 2013. **32**(21): p. 2670-81.
189. Eskilsson, E., et al., *EGFRvIII mutations can emerge as late and heterogenous events in glioblastoma development and promote angiogenesis through Src activation.* Neuro Oncol, 2016. **18**(12): p. 1644-1655.
190. Pandita, A., et al., *Contrasting in vivo and in vitro fates of glioblastoma cell subpopulations with amplified EGFR.* Genes Chromosomes Cancer, 2004. **39**(1): p. 29-36.
191. Reuss, D.E., et al., *Adult IDH wild type astrocytomas biologically and clinically resolve into other tumor entities.* Acta Neuropathol, 2015. **130**(3): p. 407-17.
192. Lee, M., et al., *Intramedullary spinal cord tumors in neurofibromatosis.* Neurosurgery, 1996. **38**(1): p. 32-7.
193. Horbinski, C., et al., *Association of molecular alterations, including BRAF, with biology and outcome in pilocytic astrocytomas.* Acta Neuropathol, 2010. **119**(5): p. 641-9.
194. Burrell, R.A., et al., *Replication stress links structural and numerical cancer chromosomal instability.* Nature, 2013. **494**(7438): p. 492-496.
195. Campbell, P.J., et al., *The patterns and dynamics of genomic instability in metastatic pancreatic cancer.* Nature, 2010. **467**(7319): p. 1109-13.

196. de Bruin, E.C., et al., *Spatial and temporal diversity in genomic instability processes defines lung cancer evolution.* Science, 2014. **346**(6206): p. 251-6.
197. Gordon, D.J., B. Resio, and D. Pellman, *Causes and consequences of aneuploidy in cancer.* Nat Rev Genet, 2012. **13**(3): p. 189-203.
198. Armenia, J., et al., *The long tail of oncogenic drivers in prostate cancer.* Nat Genet, 2018. **50**(5): p. 645-651.
199. Wang, W., et al., *Chromosomal transposition of PiggyBac in mouse embryonic stem cells.* Proc Natl Acad Sci U S A, 2008. **105**(27): p. 9290-5.
200. Liang, Q., et al., *Chromosomal mobilization and reintegration of Sleeping Beauty and PiggyBac transposons.* Genesis, 2009. **47**(6): p. 404-8.
201. Li, M.A., et al., *The piggyBac transposon displays local and distant reintegration preferences and can cause mutations at noncanonical integration sites.* Mol Cell Biol, 2013. **33**(7): p. 1317-30.
202. Simon, M., et al., *Functional evidence for a role of combined CDKN2A (p16-p14(ARF))/CDKN2B (p15) gene inactivation in malignant gliomas.* Acta Neuropathol, 1999. **98**(5): p. 444-52.
203. Wakioka, T., et al., *Spred is a Sprouty-related suppressor of Ras signalling.* Nature, 2001. **412**(6847): p. 647-51.
204. Sorensen, S.A., J.J. Mulvihill, and A. Nielsen, *On the natural history of von Recklinghausen neurofibromatosis.* Ann N Y Acad Sci, 1986. **486**: p. 30-7.
205. Messiaen, L., et al., *Clinical and mutational spectrum of neurofibromatosis type 1-like syndrome.* JAMA, 2009. **302**(19): p. 2111-8.
206. Wang, S.I., et al., *Somatic mutations of PTEN in glioblastoma multiforme.* Cancer Res, 1997. **57**(19): p. 4183-6.
207. Quayle, S.N., et al., *Somatic mutations of PIK3R1 promote gliomagenesis.* PLoS One, 2012. **7**(11): p. e49466.
208. Ozawa, T., et al., *PDGFRA gene rearrangements are frequent genetic events in PDGFRA-amplified glioblastomas.* Genes Dev, 2010. **24**(19): p. 2205-18.
209. Azim, E., et al., *SOX6 controls dorsal progenitor identity and interneuron diversity during neocortical development.* Nat Neurosci, 2009. **12**(10): p. 1238-47.
210. Uittenbogaard, M. and A. Chiaramello, *Expression of the bHLH transcription factor Tcf12 (ME1) gene is linked to the expansion of precursor cell populations during neurogenesis.* Brain Res Gene Expr Patterns, 2002. **1**(2): p. 115-21.
211. Cohen-Dvashi, H., et al., *Navigator-3, a modulator of cell migration, may act as a suppressor of breast cancer progression.* EMBO Mol Med, 2015. **7**(3): p. 299-314.
212. Takata, M., et al., *The Rad51 paralog Rad51B promotes homologous recombinational repair.* Mol Cell Biol, 2000. **20**(17): p. 6476-82.
213. Zheng, H., et al., *PLAGL2 regulates Wnt signaling to impede differentiation in neural stem cells and gliomas.* Cancer Cell, 2010. **17**(5): p. 497-509.
214. Zhang, N., et al., *FoxM1 promotes beta-catenin nuclear localization and controls Wnt target-gene expression and glioma tumorigenesis.* Cancer Cell, 2011. **20**(4): p. 427-42.
215. Wu, G., et al., *Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas.* Nat Genet, 2012. **44**(3): p. 251-3.
216. Jenkins, R.B., et al., *A cytogenetic study of 53 human gliomas.* Cancer Genet Cytogenet, 1989. **39**(2): p. 253-79.
217. Mertens, F., et al., *Chromosomal imbalance maps of malignant solid tumors: a cytogenetic survey of 3185 neoplasms.* Cancer Res, 1997. **57**(13): p. 2765-80.

218. Saitoh, Y., et al., *Identification of allelic loss on chromosome arm 6p in human astrocytomas by arbitrarily primed polymerase chain reaction.* Genes Chromosomes Cancer, 1998. **22**(3): p. 165-70.
219. Fults, D., et al., *Allelotype of human malignant astrocytoma.* Cancer Res, 1990. **50**(18): p. 5784-9.
220. Teyssier, J.R. and D. Ferre, *Identification of a clustering of chromosomal breakpoints in the analysis of 203 human primary solid tumor non specific karyotypic rearrangements.* Anticancer Res, 1992. **12**(3): p. 997-1004.
221. Ransom, D.T., et al., *Cytogenetic and loss of heterozygosity studies in ependymomas, pilocytic astrocytomas, and oligodendrogliomas.* Genes Chromosomes Cancer, 1992. **5**(4): p. 348-56.
222. Thiel, G., et al., *Karyotypes in 90 human gliomas.* Cancer Genet Cytogenet, 1992. **58**(2): p. 109-20.
223. Miyakawa, A., et al., *Multiple deleted regions on the long arm of chromosome 6 in astrocytic tumours.* Br J Cancer, 2000. **82**(3): p. 543-9.
224. King, H.O., et al., *RAD51 Is a Selective DNA Repair Target to Radiosensitize Glioma Stem Cells.* Stem Cell Reports, 2017. **8**(1): p. 125-139.
225. Ishii, N., et al., *Frequent co-alterations of TP53, p16/CDKN2A, p14ARF, PTEN tumor suppressor genes in human glioma cell lines.* Brain Pathol, 1999. **9**(3): p. 469-79.
226. Pore, N., et al., *PTEN mutation and epidermal growth factor receptor activation regulate vascular endothelial growth factor (VEGF) mRNA expression in human glioblastoma cells by transactivating the proximal VEGF promoter.* Cancer Res, 2003. **63**(1): p. 236-41.
227. Ermoian, R.P., et al., *Dysregulation of PTEN and protein kinase B is associated with glioma histology and patient survival.* Clin Cancer Res, 2002. **8**(5): p. 1100-6.
228. Smith, J.S., et al., *PTEN mutation, EGFR amplification, and outcome in patients with anaplastic astrocytoma and glioblastoma multiforme.* J Natl Cancer Inst, 2001. **93**(16): p. 1246-56.
229. Li, J., et al., *PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer.* Science, 1997. **275**(5308): p. 1943-7.
230. Kwon, C.H., et al., *Pten haploinsufficiency accelerates formation of high-grade astrocytomas.* Cancer Res, 2008. **68**(9): p. 3286-94.
231. Xiao, A., et al., *Astrocyte inactivation of the pRb pathway predisposes mice to malignant astrocytoma development that is accelerated by PTEN mutation.* Cancer Cell, 2002. **1**(2): p. 157-68.
232. Chalhoub, N. and S.J. Baker, *PTEN and the PI3-kinase pathway in cancer.* Annu Rev Pathol, 2009. **4**: p. 127-50.
233. Hayashi, Y., et al., *Association of EGFR gene amplification and CDKN2 (p16/MTS1) gene deletion in glioblastoma multiforme.* Brain Pathol, 1997. **7**(3): p. 871-5.
234. An, Z., et al., *Epidermal growth factor receptor and EGFRvIII in glioblastoma: signaling pathways and targeted therapies.* Oncogene, 2018. **37**(12): p. 1561-1575.
235. Lai, T., et al., *SOX5 controls the sequential generation of distinct corticofugal neuron subtypes.* Neuron, 2008. **57**(2): p. 232-47.
236. Lin, D., et al., *ASAP1, a gene at 8q24, is associated with prostate cancer metastasis.* Cancer Res, 2008. **68**(11): p. 4352-9.
237. Levine, A.J., J. Momand, and C.A. Finlay, *The p53 tumour suppressor gene.* Nature, 1991. **351**(6326): p. 453-6.
238. Hollstein, M., et al., *p53 mutations in human cancers.* Science, 1991. **253**(5015): p. 49-53.
239. Malkin, D., et al., *Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms.* Science, 1990. **250**(4985): p. 1233-8.
240. Srivastava, S., et al., *Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li-Fraumeni syndrome.* Nature, 1990. **348**(6303): p. 747-9.

241. Ziegler, A., et al., *Mutation hotspots due to sunlight in the p53 gene of nonmelanoma skin cancers.* Proc Natl Acad Sci U S A, 1993. **90**(9): p. 4216-20.
242. Rodrigues, N.R., et al., *p53 mutations in colorectal cancer.* Proc Natl Acad Sci U S A, 1990. **87**(19): p. 7555-9.
243. Momand, J., et al., *The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation.* Cell, 1992. **69**(7): p. 1237-45.
244. Oliner, J.D., et al., *Oncoprotein MDM2 conceals the activation domain of tumour suppressor p53.* Nature, 1993. **362**(6423): p. 857-60.
245. Chen, Z., et al., *Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis.* Nature, 2005. **436**(7051): p. 725-30.
246. Sigal, A. and V. Rotter, *Oncogenic mutations of the p53 tumor suppressor: the demons of the guardian of the genome.* Cancer Res, 2000. **60**(24): p. 6788-93.
247. Lang, G.A., et al., *Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome.* Cell, 2004. **119**(6): p. 861-72.
248. Liu, D.P., H. Song, and Y. Xu, *A common gain of function of p53 cancer mutants in inducing genetic instability.* Oncogene, 2010. **29**(7): p. 949-56.
249. Dittmer, D., et al., *Gain of function mutations in p53.* Nat Genet, 1993. **4**(1): p. 42-6.
250. Iggo, R., et al., *Increased expression of mutant forms of p53 oncogene in primary lung cancer.* Lancet, 1990. **335**(8691): p. 675-9.
251. Song, H., M. Hollstein, and Y. Xu, *p53 gain-of-function cancer mutants induce genetic instability by inactivating ATM.* Nat Cell Biol, 2007. **9**(5): p. 573-80.
252. Olivier, M., M. Hollstein, and P. Hainaut, *TP53 mutations in human cancers: origins, consequences, and clinical use.* Cold Spring Harb Perspect Biol, 2010. **2**(1): p. a001008.
253. Bieging, K.T., S.S. Mello, and L.D. Attardi, *Unravelling mechanisms of p53-mediated tumour suppression.* Nat Rev Cancer, 2014. **14**(5): p. 359-70.
254. Adesina, A.M., J. Nalbantoglu, and W.K. Cavenee, *p53 gene mutation and mdm2 gene amplification are uncommon in medulloblastoma.* Cancer Res, 1994. **54**(21): p. 5649-51.
255. Saylors, R.L., 3rd, et al., *Infrequent p53 gene mutations in medulloblastomas.* Cancer Res, 1991. **51**(17): p. 4721-3.
256. Packer, R.J. and G. Vezina, *Management of and prognosis with medulloblastoma: therapy at a crossroads.* Arch Neurol, 2008. **65**(11): p. 1419-24.
257. Thompson, M.C., et al., *Genomics identifies medulloblastoma subgroups that are enriched for specific genetic alterations.* J Clin Oncol, 2006. **24**(12): p. 1924-31.
258. Kool, M., et al., *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, 2012. **123**(4): p. 473-84.
259. Gajjar, A.J. and G.W. Robinson, *Medulloblastoma-translating discoveries from the bench to the bedside.* Nat Rev Clin Oncol, 2014. **11**(12): p. 714-22.
260. Jones, D.T., et al., *Dissecting the genomic complexity underlying medulloblastoma.* Nature, 2012. **488**(7409): p. 100-5.
261. Zhukova, N., et al., *Subgroup-specific prognostic implications of TP53 mutation in medulloblastoma.* J Clin Oncol, 2013. **31**(23): p. 2927-35.
262. Hart, M.N. and K.M. Earle, *Primitive neuroectodermal tumors of the brain in children.* Cancer, 1973. **32**(4): p. 890-7.
263. Shingu, T., et al., *Qki deficiency maintains stemness of glioma stem cells in suboptimal environment by downregulating endolysosomal degradation.* Nat Genet, 2017. **49**(1): p. 75-86.

264. Zuckermann, M., et al., *Somatic CRISPR/Cas9-mediated tumour suppressor disruption enables versatile brain tumour modelling.* Nat Commun, 2015. **6**: p. 7391.
265. Chow, R.D., et al., *AAV-mediated direct in vivo CRISPR screen identifies functional suppressors in glioblastoma.* Nat Neurosci, 2017. **20**(10): p. 1329-1341.
266. Chavez, A., et al., *Comparison of Cas9 activators in multiple species.* Nat Methods, 2016. **13**(7): p. 563-567.
267. Blakely, C.M., et al., *Evolution and clinical impact of co-occurring genetic alterations in advanced-stage EGFR-mutant lung cancers.* Nat Genet, 2017. **49**(12): p. 1693-1704.
268. Morrissy, A.S., et al., *Divergent clonal selection dominates medulloblastoma at recurrence.* Nature, 2016. **529**(7586): p. 351-7.
269. Chapeau, E.A., et al., *Resistance mechanisms to TP53-MDM2 inhibition identified by in vivo piggyBac transposon mutagenesis screen in an Arf(-/-) mouse model.* Proc Natl Acad Sci U S A, 2017. **114**(12): p. 3151-3156.
270. Fan, P.D., et al., *YES1 amplification is a mechanism of acquired resistance to EGFR inhibitors identified by transposon mutagenesis and clinical genomics.* Proc Natl Acad Sci U S A, 2018. **115**(26): p. E6030-E6038.
271. Guo, Y., et al., *Comprehensive Ex Vivo Transposon Mutagenesis Identifies Genes That Promote Growth Factor Independence and Leukemogenesis.* Cancer Res, 2016. **76**(4): p. 773-86.
272. Molyneux, S.D., et al., *Human somatic cell mutagenesis creates genetically tractable sarcomas.* Nat Genet, 2014. **46**(9): p. 964-72.
273. Kodama, T., et al., *Transposon mutagenesis identifies genes and cellular processes driving epithelial-mesenchymal transition in hepatocellular carcinoma.* Proc Natl Acad Sci U S A, 2016. **113**(24): p. E3384-93.
274. Chen, H.J., et al., *A recellularized human colon model identifies cancer driver genes.* Nat Biotechnol, 2016. **34**(8): p. 845-51.
275. Uhm, J.H., et al., *Phase II evaluation of gefitinib in patients with newly diagnosed Grade 4 astrocytoma: Mayo/North Central Cancer Treatment Group Study N0074.* Int J Radiat Oncol Biol Phys, 2011. **80**(2): p. 347-53.
276. Reardon, D.A., et al., *Phase I/randomized phase II study of afatinib, an irreversible ErbB family blocker, with or without protracted temozolomide in adults with recurrent glioblastoma.* Neuro Oncol, 2015. **17**(3): p. 430-9.
277. Reardon, D.A., et al., *A phase I/II trial of pazopanib in combination with lapatinib in adult patients with relapsed malignant glioma.* Clin Cancer Res, 2013. **19**(4): p. 900-8.
278. Lassman, A.B., et al., *Molecular study of malignant gliomas treated with epidermal growth factor receptor inhibitors: tissue analysis from North American Brain Tumor Consortium Trials 01-03 and 00-01.* Clin Cancer Res, 2005. **11**(21): p. 7841-50.
279. Martens, T., et al., *Inhibition of glioblastoma growth in a highly invasive nude mouse model can be achieved by targeting epidermal growth factor receptor but not vascular endothelial growth factor receptor-2.* Clin Cancer Res, 2008. **14**(17): p. 5447-58.
280. Neyns, B., et al., *Stratified phase II trial of cetuximab in patients with recurrent high-grade glioma.* Ann Oncol, 2009. **20**(9): p. 1596-603.
281. Hasselbalch, B., et al., *Cetuximab, bevacizumab, and irinotecan for patients with primary glioblastoma and progression after radiation therapy and temozolomide: a phase II trial.* Neuro Oncol, 2010. **12**(5): p. 508-16.
282. Talavera, A., et al., *Nimotuzumab, an antitumor antibody that targets the epidermal growth factor receptor, blocks ligand binding while permitting the active receptor conformation.* Cancer Res, 2009. **69**(14): p. 5851-9.

283. Westphal, M., et al., *A randomised, open label phase III trial with nimotuzumab, an anti-epidermal growth factor receptor monoclonal antibody in the treatment of newly diagnosed adult glioblastoma.* Eur J Cancer, 2015. **51**(4): p. 522-32.
284. Bode, U., et al., *Nimotuzumab treatment of malignant gliomas.* Expert Opin Biol Ther, 2012. **12**(12): p. 1649-59.
285. Lim, W.A. and C.H. June, *The Principles of Engineering Immune Cells to Treat Cancer.* Cell, 2017. **168**(4): p. 724-740.
286. Morgan, R.A., et al., *Recognition of glioma stem cells by genetically modified T cells targeting EGFRvIII and development of adoptive cell therapy for glioma.* Hum Gene Ther, 2012. **23**(10): p. 1043-53.
287. Swartz, A.M., Q.J. Li, and J.H. Sampson, *Rindopepimut: a promising immunotherapeutic for the treatment of glioblastoma multiforme.* Immunotherapy, 2014. **6**(6): p. 679-90.
288. Malkki, H., *Trial Watch: Glioblastoma vaccine therapy disappointment in Phase III trial.* Nat Rev Neurol, 2016. **12**(4): p. 190.
289. Wikstrand, C.J., et al., *The class III variant of the epidermal growth factor receptor (EGFRvIII): characterization and utilization as an immunotherapeutic target.* J Neurovirol, 1998. **4**(2): p. 148-58.
290. Cho, J., et al., *Glioblastoma-derived epidermal growth factor receptor carboxyl-terminal deletion mutants are transforming and are sensitive to EGFR-directed therapies.* Cancer Res, 2011. **71**(24): p. 7587-96.
291. Patel, A.P., et al., *Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma.* Science, 2014. **344**(6190): p. 1396-401.
292. Snuderl, M., et al., *Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma.* Cancer Cell, 2011. **20**(6): p. 810-7.
293. Lee, J.K., et al., *Pharmacogenomic landscape of patient-derived tumor cells informs precision oncology therapy.* Nat Genet, 2018. **50**(10): p. 1399-1411.
294. Weller, M., et al., *Molecular classification of diffuse cerebral WHO grade II/III gliomas using genome- and transcriptome-wide profiling improves stratification of prognostically distinct patient groups.* Acta Neuropathol, 2015. **129**(5): p. 679-93.
295. Sturm, D., et al., *New Brain Tumor Entities Emerge from Molecular Classification of CNS-PNETs.* Cell, 2016. **164**(5): p. 1060-1072.
296. Guo, X.E., et al., *Targeting tumor suppressor networks for cancer therapeutics.* Curr Drug Targets, 2014. **15**(1): p. 2-16.
297. Issaeva, N., et al., *Small molecule RITA binds to p53, blocks p53-HDM-2 interaction and activates p53 function in tumors.* Nat Med, 2004. **10**(12): p. 1321-8.
298. Lee, Y.R., M. Chen, and P.P. Pandolfi, *The functions and regulation of the PTEN tumour suppressor: new modes and prospects.* Nat Rev Mol Cell Biol, 2018. **19**(9): p. 547-562.

Supplementary Tables

#Gene	Indels	SNVs	Tot Muts	Covd Bps	Muts pMbp	P-value LRT	P-value CT	FDR LRT
Sub1	3	3	6	25530	235.02	1.11E-16	4.67E-18	2.27E-12
Trp53	0	5	5	71696	69.74	1.13E-12	8.13E-14	7.75E-09
Ces1c	0	4	4	95152	42.04	7.85E-12	2.01E-10	4.02E-08
Tead2	4	0	4	74211	53.9	2.80E-11	2.10E-09	1.15E-07
Nt5c2	0	4	4	108022	37.03	1.95E-10	1.40E-08	6.66E-07
Apol7a	0	3	3	46603	64.37	6.04E-10	1.18E-08	1.77E-06
Gm10696	0	4	4	44400	90.09	8.90E-10	7.98E-11	2.02E-06
Tcp10a	0	4	4	325160	12.3	8.80E-10	9.10E-08	2.02E-06
Tmsb15l	0	3	3	26472	113.33	1.52E-09	4.70E-08	3.11E-06
Mrpl47	0	3	3	48840	61.43	2.05E-09	2.77E-09	3.82E-06
Pabpc4	0	3	3	114225	26.26	7.28E-09	1.25E-07	1.24E-05
Skint6	0	4	4	231575	17.27	9.29E-09	3.71E-07	1.46E-05
1110059E24Rik	2	1	3	25361	118.29	1.57E-08	3.32E-08	2.30E-05
Uimc1	3	0	3	140324	21.38	1.39E-07	3.94E-06	0.00019017
Atp5e	0	2	2	11546	173.22	1.98E-07	3.35E-06	0.00025286
Itga6	3	0	3	174365	17.21	2.70E-07	8.65E-06	0.00032504
Nckap5	0	2	2	263359	7.59	6.15E-07	2.15E-05	0.00069978
Sypl2	0	2	2	47299	42.28	6.65E-07	3.96E-06	0.00071637
Mrgprb5	0	2	2	39627	50.47	7.98E-07	2.81E-06	0.00081766
Gpr85	0	2	2	44395	45.05	1.03E-06	1.15E-05	0.00100761
Olfr955	2	0	2	26521	75.41	1.26E-06	1.35E-05	0.00116895
H2-Q7	0	2	2	64637	30.94	1.37E-06	2.09E-06	0.00122437
Olfr514	2	0	2	35388	56.52	2.28E-06	2.29E-05	0.00188394
Olfr907	2	0	2	35520	56.31	2.30E-06	2.47E-05	0.00188394
4930433I11Rik	0	2	2	75964	26.33	3.06E-06	1.94E-05	0.00222737
Vmn2r111	0	2	2	88800	22.52	3.36E-06	2.08E-05	0.00222737
Olfr812	0	2	2	35431	56.45	3.14E-06	3.61E-05	0.00222737
Olfr1032	0	2	2	35484	56.36	3.11E-06	4.15E-05	0.00222737
Olfr1484	0	2	2	35508	56.33	3.37E-06	4.05E-05	0.00222737
Olfr213	0	2	2	39958	50.05	3.32E-06	4.25E-05	0.00222737
Vmn1r234	0	2	2	39790	50.26	4.27E-06	5.11E-05	0.00273133
Dagla	0	2	2	153765	13.01	6.46E-06	4.84E-05	0.00389161
Adipor1	2	0	2	58441	34.22	6.45E-06	0.00011924	0.00389161
Nfs1	2	0	2	72239	27.69	1.00E-05	0.00016899	0.0058524

Wsb1	0	2	2	70047	28.55	1.23E-05	0.00018179	0.00701854
Esyt2	0	2	2	140442	14.24	1.32E-05	0.00014772	0.00715512
Fntb	2	0	2	82846	24.14	1.33E-05	0.00025383	0.00715512
Gpbp1l1	0	2	2	78604	25.44	1.55E-05	0.00019417	0.00816374
Ccnb3	1	2	3	172134	17.43	1.68E-05	2.06E-05	0.00836098
Nlrp1b	0	3	3	226414	13.25	1.63E-05	2.99E-05	0.00836098
Olfr1076	0	2	2	35482	56.37	1.76E-05	1.10E-05	0.00836098
Olfr1286	1	1	2	31080	64.35	1.74E-05	1.07E-05	0.00836098
Cyp4a32	0	2	2	91020	21.97	1.86E-05	0.00024916	0.00866047
Ptp4a2	1	1	2	30811	64.91	2.02E-05	2.10E-05	0.00918145
Rpl32	1	1	2	28860	69.3	2.13E-05	3.27E-05	0.0094831
Zfp598	2	0	2	107068	18.68	2.26E-05	0.00035395	0.00984101
Olfr1000	0	2	2	35520	56.31	2.73E-05	1.49E-05	0.01142181
Nom1	2	0	2	116822	17.12	2.70E-05	0.00036471	0.01142181
Caprin1	2	0	2	119599	16.72	2.84E-05	0.00036603	0.01162296
Apbb2	2	0	2	132178	15.13	3.49E-05	0.00049655	0.01400698
Tgs1	0	2	2	113059	17.69	3.66E-05	0.00043003	0.01441445
Kdm1b	2	0	2	147574	13.55	4.38E-05	0.00052031	0.01660815
Arid5b	2	0	2	149964	13.34	4.52E-05	0.00075984	0.01682885
Prex2	2	1	3	271570	11.05	5.20E-05	8.41E-05	0.01868531
Cul2	0	2	2	135008	14.81	5.82E-05	0.00062372	0.01988478
Fstl5	0	2	2	137289	14.57	5.72E-05	0.00062049	0.01988478
Plk4	0	2	2	141761	14.11	5.83E-05	0.00062697	0.01988478
Ppp4r4	0	2	2	159182	12.56	7.19E-05	0.00087433	0.02375467
Vmn2r96	0	2	2	84360	23.71	7.88E-05	4.80E-05	0.02547925
Aurkc	1	1	2	43207	46.29	8.34E-05	8.50E-05	0.02547925
Fbxo22	1	1	2	59719	33.49	8.09E-05	0.00013616	0.02547925
Baz1b	2	0	2	198592	10.07	8.07E-05	0.00088071	0.02547925
Impg2	0	2	2	177319	11.28	8.25E-05	0.00094519	0.02547925
Cxadr	1	1	2	63453	31.52	9.27E-05	0.00013583	0.02793116
Vmn2r88	1	1	2	82116	24.36	0.0001093	6.79E-05	0.03244239
Magee2	1	1	2	67273	29.73	0.00011598	0.0001904	0.03393455
Cd244	0	2	2	68544	29.18	0.00012714	0.00015919	0.03667443
BC003331	1	1	2	83050	24.08	0.0001344	0.00010416	0.03801668
Nek10	0	2	2	222035	9.01	0.0001355	0.00145896	0.03801668
Sap30bp	1	1	2	63825	31.34	0.00017558	0.00019716	0.04670197
Psat1	1	1	2	66414	30.11	0.00020339	0.00021631	0.05340473
Plch1	1	1	2	248379	8.05	0.00023522	0.00019936	0.05947572

Pes1	1	1	2	104269	19.18	0.00023501	0.00029769	0.05947572
Slfn9	0	2	2	111000	18.02	0.00025808	0.00020231	0.06368252
Muc2	0	2	2	375826	5.32	0.00025523	0.00214414	0.06368252
Sp140	0	2	2	123007	16.26	0.00031697	0.00026102	0.07728325
Itgam	1	1	2	212715	9.4	0.00038295	0.00042823	0.08912649
Ccar1	1	1	2	160908	12.43	0.00051532	0.00045236	0.11472068
4930595M18Rik	1	1	2	96686	20.69	0.00053353	0.00042076	0.11749746
Dgkb	1	1	2	158915	12.59	0.00067434	0.00075238	0.13950718
Pclo	2	0	2	607388	3.29	0.00076614	0.00780049	0.1508779
Pign	0	2	2	185089	10.81	0.00084595	0.00082038	0.15750831
Zfp106	0	2	2	262045	7.63	0.00087769	0.00099362	0.16079591
Tex16	1	1	2	134405	14.88	0.00099529	0.00080426	0.16535887
Nf1	0	2	2	598238	3.34	0.00104908	0.01126867	0.16535887

Supplementary Table 1. Significantly mutated genes in *EGFRvIII*-only tumors.

Gene name	Source	Base Mean	log_2FoldChange	lfcSE	stat	p value, adjusted
Hoxa3	ensembl_havana	278.619316	9.39986332	0.94367147	9.96094894	1.5194E-21
Hoxa5	ensembl_havana	477.364085	9.28232077	0.83501008	11.1164176	1.1646E-26
Hoxd8	ensembl_havana	667.194941	8.86533997	0.72585202	12.2137016	4.9868E-32
Hoxc8	ensembl_havana	379.165958	8.84546772	0.95090811	9.30212667	6.6276E-19
Gm12688	havana	285.799904	8.70303756	1.02854526	8.46150181	7.7193E-16
Hoxa7	ensembl_havana	745.345625	8.65253777	0.73188103	11.8223282	4.909E-30
Nkx3-2	ensembl_havana	144.096303	8.55650963	0.95135725	8.99400263	9.676E-18
Irx2	ensembl_havana	312.594203	8.42995929	0.81195854	10.3822533	2.4669E-23
Gm13394	havana	823.394729	8.28018082	0.48669292	17.0131524	5.5164E-62
Hoxa2	ensembl_havana	114.987993	8.27825959	0.95132149	8.70185281	1.1033E-16
Hoxc4	ensembl_havana	229.118566	8.26549812	0.8710818	9.48877377	1.2203E-19
En1	ensembl_havana	146.963741	8.24012842	1.02912461	8.00692971	2.7198E-14
Hoxc5	ensembl_havana	172.510409	8.17259608	0.93944569	8.69938108	1.1255E-16
Hoxd9	ensembl_havana	75.5440359	7.80131307	0.93853964	8.31218285	2.5251E-15
Foxd3	ensembl_havana	78.5205631	7.75481063	0.9671947	8.01783819	2.5036E-14
Hoxc9	ensembl_havana	133.809793	7.74510984	0.96574263	8.01984877	2.4678E-14
Hoxc6	ensembl_havana	216.077503	7.72351624	0.8745156	8.83176501	3.7549E-17
Hoxa1	ensembl_havana	67.6822943	7.71842263	0.92145725	8.37632203	1.5178E-15
Runx3	ensembl_havana	277.411973	7.67767557	0.75117855	10.2208397	1.2266E-22
Hoxd3os1	ensembl_havana	60.5469758	7.51684258	0.94188607	7.98062828	3.3212E-14
Tmem26	ensembl	221.928053	7.37459785	0.76080373	9.69316737	1.9311E-20
Irx5	ensembl_havana	247.156103	7.33399821	0.75691577	9.68931883	2.0002E-20
Hoxc10	ensembl_havana	133.163785	7.30981784	0.93750462	7.79710057	1.3291E-13
Hoxb3os	havana	65.390674	7.29297236	1.03441067	7.05036461	2.5051E-11
Hoxb3	ensembl_havana	114.264545	7.19355565	0.92350521	7.78940454	1.4014E-13
Top2a	ensembl_havana	1390.25087	7.18853792	0.8100619	8.87406	2.6651E-17
Hoxb4	ensembl_havana	65.0687333	7.17614058	0.93018682	7.71473042	2.4188E-13
Hotairm1	havana	44.2520468	7.17170634	0.92728551	7.73408646	2.1046E-13
Foxa3	ensembl_havana	54.1219821	7.03580637	1.04322064	6.74431284	1.8327E-10
Hoxa6	ensembl_havana	41.1412123	6.99500031	0.95673845	7.31129839	4.2691E-12
2810417H13Rik	ensembl_havana	299.922819	6.92903069	0.87338938	7.93349546	4.7544E-14
Pnlip	ensembl	2059.06639	6.92088489	0.58750725	11.7800842	7.8121E-30
Mmp3	ensembl	98.1334985	6.90273503	1.04076939	6.63233866	3.7361E-10
Hoxb2	ensembl_havana	85.6186247	6.89935571	0.92368175	7.46940786	1.4142E-12
Gm42909	havana	29.7030621	6.87832966	0.84601209	8.13029711	1.0546E-14
Rps2-ps10	havana	130.788751	6.86533583	1.01396852	6.77075834	1.553E-10

Hoxd4	ensembl_havana	34.4287872	6.85429504	0.92903854	7.37783711	2.6976E-12
Hoxa4	ensembl_havana	97.2925925	6.84629221	0.85847054	7.97498797	3.463E-14
Pbk	ensembl_havana	331.443199	6.84535611	0.78498875	8.72032386	9.5061E-17
C1ql2	ensembl_havana	1167.04811	6.82834509	0.58451716	11.6820267	2.3921E-29
Meox1	ensembl_havana	1310.01564	6.79808426	0.62127091	10.942222	7.4602E-26
Hoxaas2	ensembl_havana	30.2070096	6.74409868	0.90958342	7.41449164	2.0965E-12
Cdk1	ensembl_havana	361.301216	6.63727972	0.80466839	8.24846583	4.1736E-15
Mki67	ensembl	888.742681	6.62808981	0.77817333	8.51749805	4.9398E-16
H19	ensembl_havana	792.555329	6.61084878	0.82957249	7.96898263	3.625E-14
Bricd5	ensembl	539.499634	6.51187305	0.59424866	10.958162	6.3251E-26
Cdkn2a	ensembl_havana	429.49757	6.501445	0.76049482	8.5489668	3.837E-16
Fam64a	ensembl_havana	212.381259	6.47323083	0.88457545	7.31789565	4.084E-12
Gm27477	ensembl	32.4524792	6.46207508	0.88958325	7.26416005	5.9109E-12
Itih5l-ps	havana	99.5372195	6.44883644	1.09602192	5.88385719	3.105E-08
Otos	ensembl_havana	123.706083	6.44324339	0.82239049	7.83477374	9.9976E-14
Gsx1	ensembl_havana	386.554017	6.37712526	0.61098899	10.4373816	1.4072E-23
Ube2c	ensembl_havana	337.358203	6.3525281	0.77740051	8.17149975	7.6281E-15
Irx1	ensembl_havana	628.108931	6.32608261	0.59084659	10.7068107	8.6623E-25
Kif18b	ensembl_havana	245.240142	6.2810011	0.85385742	7.35603032	3.1428E-12
5730596B20Rik	ensembl	19.8677121	6.27449212	0.88617854	7.08039278	2.051E-11
Gm10260	ensembl	201.322175	6.26873541	0.8109035	7.73055663	2.1545E-13
Ccdc178	ensembl	25.6149288	6.21099515	1.02465706	6.06153552	1.1442E-08
Mroh3	havana	440.162975	6.19885584	0.59786151	10.3683809	2.813E-23
Stk19-ps1	havana	20.2168019	6.19030885	0.92981673	6.65755803	3.1875E-10
Pdx1	ensembl_havana	26.3671588	6.18838281	1.04231393	5.93715831	2.3057E-08
Fam89a	ensembl	759.103403	6.17195203	0.54201548	11.3870401	6.4171E-28
Hmga2	ensembl_havana	149.985294	6.1681856	1.01161674	6.09735424	9.2756E-09
Col20a1	ensembl_havana	5750.03172	6.14368683	0.69633696	8.8228648	4.0346E-17
Gdf3	ensembl_havana	23.2854993	6.11767411	1.01823222	6.0081325	1.5421E-08
Hoxaas3	havana	17.6441879	6.05863275	0.91471836	6.62349532	3.9402E-10
A930009A15Rik	ensembl_havana	331.153537	6.05269226	0.54870978	11.0307715	2.9056E-26
Piwil4	ensembl_havana	46.7599306	6.02970066	0.84766109	7.11333894	1.6481E-11
Nxf3	ensembl_havana	180.268561	6.0102255	0.91978723	6.53436503	6.8953E-10
Neil3	ensembl_havana	95.2373242	6.00804275	0.88227561	6.80971195	1.2041E-10
Cxcl13	ensembl_havana	37.829686	6.00504176	1.15514146	5.19853366	1.137E-06
Lbx1	ensembl_havana	162.320467	5.99610607	1.11449172	5.38012615	4.5479E-07
Troap	ensembl	161.105017	5.99042085	0.80792845	7.4145438	2.0965E-12
Gata6	ensembl	256.76976	5.97655969	0.6404476	9.33184803	5.0816E-19

Ticrr	ensembl_havana	60.680563	5.9740599	0.86307043	6.92186836	5.8556E-11
Ctse	ensembl_havana	32.7112022	5.97078066	0.8723589	6.84440852	9.6728E-11
Gm8909	ensembl_havana	20.1871416	5.96865792	1.01076463	5.90509178	2.764E-08
Serpina3f	ensembl_havana	30.8922472	5.96081484	1.0205813	5.84060756	3.9549E-08
Cd300lf	ensembl_havana	86.870023	5.9564507	0.67169845	8.86774514	2.8071E-17
Gm20554	havana	14.4395288	5.91830037	0.864223	6.84811715	9.4456E-11
Hils1	ensembl_havana	15.2788491	5.89221197	0.90817515	6.48796874	9.1478E-10
Cnpy1	ensembl_havana	53.7763457	5.89145194	0.94360736	6.24354175	3.9821E-09
C4b	ensembl_havana	13627.4982	5.87681987	0.43518764	13.5041055	4.5814E-39
Gsc	ensembl_havana	18.6462006	5.87503029	1.01020496	5.81568149	4.5405E-08
Prc1	ensembl_havana	567.625033	5.83958908	0.72470462	8.05788856	1.8384E-14
Nr5a2	ensembl_havana	38.6359898	5.79293143	0.83927497	6.90230456	6.6476E-11
Mmp13	ensembl	138.986174	5.78830511	1.03856267	5.57338064	1.6772E-07
Gm8210	ensembl_havana	43.3266638	5.76545434	0.90575031	6.36539043	1.9212E-09
Klra2	ensembl_havana	117.665062	5.76391779	0.64093698	8.99295555	9.7351E-18
Cenpf	ensembl_havana	288.733202	5.759621	0.71997139	7.99979151	2.8709E-14
Snx22	ensembl	4617.46451	5.7505615	0.51755015	11.1111194	1.23E-26
Frmd7	ensembl_havana	415.987103	5.74870418	0.6196834	9.27684067	8.3029E-19
Depdc1a	ensembl_havana	54.4349378	5.74707166	0.97560575	5.89077265	2.9896E-08
Epyc	ensembl_havana	37.6134225	5.71493396	0.85610787	6.67548352	2.834E-10
Prokr1	ensembl_havana	75.9953108	5.70832169	0.81286163	7.02250111	3.0251E-11
Irx3os	ensembl_havana	36.4313808	5.70665201	0.85243049	6.69456585	2.521E-10
Gm27861	ensembl	13.9635608	5.69710323	0.94453899	6.03162316	1.3546E-08
Mis18bp1	ensembl_havana	133.236281	5.69460596	0.77421006	7.35537583	3.1542E-12
Slamf7	ensembl_havana	21.109607	5.68201926	0.98645796	5.76002172	6.1547E-08
5033426O07Rik	havana	152.51473	5.66820211	0.7022111	8.0719347	1.6421E-14

Supplementary Table 2. Top 100 most upregulated genes in brain tumors.

gene_name	source	Base Mean	log2FoldChange	lfcSE	stat	p value, adjusted
9130024F11Rik	havana	328.610907	-10.401713	0.50638195	-20.541239	1.9728E-90
Tmem215	ensembl_havana	285.466843	-9.7092875	0.4679682	-20.747751	3.0243E-92
Igfn1	ensembl_havana	489.933872	-9.2893857	0.50674272	-18.331562	5.4547E-72
Gm37717	havana	48.5168346	-8.8808101	0.61334563	-14.479292	7.7113E-45
Ovol2	ensembl_havana	62.2791155	-8.7457129	0.60037269	-14.56714	2.2767E-45
Gm28928	havana	32.4046988	-8.578732	0.61395873	-13.972816	8.9622E-42
Gm38073	havana	48.7213996	-8.0716613	0.6329517	-12.752413	7.3355E-35
Mal2	ensembl	2227.29391	-7.8094117	0.88107925	-8.8634612	2.9032E-17
Gm765	ensembl_havana	220.771733	-7.4360853	0.79855128	-9.3119696	6.066E-19
Satb2	ensembl_havana	645.087444	-7.3525292	0.22088118	-33.28726	1.39E-238
Cldn22	ensembl	15.039166	-7.3236912	0.73933194	-9.9058227	2.6044E-21
Olfr464	ensembl_havana	16.2173934	-7.3074347	0.66273018	-11.026259	3.0412E-26
Myh2	ensembl_havana	12.0967219	-7.2398328	0.610278	-11.863172	3.1368E-30
Gpr88	ensembl_havana	1164.46617	-7.1744123	0.60730169	-11.813589	5.3547E-30
Tbata	ensembl_havana	21.5448992	-7.0267155	0.59627953	-11.784264	7.4843E-30
4921539H07Rik	havana	25.7827694	-6.8266217	0.5523346	-12.359576	8.93E-33
Alox12b	ensembl_havana	57.5131834	-6.6584513	0.38624833	-17.238783	1.2302E-63
Neurod6	ensembl_havana	507.595925	-6.61052	1.07181118	-6.1676162	6.1887E-09
Chat	ensembl	12.0514092	-6.5893376	0.84440625	-7.803516	1.2655E-13
Gm13446	havana	53.856653	-6.5694146	0.3839547	-17.109869	1.0924E-62
Panct2	havana	32.5265576	-6.5311441	0.45893453	-14.231102	2.5612E-43
Gm4081	havana	7.09886203	-6.4737879	0.6391159	-10.129286	2.9581E-22
Rspo2	ensembl	92.4052375	-6.4705881	0.47357406	-13.663308	5.6557E-40
Slc26a4	ensembl	14.6712495	-6.4697873	0.61540997	-10.512971	6.5159E-24
Gm13306	ensembl_havana	22.383638	-6.4184695	0.53786178	-11.933306	1.4153E-30
Gm42954	havana	16.4668927	-6.4111641	0.5759231	-11.131979	1.0019E-26
Scn4b	ensembl_havana	2353.5171	-6.4110256	0.84916408	-7.5498078	7.955E-13
Vmn2r84	ensembl_havana	33.138937	-6.4053594	0.46568193	-13.754795	1.6501E-40
Tmco5	ensembl_havana	9.40040904	-6.3614254	0.62932755	-10.10829	3.6214E-22
Gm17916	havana	7.76238383	-6.3460325	0.63746995	-9.9550299	1.6035E-21
Gm12300	havana	6.55185734	-6.3329083	0.66560871	-9.5144613	9.6415E-20
Gm14340	ensembl_havana	9.25109693	-6.3235291	0.64077561	-9.8685546	3.7056E-21
Robo3	ensembl_havana	628.23625	-6.3179646	0.93534211	-6.7547099	1.7156E-10
Gm13601	havana	40.9837028	-6.2908201	0.65930069	-9.5416555	7.5041E-20
Gm20752	havana	9.01796956	-6.2842346	0.64209679	-9.7870518	7.9283E-21
Sstr2	ensembl_havana	187.069725	-6.269868	0.40205665	-15.594489	5.0492E-52
Dmrt2	ensembl	6.90271924	-6.201255	0.62399487	-9.9379905	1.8921E-21

Gm14015	havana	5.88461944	-6.1998208	0.65640023	-9.4451837	1.7973E-19
Prss16	ensembl_havana	23.4789309	-6.1670186	0.52173669	-11.820174	4.9852E-30
Kcnk4	ensembl	211.144494	-6.1595356	0.79246683	-7.77261	1.5802E-13
Gm16339	havana	17.036485	-6.0432789	0.53023358	-11.39739	5.763E-28
Mas1	ensembl_havana	156.725968	-6.0336456	0.85041697	-7.0949262	1.8589E-11
Caln1	ensembl_havana	1372.99913	-6.0192308	0.7847716	-7.6700416	3.3592E-13
Gm14317	havana	8.75801049	-6.0173202	0.61918142	-9.7181859	1.5226E-20
Gm12930	havana	5.07236216	-5.9976188	0.65696903	-9.1292261	3.0133E-18
Clcnka	ensembl_havana	24.049661	-5.9682864	0.91400711	-6.5298031	7.0956E-10
Olfr316	ensembl_havana	7.28392461	-5.9552591	0.66924662	-8.8984522	2.1643E-17
1700047F07Rik	havana	13.7723385	-5.9351954	0.55905442	-10.616489	2.219E-24
Tfap2d	ensembl_havana	6.29426849	-5.8654154	0.77347454	-7.5832042	6.274E-13
Gm12576	havana	6.44866974	-5.8253243	0.63764919	-9.1356257	2.8456E-18
Agmat	ensembl_havana	25.6188271	-5.8226655	0.4446211	-13.095792	9.2058E-37
9130227L01Rik	havana	4.37178497	-5.8022518	0.65418659	-8.869414	2.7697E-17
Gm12371	havana	63.7687672	-5.7843243	0.67484552	-8.5713309	3.211E-16
6530403H02Rik	havana	6.82852983	-5.7390602	0.76140799	-7.5374309	8.7002E-13
Gm15870	havana	7.58701115	-5.7245204	0.68942882	-8.3032798	2.6971E-15
RP23-458C8.3	havana	5.03258244	-5.7092276	0.67166408	-8.5001234	5.6653E-16
1700007P06Rik	havana	4.15958154	-5.709198	0.6750915	-8.4569246	7.9985E-16
Gm10334	ensembl_havana	4.1919787	-5.7020411	0.68610702	-8.3107168	2.5535E-15
A330070K13Rik	ensembl_havana	4.20444028	-5.6853874	0.70096219	-8.1108332	1.2189E-14
Chn1os1	havana	12.1718984	-5.5925965	0.52094929	-10.735395	6.4597E-25
Npy2r	ensembl_havana	17.6308782	-5.5684039	0.76021047	-7.3248188	3.9001E-12
Efcab6	ensembl_havana	54.8492386	-5.5545705	0.57619614	-9.6400689	3.1094E-20
2900064K03Rik	havana	6.39360551	-5.5545432	0.64204537	-8.6513251	1.6691E-16
Gm5912	havana	7.64709081	-5.5495569	0.64139353	-8.6523431	1.6566E-16
Tex40	ensembl	121.129807	-5.5421331	0.421603	-13.145383	5.0099E-37
Kcnk9	ensembl	7.46334135	-5.5179004	0.95170339	-5.7979203	5.0084E-08
RP24-134N2.1	havana	186.219208	-5.5175771	0.44066699	-12.520968	1.3104E-33
Ascl5	havana	9.00003831	-5.5146192	0.6147465	-8.9705581	1.1653E-17
Allc	ensembl_havana	4.28148292	-5.4974355	0.66690624	-8.2431909	4.3331E-15
A830029E22Rik	havana	3.4916202	-5.4711358	0.67575957	-8.0962757	1.3613E-14
Gm15997	havana	4.2527376	-5.4666571	0.68135751	-8.0231846	2.4089E-14
Gm15645	havana	4.99026312	-5.4490303	0.66040474	-8.2510467	4.0933E-15
Theg	ensembl_havana	3.43825102	-5.4362416	0.68550607	-7.9302604	4.8707E-14
Rph3a	ensembl_havana	4564.25757	-5.4216625	0.84432074	-6.4213304	1.3708E-09
Sh3rf2	ensembl_havana	31.6797315	-5.4094324	0.79547436	-6.8002598	1.2791E-10

Neurod2	ensembl_havana	1052.92045	-5.3636111	1.09502475	-4.8981643	4.7975E-06
Cd6	ensembl_havana	69.9281626	-5.3084948	0.66018553	-8.0409136	2.0972E-14
RP23-16L19.5	havana	3.69455063	-5.2900458	0.66545594	-7.9495057	4.2145E-14
Myoc	ensembl_havana	288.396047	-5.2848198	0.73232726	-7.2164729	8.183E-12
Gm42730	havana	37.2603256	-5.2804433	0.71484637	-7.3868226	2.5378E-12
Gm9946	ensembl_havana	7.48787626	-5.2792114	0.60523416	-8.7225932	9.3311E-17
Glt8d2	ensembl_havana	162.263962	-5.2742957	0.51322057	-10.276859	7.0547E-23
Gm17751	havana	4.45235218	-5.2639116	0.68803747	-7.6506177	3.8533E-13
Hkdc1	ensembl_havana	148.950718	-5.2622023	0.31022343	-16.962621	1.2604E-61
A530021J07Rik	ensembl_havana	3.29985213	-5.2514687	0.78320743	-6.7050803	2.3577E-10
Btn2a2	ensembl_havana	14.3005286	-5.2460264	0.44356746	-11.826896	4.7089E-30
Mfsd4	ensembl_havana	779.63276	-5.2102415	0.58803911	-8.8603656	2.9755E-17
Gm27004	havana	24.9763285	-5.2062776	0.4665565	-11.158943	7.547E-27
Gm10421	havana	7.08083562	-5.1960701	0.61056108	-8.5103199	5.2151E-16
Gm11549	havana	497.922529	-5.1873423	0.99966602	-5.1890753	1.1906E-06
Nhlh1	ensembl_havana	5.89861678	-5.1863525	0.63813756	-8.1273268	1.0785E-14
Bglap3	ensembl_havana	16.2151169	-5.1818598	0.55687104	-9.3053137	6.445E-19
1700108N11Rik	havana	3.01402095	-5.1814133	0.7484718	-6.9226567	5.8296E-11
Myh1	ensembl_havana	8.14133133	-5.1789594	0.57716254	-8.9731385	1.1422E-17
AI593442	ensembl	3002.7825	-5.1777766	0.54452988	-9.5087098	1.0144E-19
Gm22389	ensembl	2.91647092	-5.1562054	0.73722438	-6.9940788	3.6272E-11
Klhdc7b	ensembl_havana	8.69089108	-5.1517311	0.66461751	-7.7514225	1.8521E-13
Gng13	ensembl_havana	149.35804	-5.1408605	0.77595323	-6.6252195	3.9E-10
1700048F04Rik	havana	2.74397198	-5.1333382	0.68657508	-7.4767325	1.3436E-12
Kctd16	ensembl	52.0300938	-5.115876	0.54053923	-9.4643936	1.5139E-19

Supplementary Table 3. Top 100 most downregulated genes in brain tumors.

Gene name	source	Base Mean	log_2FoldChange	lfcSE	stat	p value adjusted
Pnlip	ensembl	2122.17388	9.11722837	0.43335277	21.0388141	1.8435E-95
Gm13394	havana	530.838672	8.63230606	0.48070325	17.957661	8.4148E-70
Nkx3-2	ensembl_havana	115.220138	8.24714202	0.77302513	10.6686596	3.4546E-25
Nxf3	ensembl_havana	209.615669	7.73140566	0.82239877	9.4010423	8.9323E-20
Foxa3	ensembl_havana	45.0894673	7.12157996	0.90780186	7.84486164	4.4792E-14
Gsx1	ensembl_havana	378.301174	7.1044325	0.40660711	17.4724747	3.8949E-66
Fgb	ensembl_havana	52.847471	6.78044122	1.00767674	6.72878609	1.2559E-10
Cdkn2a	ensembl_havana	475.57514	6.77288936	0.3660268	18.5038075	4.8406E-74
Gm10260	ensembl	144.548889	6.7093184	0.77610038	8.64491059	7.0915E-17
Apoc4	ensembl_havana	22.7722698	6.5671452	0.81465657	8.06124377	8.3141E-15
Stk19-ps1	havana	20.1555172	6.41862304	0.81350134	7.89011982	3.1554E-14
Bricd5	ensembl	544.73407	6.11740057	0.38711281	15.8026303	3.1559E-54
Gm12128	havana	19.0077542	6.11294048	0.89607395	6.82191518	6.8078E-11
Btnl2	ensembl_havana	23.6375275	6.07211866	0.84961787	7.14688203	7.4651E-12
Postn	ensembl_havana	3843.55416	6.05630423	0.50470529	11.9996844	1.3122E-31
Otos	ensembl_havana	244.262635	6.0323201	0.63543501	9.49321328	3.8047E-20
Gbp2b	ensembl_havana	18.6069028	6.0221023	0.91695815	6.56747782	3.5774E-10
Rps2-ps10	havana	112.434467	5.97119624	1.00424348	5.94596464	1.6332E-08
Pdx1	ensembl_havana	29.4843967	5.95316832	1.04708172	5.68548588	7.2285E-08
H2-M2	ensembl_havana	37.4922996	5.95021181	0.74769094	7.95811677	1.867E-14
H2-Ea-ps	havana	19.9094435	5.86080867	0.98606384	5.9436402	1.6553E-08
Gm8909	ensembl_havana	26.5792426	5.80674115	0.95759619	6.06387243	8.1659E-09
Clec4n	ensembl_havana	150.582035	5.7803685	0.62332295	9.27347288	2.8696E-19
Gm42909	havana	27.4235865	5.68392074	0.69775212	8.14604581	4.2306E-15
3100002H09Rik	ensembl	56.3111606	5.60308853	0.58312398	9.60874315	1.296E-20
Adamdec1	ensembl	41.3453354	5.58852252	0.68718183	8.13252375	4.7185E-15
Gm14165	havana	52.3439979	5.56588245	0.75195394	7.40189279	1.2194E-12
Meox1	ensembl_havana	1939.20496	5.56393942	0.28089208	19.8081033	1.0009E-84
Adh6b	ensembl_havana	45.6285809	5.55052392	0.85697183	6.47690358	6.386E-10
Ccdc178	ensembl	36.106715	5.48208783	0.65693439	8.34495482	8.4737E-16
1010001N08Rik	ensembl	11.0013494	5.47045619	0.88761631	6.1630866	4.4974E-09
Tlx1	ensembl_havana	24.8802609	5.46577543	1.08598033	5.03303354	2.2454E-06
Cd5l	ensembl_havana	52.9462173	5.45953538	0.79925352	6.83079306	6.4181E-11
Gm18734	havana	13.1076946	5.45531311	0.96311392	5.6642449	8.1329E-08
Kng1	ensembl_havana	66.0618916	5.42865253	0.89808334	6.0447091	9.1512E-09
Xcr1	ensembl_havana	15.9916877	5.41971456	0.90728548	5.97354935	1.3915E-08
Gm34342	havana	27.6670873	5.41175914	0.6565954	8.24215207	1.953E-15

Gm25631	ensembl	9.10154128	5.39097758	0.82509907	6.53373366	4.4466E-10
BC023719	havana	70.7329181	5.36206439	0.7610698	7.0454305	1.4967E-11
Mroh3	havana	489.854356	5.34046957	0.32135636	16.6185277	6.9719E-60
Col20a1	ensembl_havana	2695.03076	5.33316667	0.38533503	13.8403369	9.1168E-42
Trav3-3	ensembl_havana	11.3550994	5.31321601	0.9551315	5.56281098	1.4199E-07
Rlbp1	ensembl_havana	3597.93164	5.29753416	0.29323101	18.0660773	1.2201E-70
5730559C18Rik	ensembl_havana	1911.49676	5.2931441	0.27807683	19.034826	2.659E-78
Il1f10	ensembl_havana	13.2409243	5.28041484	1.01543949	5.20012755	9.6826E-07
Gm6614	ensembl_havana	45.6973149	5.26837287	0.93593612	5.62898769	9.8768E-08
Mir6358	mirbase	11.2192421	5.25722025	0.81055016	6.48599004	6.0338E-10
Gata6	ensembl	176.101804	5.24649651	0.48058844	10.916818	2.5221E-26
1700040K01Rik	havana	11.7557732	5.22128656	0.85425119	6.11212089	6.1164E-09
Runx3	ensembl_havana	251.798295	5.21854155	0.43843487	11.9026609	4.1014E-31
B9d1os	ensembl_havana	243.343729	5.19377983	0.40624028	12.7849947	9.2608E-36
Rbpjl	ensembl_havana	607.428104	5.18717996	0.31858654	16.2818554	1.6374E-57
Gm42825	havana	8.35557643	5.17049351	0.87225947	5.92770117	1.8159E-08
A930009A15Rik	ensembl_havana	413.233455	5.16413175	0.32194221	16.0405552	7.5785E-56
Matn4	ensembl_havana	6785.71567	5.13279963	0.26327656	19.4958475	4.0378E-82
Gm8730	ensembl	226.505443	5.09169865	0.54121576	9.4078906	8.4049E-20
Gucy2f	ensembl_havana	130.06289	5.08456417	0.51070865	9.95589987	4.7049E-22
Itih5l-ps	havana	86.7494075	5.08340469	0.99821305	5.09250475	1.6701E-06
Mmp3	ensembl	61.8930553	5.07136885	1.02546752	4.94542126	3.4671E-06
Gsc	ensembl_havana	12.6765523	5.04171382	0.94600764	5.32946414	4.9592E-07
Gm9008	ensembl_havana	30.5747769	5.03763879	0.96295762	5.23142315	8.2502E-07
RP23-152B12.3	havana	6.92336909	5.01717497	0.83758841	5.99002432	1.2642E-08
Itgax	ensembl_havana	593.5313	4.95649333	0.43324351	11.4404331	8.2497E-29
Prok2	ensembl_havana	138.456647	4.93714982	1.02771532	4.80400528	6.8562E-06
Adam12	ensembl_havana	1621.46038	4.92051801	0.36997515	13.2995906	1.2563E-38
Rep15	ensembl_havana	355.802841	4.91749334	0.27925246	17.60949	3.7567E-67
Gm8210	ensembl_havana	29.8284576	4.89887213	0.95238947	5.14376974	1.2897E-06
Vmn1r181	ensembl_havana	19.3547533	4.89205671	0.84680795	5.7770557	4.3209E-08
Snx22	ensembl	6030.65755	4.88863783	0.27675016	17.6644445	1.4374E-67
Fbp1	ensembl_havana	30.5256944	4.88183094	0.59502323	8.20443753	2.6495E-15
Gm15698	ensembl_havana	11.4454052	4.85851875	0.95935653	5.0643516	1.9226E-06
Nr2e1	ensembl_havana	8.37847756	4.85740055	0.98643703	4.92418715	3.8431E-06
Klk13	ensembl_havana	9.92132172	4.83529834	0.92099329	5.25009073	7.4941E-07
Ascl1	ensembl_havana	1469.64274	4.82351761	0.30618842	15.7534294	6.7882E-54
RP23-52F4.1	havana	168.938605	4.81865162	0.40005942	12.0448397	7.6964E-32

C1ql2	ensembl_havana	1273.65558	4.8150795	0.37977152	12.6788854	3.5059E-35
Ms4a7	ensembl_havana	397.027474	4.8073631	0.42544817	11.299527	3.9501E-28
Sapcd2	ensembl_havana	1023.58856	4.79971641	0.32183065	14.913795	2.2666E-48
BC018473	havana	41.291218	4.79584984	0.91817359	5.22324959	8.5973E-07
Sox3	ensembl_havana	863.529978	4.7813043	0.32523831	14.7009258	5.0222E-47
C1rb	havana	11.8119049	4.76470104	0.8330362	5.71968067	5.9683E-08
6030408B16Rik	ensembl_havana	22.2362551	4.76239755	0.91761295	5.18998511	1.0198E-06
Tsga13	ensembl_havana	21.4489709	4.76229631	0.6750352	7.05488591	1.4053E-11
Fam89a	ensembl	762.382645	4.75309406	0.31276814	15.1968612	3.2975E-50
Fcgr4	ensembl_havana	162.69426	4.74833026	0.36604548	12.9719681	8.6963E-37
Gm3375	havana	54.8469014	4.73649036	0.8512257	5.56431786	1.4086E-07
Gm1821	havana	378.140291	4.72232109	0.60750262	7.77333453	7.7161E-14
Ccnd1	ensembl_havana	23307.8869	4.7020004	0.26341091	17.8504393	5.4555E-69
Gm42803	havana	7.01984644	4.69635779	0.79364053	5.91748736	1.9261E-08
Vcan	ensembl_havana	10739.2669	4.69224123	0.29780292	15.7561962	6.5193E-54
C6	ensembl_havana	75.9044713	4.6907965	0.67219108	6.97836763	2.378E-11
Gm43820	havana	42.8684932	4.68963662	0.43349553	10.818189	7.1606E-26
Aldh1a3	ensembl_havana	954.672108	4.68127071	0.69455846	6.73992327	1.1672E-10
Pbk	ensembl_havana	276.360959	4.67886584	0.50826086	9.20563873	5.2904E-19
Pdgfra	ensembl_havana	15744.0023	4.66121349	0.31608789	14.7465742	2.596E-47
Neu4	ensembl_havana	4267.26649	4.66057454	0.31606148	14.7457847	2.6198E-47
Emx2	ensembl_havana	17.4713707	4.65062	0.97832144	4.7536728	8.6871E-06
Kng2	ensembl_havana	288.760238	4.64849735	0.73669655	6.3099214	1.8264E-09
Epyc	ensembl_havana	116.767151	4.62464243	1.09643195	4.21790193	9.3512E-05
Timp1	ensembl_havana	383.271413	4.60615938	0.57541756	8.00489885	1.2957E-14

Supplementary Table 4. Top 100 genes most upregulated in spinal tumors.

Gene name	source	Base Mean	log_2FoldChange	lfcSE	stat	p value adjusted
Chat	ensembl	168.142906	-9.9307336	0.65125954	-15.248504	1.5189E-50
Slc6a5	ensembl_havana	2400.88165	-9.7705032	0.88078998	-11.092886	3.7737E-27
Htr2c	ensembl_havana	702.867929	-9.7401835	0.600382	-16.22331	4.173E-57
Lamp5	ensembl_havana	1936.54273	-9.4728419	0.53596459	-17.67438	1.2124E-67
Gm14204	havana	135.08327	-9.4648371	0.59884455	-15.805165	3.0418E-54
Otp	ensembl_havana	75.0771559	-9.4548965	0.62546163	-15.116669	1.1082E-49
Slc30a3	ensembl_havana	347.996674	-9.3048006	0.37857126	-24.578729	3.025E-130
Gm42756	havana	67.6884279	-9.1936669	0.67501628	-13.619919	1.7915E-40
Glra1	ensembl_havana	934.03217	-9.0497987	0.73393324	-12.330548	2.5131E-33
Slc32a1	ensembl_havana	1909.41572	-9.027482	0.86645191	-10.418907	4.599E-24
Slc18a3	havana	394.057024	-9.0052437	0.46015442	-19.570047	9.9846E-83
Lhx5	ensembl_havana	76.4710846	-8.9569354	0.61915764	-14.466325	1.4586E-45
Trhr	ensembl	95.6679787	-8.702169	0.58727331	-14.81792	9.2366E-48
Gm42500	havana	38.895394	-8.5643227	0.6284921	-13.626779	1.637E-40
Magel2	ensembl_havana	78.8094562	-8.4777144	0.57310014	-14.792728	1.326E-47
Nkx2-9	ensembl_havana	47.7728891	-8.454287	0.67849364	-12.460378	5.2192E-34
Pnoc	ensembl	89.509404	-8.4505798	0.53447059	-15.811122	2.7863E-54
Cck	ensembl	176.459164	-8.4388602	0.39394707	-21.421305	6.283E-99
Gpr101	ensembl_havana	101.673629	-8.427774	0.51099541	-16.492856	5.4362E-59
Gm42495	havana	32.8287183	-8.3427731	0.62429245	-13.363566	5.4225E-39
Pax2	ensembl_havana	354.690736	-8.2214517	0.26603278	-30.903904	1.101E-205
Tekt5	ensembl_havana	241.057451	-8.0587408	0.3095274	-26.03563	4.697E-146
Slc10a4	ensembl_havana	204.21927	-8.0306858	0.37517745	-21.405033	8.702E-99
Fam216b	ensembl_havana	63.3211298	-7.983239	0.66005441	-12.09482	4.2284E-32
Ermn	ensembl_havana	2122.74596	-7.9070057	0.45040187	-17.555446	9.472E-67
Neurod2	ensembl_havana	57.0146531	-7.8408933	0.5335539	-14.695597	5.4196E-47
Foxb1	ensembl	30.4023426	-7.7084397	0.62341113	-12.364938	1.6563E-33
Uts2	ensembl_havana	55.7462305	-7.6641837	0.61372302	-12.488017	3.7452E-34
Sst	ensembl	526.739301	-7.6345822	0.73598966	-10.373219	7.3104E-24
Neurod6	ensembl_havana	28.274478	-7.6194762	0.61708845	-12.347462	2.0423E-33
Aqp6	ensembl	60.0190457	-7.5293579	0.49845814	-15.105296	1.3096E-49
Mobp	ensembl_havana	27101.6308	-7.470311	0.66449279	-11.242125	7.4669E-28
RP23-53O7.3	havana	15.875006	-7.3812213	0.62134356	-11.879452	5.3783E-31
Oprk1	ensembl_havana	71.994221	-7.3777862	0.43703755	-16.881355	9.0392E-62
Gm13112	havana	47.2586371	-7.3739839	0.5150509	-14.317	1.2096E-44
Gm27199	havana	15.4763399	-7.3472801	0.62246811	-11.803464	1.2985E-30
Slc27a2	ensembl_havana	127.825247	-7.2903468	0.42968325	-16.966793	2.1797E-62

Hsd17b2	ensembl	22.7452619	-7.2850421	0.63458937	-11.479931	5.3087E-29
Nmur2	ensembl_havana	26.9660517	-7.2715069	0.62240322	-11.682952	5.1984E-30
B230323A14Rik	ensembl	22.3733119	-7.2692702	0.63547442	-11.439123	8.3584E-29
Gm37111	havana	109.029451	-7.2261236	0.43507761	-16.608815	8.1266E-60
Tfap2b	ensembl_havana	41.7032549	-7.1903365	0.52196768	-13.775444	2.2041E-41
Ghsr	ensembl_havana	17.2407408	-7.1800658	0.64310118	-11.164753	1.7252E-27
Gm28322	havana	22.1545533	-7.1729369	0.67657234	-10.601877	6.9504E-25
Tmem215	ensembl_havana	24.6771263	-7.1576716	0.6176296	-11.588939	1.5303E-29
Fndc9	ensembl_havana	16.1556394	-7.1191409	0.62230488	-11.439957	8.2868E-29
Mnx1	ensembl_havana	19.4123437	-7.0700692	0.63536422	-11.127585	2.5909E-27
A730046J19Rik	havana	22.9949422	-7.0531294	0.62235191	-11.333024	2.73E-28
Grm1	ensembl_havana	262.687086	-7.0517268	0.40136351	-17.569427	7.4881E-67
Doc2a	ensembl_havana	306.893653	-7.0330003	0.19933711	-35.281942	1.733E-268
Gm27544	ensembl	18.2146526	-6.9994484	0.63252391	-11.065903	5.038E-27
Lhx1	ensembl_havana	63.1582101	-6.9793515	0.40607871	-17.187189	5.1667E-64
Cdh7	ensembl_havana	113.542787	-6.974702	0.37919959	-18.393221	3.5746E-73
Nrsn2	ensembl_havana	1015.87245	-6.953626	0.61945725	-11.225352	8.9522E-28
Dmrt3	ensembl	14.3995235	-6.9479374	0.63435423	-10.952772	1.7218E-26
A530058N18Rik	havana	60.9901103	-6.9408801	0.39568586	-17.54139	1.1997E-66
Rmst	ensembl_havana	68.4081864	-6.887144	0.39697518	-17.349055	3.2754E-65
Klhl14	ensembl_havana	59.3001024	-6.8800468	0.41343618	-16.641134	4.8018E-60
Ccdc108	ensembl_havana	136.396446	-6.8656725	0.52391713	-13.104501	1.587E-37
Dao	ensembl_havana	260.240716	-6.8438062	0.48214726	-14.194431	6.8615E-44
Evx1	ensembl_havana	16.3501841	-6.8335151	0.64525689	-10.590379	7.8294E-25
Nxph2	ensembl_havana	36.0551604	-6.8301077	0.48918486	-13.962222	1.7191E-42
Gm38103	havana	14.1943485	-6.8246265	0.68614708	-9.9463027	5.1583E-22
Gpr12	ensembl_havana	27.4537955	-6.8137635	0.55286639	-12.324431	2.7007E-33
Klk6	ensembl_havana	830.777346	-6.8082237	0.62760994	-10.847858	5.2208E-26
Sp8	ensembl	13.9386855	-6.8055559	0.69040714	-9.8573082	1.2222E-21
Pax8	ensembl_havana	137.312869	-6.7684095	0.62869896	-10.76574	1.2459E-25
Gng13	ensembl_havana	65.2511109	-6.7467524	0.37345042	-18.065992	1.2201E-70
Vip	ensembl	15.1951773	-6.7441492	0.6361585	-10.601366	6.9778E-25
Slc12a5	ensembl_havana	5395.9174	-6.7151909	0.3513302	-19.113617	6.0684E-79
Klhl1	ensembl_havana	133.839853	-6.7142321	0.47174368	-14.232797	4.0202E-44
Zdbf2	ensembl_havana	178.739057	-6.6912669	0.22370153	-29.911583	8.68E-193
Hapln2	ensembl_havana	1362.4187	-6.6719764	0.36030768	-18.517442	3.8198E-74
Sox1	havana	186.723314	-6.6651267	0.38188454	-17.453251	5.368E-66
Fam163b	ensembl_havana	464.282384	-6.649801	0.49525061	-13.427144	2.3612E-39

Zfp474	ensembl	10.4302905	-6.6371604	0.71901761	-9.2308733	4.2106E-19
Smim17	ensembl_havana	498.004305	-6.6218026	0.13767059	-48.098892	0
4930426D05Rik	ensembl_havana	41.5479576	-6.6194345	0.42677729	-15.510278	2.8203E-52
Slc24a4	ensembl_havana	67.247096	-6.6179646	0.35966696	-18.400257	3.1868E-73
Evx2	ensembl_havana	9.09624536	-6.6121201	0.63720692	-10.376724	7.0521E-24
Ctxn2	ensembl	69.7984577	-6.5960689	0.35617523	-18.519168	3.7299E-74
A930006I01Rik	havana	27.2227317	-6.5942421	0.66338727	-9.9402603	5.4637E-22
Uncx	ensembl_havana	34.5454092	-6.5626998	0.51326472	-12.786189	9.1333E-36
D930020B18Rik	ensembl_havana	9.07059971	-6.5623834	0.65440167	-10.028066	2.3185E-22
Crhbp	ensembl	33.2384462	-6.5602595	0.47174778	-13.906286	3.7014E-42
Grm2	ensembl_havana	29.4831589	-6.5566565	0.49025626	-13.373937	4.7334E-39
Tac2	ensembl	169.987917	-6.5485075	0.93793523	-6.9818334	2.323E-11
Slc26a3	ensembl_havana	21.0524574	-6.5418704	0.66376191	-9.8557485	1.2405E-21
Sox14	ensembl_havana	8.75494631	-6.5299443	0.65519653	-9.9663902	4.2553E-22
Cacng3	ensembl_havana	158.687756	-6.5253917	0.26441866	-24.678258	2.856E-131
Gm29771	havana	10.4678782	-6.4784763	0.78838426	-8.2174095	2.3878E-15
Gabra5	ensembl_havana	536.483552	-6.4700451	0.29209795	-22.150259	1.011E-105
Bbox1	ensembl_havana	181.276625	-6.4619396	0.43815431	-14.748091	2.5449E-47
Wnt8b	ensembl	9.99048976	-6.4494139	0.64218032	-10.042995	2.0021E-22
Gm8104	havana	20.8801195	-6.4092578	0.57072872	-11.229955	8.5059E-28
Alox8	ensembl_havana	118.867046	-6.4016411	0.31312561	-20.444323	3.4517E-90
Gm15478	havana	10.6350465	-6.3959331	0.71739635	-8.9154804	6.9671E-18
Chodl	ensembl_havana	277.124519	-6.3717871	0.49190971	-12.953164	1.1043E-36
Olfm3	ensembl_havana	309.651061	-6.3620131	0.29562281	-21.520711	7.581E-100
Pkd2l1	ensembl_havana	232.450411	-6.351742	0.35557474	-17.863311	4.3856E-69

Supplementary Table 5. Top 100 most downregulated genes in spinal tumors.

gene	chromosome	start	end	total IS	total samples	total reads	**corrected p value**	Shared IS	Brain only IS	Spine only IS
Cdkn2a	4	89096058	89489079	99	47	233719	0	1		
Nf1	11	79287193	79619803	47	29	18100	0	1		
Ppp1r14c	10	3309887	3724803	43	29	3155	0	1		
Pten	19	32678851	32905851	30	20	110519	0	1		
Sox6	7	115548304	115871683	29	19	8971	0	1		
Map7	10	20073797	20308872	22	20	15183	0	1		
Adgrl3	5	81048981	81303956	20	19	6741	5.5511E-16	1		
Asap1	15	64110005	64371418	20	17	2096	1.1102E-16	1		
Sox5	6	143942931	144227797	20	19	9252	3.3694E-07	1		
Csmd3	15	47486085	47660361	18	18	21497	0	1		
Exosc9	3	36447497	36668433	18	18	9996	0	1		
Spred1	2	117042029	117209386	18	13	3570	0	1		
Clcn3	8	60850884	61027199	17	14	1925	0	1		
Ctnnd2	15	30484570	30707255	17	15	9065	0	1		
Pik3r1	13	101593163	101847492	17	16	2457	0	1		
Ust	10	8261283	8476769	17	16	2330	0	1		
Snx29	16	11314621	11548051	16	15	10056	0	1		
Dmd	X	84744328	85004890	15	15	14998	0	1		
Slc8a1	17	81501469	81695955	15	15	5217	0	1		
Tcf12	9	71862595	72068316	15	13	1263	1.6817E-05	1		
Zfat	15	67719615	67820188	15	11	14253	0	1		
Zfhx4	3	5239819	5286895	15	8	12633	0	1		
Csnk1g3	18	53839007	54013921	14	13	805	2.2196E-05	1		
Nova1	12	46671946	46867497	14	14	11444	2.8796E-08	1		
Nrip1	16	76237345	76451322	14	14	10157	0	1		
Phlda1	10	111429778	111566905	14	14	4974	0	1		
Tnr	1	159624066	159831238	14	13	2494	0	1		
Asb16	11	102211607	102341521	13	10	16049	0	1		
Epn2	11	61453428	61639298	13	12	1885	0	1		
Nav3	10	109682551	109747074	13	13	20475	2.4789E-06	1		
Ptprj	2	90471620	90658666	13	13	2422	0	1		
Qk	17	10222528	10426738	13	13	1014	9.0392E-05	1		
Tcf4	18	69435503	69610417	13	12	1290	1.8171E-05	1		
Tmub2	11	102200050	102358455	13	10	16049	0	1		
Zcchc11	4	108393432	108530990	13	12	15557	0	1		
9430076C15Rik	6	53537831	53688058	12	12	1163	2.7605E-12	1		

Crebbp	16	4068564	4214458	12	11	5279	3.2305E-05	1		
Klhl13	X	23136091	23319449	12	11	819	2.6341E-09	1		
Myo7b	18	31867868	32062217	12	12	1054	2.2919E-12	1		
Rad51b	12	79201818	79368036	12	11	9554	0	1		
Tmtc2	10	105535455	105582382	12	11	13401	7.6176E-05	1		
Trio	15	27734894	27938215	12	11	1577	7.6712E-07	1		
Zeb2	2	45034252	45184826	12	10	917	5.3173E-06	1		
Zfyve26	12	79249807	79335580	12	11	9554	0	1		
Ankrd28	14	31613680	31756906	11	10	1364	3.9097E-06	1		
Decr1	4	15817126	16013637	11	11	11166	4.4614E-06	1		
Diaph3	14	86812929	86994349	11	11	4533	1.8874E-15	1		
Gtdc1	2	44505726	44600169	11	11	10641	3.4515E-06	1		
Nlgn1	3	25889316	26066141	11	10	615	2.6795E-05	1		
Zwint	10	72630894	72673870	11	11	8769	0	1		
1110015O18Rik	3	4782549	4808029	10	7	1071	2.7404E-06	1		
Btd	14	31618262	31738328	10	9	1336	0	1		
Ddah1	3	145717448	145923743	10	10	2287	0	1		
Fam118b	9	35165806	35312750	10	10	3708	8.121E-06	1		
Lims2	18	31892312	32026045	10	10	713	0	1		
Map2	1	66248995	66367379	10	9	483	1.7548E-05	1		
Mmp16	4	18018049	18135955	10	9	682	3.6597E-05	1		
Pcdh7	5	57807058	57954158	10	10	1901	0.00011275	1		
Usp9y	Y	1069938	1414895	10	5	1517	1.1102E-16	1		
Zfhx3	8	108708891	108849884	10	10	12239	0	1		
Cask	X	13649735	13833093	9	9	3131	1.2054E-07	1		
Ddx3y	Y	1073714	1340359	9	4	1488	0	1		
Gria4	9	4591680	4709235	9	8	2077	3.5137E-05	1		
Iyd	10	3490141	3557502	9	8	459	5.0069E-07	1		
Map3k1	13	111776572	111817509	9	7	487	7.9972E-07	1		
Mir6400	4	15893410	15989584	9	9	11116	3.0865E-07	1		
Mir99ahg	16	77404495	77521210	9	8	2600	0.00014562	1		
Mis18bp1	12	65102607	65219938	9	9	16655	7.57E-08	1		
Mtus1	8	40937098	41035051	9	9	4923	7.2205E-10	1		
Pias1	9	62859832	62967591	9	7	1070	2.9582E-05	1		
Ptpro	6	137401664	137472271	9	8	15780	1.0596E-05	1		
Selenok	14	29907290	29987375	9	8	14224	6.3017E-09	1		
Sema3d	5	12349761	12473276	9	9	620	5.6819E-06	1		
1810041L15Rik	15	84413067	84519569	8	7	11988	0	1		

Abhd2	7	79251458	79369051	8	8	831	8.4939E-07	1		
AI838599	4	3232361	3251950	8	6	4280	0	1		
Cdc14a	3	116380149	116450767	8	8	23872	0	1		
Cngb3	4	19320602	19346068	8	7	992	1.6934E-07	1		
Cypt15	X	39373938	39400880	8	5	11211	0	1		
Dner	1	84588589	84697108	8	7	3180	6.4798E-05	1		
Dync1li2	8	104401780	104454576	8	8	2293	0	1		
Esr1	10	4889977	4965291	8	6	1025	0.00016908	1		
Gna14	19	16529747	16580612	8	7	7927	0	1		
Gpc5	14	115389812	115458420	8	8	1184	0	1		
Nbn	4	15926454	15973486	8	8	11022	3.3307E-16	0	1	
Plekhg1	10	3685807	3726789	8	8	484	1.585E-09	1		
Rpusd4	9	35214146	35308006	8	8	3329	1.3735E-05	1		
Sat1	X	155153743	155259897	8	7	1419	0	1		
St7	6	17677810	17766179	8	8	14717	1.2536E-05	1		
Stxbp6	12	44803650	44838720	8	7	23157	0	1		
Trpm8	1	88230156	88333646	8	8	397	2.3633E-05	1		
Ttc28	5	111155608	111341936	8	8	914	0	1		
Zfp458	13	67209802	67278275	8	7	657	0	1		
0610040J01Rik	5	63808770	63946064	7	7	6545	5.5218E-06	1		
Abca13	11	9295201	9309824	7	6	8545	1.1102E-16	1		
Akap13	7	75585763	75656199	7	7	390	0	1		
Asb3	11	30978506	30990192	7	5	351	2.6685E-06	0		1
Atp10a	7	58756137	58810945	7	7	4857	0	0	1	
Bmper	9	23528352	23536154	7	4	28034	6.4819E-06	0		1
Cdh11	8	102627714	102706076	7	7	267	6.5793E-05	1		
Dcc	18	72055466	72089361	7	7	8252	4.6028E-05	1		
Dlgap4	2	156676140	156745052	7	7	74607	0	1		
Gsap	5	21160657	21182161	7	6	541	2.9413E-10	0		1
Kpna2	11	106987525	107011871	7	6	56059	0	1		
Map3k4	17	12261677	12319793	7	7	1528	0	1		
Mb21d2	16	28859776	28903405	7	7	1238	0	1		
Musk	4	58304960	58337294	7	7	11817	0	1		
Nbea	3	55753466	55797616	7	7	1116	5.1135E-06	1		
Nek7	1	138489579	138532906	7	7	1436	1.2947E-06	1		
Nemp2	1	52564794	52635692	7	7	787	5.6224E-09	1		
Prex2	1	11266542	11282297	7	7	1703	4.2391E-05	1		
Rbm7	9	48447489	48509113	7	7	11035	2.5167E-08	1		

Rexo2	9	48459906	48499052	7	7	11035	0.0001147	1		
Sorcs1	19	50406272	50485396	7	7	291	3.9339E-08	1		
Tank	2	61595336	61644477	7	7	17183	5.845E-06	1		
Tdrd3	14	87554499	87579225	7	6	3877	4.1233E-05	1		
Tenm3	8	48485013	48543675	7	7	542	9.9238E-06	1		
AA545190	6	10957359	10980856	6	6	969	2.9153E-05	1		
Actbl2	13	111243423	111292208	6	6	918	1.0811E-05	1		
Bcas3	11	85666582	85686068	6	6	1666	4.7744E-05	1		
Bicc1	10	70974373	71033030	6	6	166	1.4156E-05	1		
Caln1	5	130537072	130563475	6	6	1095	1.1262E-05	1		
Ccser1	6	62151190	62162936	6	5	631	7.432E-06	1		
Clasrp	7	19553438	19629745	6	6	8970	0	0	1	
Dach1	14	97945559	97963618	6	6	8107	0	1		
Dleu2	14	61630762	61651684	6	6	1333	0	1		
Gpr155	2	73340199	73369663	6	6	5732	2.9318E-06	1		
Htr1f	16	64944590	64961997	6	6	486	1.1109E-05	1		
Ifitm1	7	140929084	140999580	6	6	9729	0	1		
Kcnh8	17	52597828	52617174	6	6	2624	1.7509E-05	1		
Nalcn	14	123620476	123636635	6	5	2352	5.473E-05	1		
Nell2	15	95415889	95433245	6	6	891	1.6427E-05	1		
Prr30	14	101048943	101064151	6	5	1301	6.6744E-05	1		
Relb	7	19558441	19628937	6	6	8970	3.6673E-05	0	1	
RP23-456L15.3	9	35249367	35312883	6	6	3239	4.9873E-08	1		
Setd5	6	113069053	113104327	6	6	1859	8.1915E-05	1		
Sp3	2	72947644	72979083	6	6	317	8.8572E-05	1		
Tbc1d8b	X	139680170	139722540	6	6	1566	5.4165E-05	1		
Tmem132b	5	125501839	125560585	6	6	1373	7.9314E-06	1		
Tspan7	X	10474178	10503045	6	6	2095	1.2259E-05	1		
1700080O16Rik	X	51992179	52026832	5	5	8557	0	1		
2610307P16Rik	13	28873488	28887127	5	5	31546	3.8598E-11	0		1
2700049A03Rik	12	71135349	71242902	5	5	302	7.1532E-06	1		
5730585A16Rik	15	103052314	103065813	5	5	477	3.8761E-09	1		
Alg14	3	121262154	121293565	5	5	4630	5.3824E-13	0	1	
Arap1	7	101359693	101418390	5	4	315	0	0		1
Baz1a	12	54883543	54900087	5	5	870	5.9099E-08	1		
Cbl	9	44179325	44192979	5	5	743	2.1882E-06	1		
Celf3	3	94468695	94490255	5	5	1295	0	0		1
Crem	18	3286413	3356329	5	5	345	8.7353E-05	1		

Eif3e	15	43258513	43264295	5	5	17142	4.0492E-05	1		
Fam46d	X	107866517	107870364	5	4	287	3.797E-14	1		
Gpc4	X	51994154	52021117	5	5	8557	3.6327E-05	1		
Gt(ROSA)26Sor	6	113068020	113080745	5	5	1735	3.2E-06	1		
Hcrtr2	9	76214423	76225151	5	5	315	2.532E-06	1		
Ighv1-53	12	115150972	115167516	5	5	5760	5.9603E-09	1		
Iqgap1	7	80731162	80829156	5	5	579	5.6024E-05	1		
Nrd1	4	108991346	109006031	5	5	2394	4.8571E-07	0		1
Olfr1385	11	49484110	49495796	5	5	2554	1.1538E-06	0	1	
Pdzrn4	15	92588555	92603009	5	5	253	2.0612E-06	1		
Pkd1l1	11	8949602	8965183	5	5	558	0	1		
Ptbp2	3	119733935	119748629	5	5	780	3.8601E-09	1		
Robo2	16	74394673	74407244	5	5	6884	1.1833E-07	1		
RP24-260O23.1	9	9455457	9470085	5	5	1040	1.1486E-06	1		
Rps18-ps2	4	9017185	9034808	5	5	275	0	0		1
Smarcc1	9	110201976	110214654	5	5	259	2.6937E-06	1		
Spag17	3	100013554	100028248	5	5	11933	4.8054E-09	1		
Tex11	X	100955040	100978133	5	4	429	1.445E-05	1		
Tmeff1	4	48590114	48609703	5	5	197	1.8149E-06	1		
Tmem56	3	121277829	121295504	5	5	4630	0.00018597	0	1	
Tomm7	5	23838187	23853820	5	5	360	1.8496E-05	1		
Tox	4	6834873	6848580	5	5	897	2.9907E-06	1		
Tpt1-ps2	4	54590149	54611696	5	5	290	2.4568E-08	1		
Vps54	11	21235629	21261949	5	5	506	4.3764E-05	1		
1700014L14Rik	8	18349901	18356728	4	4	22455	1.0031E-05	0	1	
4933439N14Rik	8	12462246	12521017	4	4	246	0.00010552	1		
A4gnt	9	99604881	99612683	4	4	820	6.5177E-06	0	1	
Agbl1	7	76174911	76188584	4	4	468	7.6613E-06	1		
Atoh8	6	72201041	72215724	4	4	2604	0	0		1
Atp2a2	5	122494917	122502734	4	4	8776	4.7205E-06	0	1	
Bcl10	3	145920673	145927530	4	4	9541	4.0375E-06	0	1	
Begain	12	109024710	109042236	4	4	6415	0	0	1	
Ccdc62	5	123948827	123958598	4	4	1939	1.0121E-05	1		
Csmd1	8	16027601	16034428	4	4	3948	2.5882E-06	0		1
Dgkb	12	38640104	38651782	4	4	29771	1.1652E-05	1		
Eif2ak3	6	70844355	70851207	4	4	16991	1.062E-05	0		1
Enc1	13	97241603	97251340	4	4	1006	2.8697E-06	1		
Enpp2	15	54889275	54892166	4	4	6147	4.5508E-05	0	1	

Erap1	13	74636251	74644041	4	4	5699	9.752E-06	0		1
F630206G17Rik	11	45776757	45784547	4	4	1321	2.8702E-05	1		
Frmpd4	X	167597497	167612893	4	4	3528	1.1607E-05	0	1	
Glp2r	11	67707304	67717043	4	4	4436	3.7951E-06	1		
Gpam	19	54943295	54953602	4	4	456	2.287E-06	0		1
Gripap1	X	7787619	7810712	4	4	502	0	1		
Hdac9	12	34524503	34530342	4	4	540	2.7449E-05	1		
Hipk2	6	38865332	38874142	4	4	640	2.3848E-06	1		
Invs	4	48285090	48288027	4	4	12184	4.0042E-05	1		
Krtap7-1	16	89528150	89532018	4	4	188	3.8388E-05	1		
Lama3	18	12549492	12559150	4	4	401	1.3944E-05	1		
Lhfpl2	13	94126746	94136483	4	4	11636	3.6042E-06	1		
Lmntd1	6	145429773	145440541	4	4	34869	8.0881E-06	0		1
Lrrc8d	5	105699725	105710473	4	4	279	5.1012E-07	1		
Magi2	5	19904424	19913218	4	4	191	1.4119E-05	0		1
Myo10	15	25703687	25726841	4	3	406	6.3252E-05	1		
Nampt	12	32821428	32830187	4	4	2596	1.1603E-05	1		
Nkain3	4	20308617	20317428	4	4	1028	1.6207E-05	1		
Pakap	4	57778979	57790727	4	4	1561	5.7348E-09	0	1	
Pde4b	4	102563855	102573645	4	4	12230	5.0818E-06	1		
Pdgfra	5	75143204	75158838	4	4	438	3.652E-07	0	1	
Picalm	7	90191959	90204650	4	4	29655	2.011E-06	1		
Pla2g2d	4	138761447	138776132	4	4	548	0	0		1
Prcc	3	87860603	87871379	4	4	11675	1.1487E-05	1		
Prdm16	4	154433167	154447853	4	4	679	2.4425E-15	1		
Prdm5	6	65850086	65869691	4	4	5111	2.4531E-05	1		
Prkca	11	108256420	108260315	4	4	318	3.8866E-05	0		1
Rab27b	18	70050432	70053329	4	4	190	5.0411E-05	1		
Rpl10l	12	66046953	66059605	4	4	1037	3.5302E-07	1		
Sel1l3	5	53113743	53123514	4	4	217	1.3822E-05	0		1
Ska1	18	74129842	74135636	4	4	2957	3.1329E-05	0		1
Slc7a11	3	49985247	49997002	4	4	3008	1.9128E-06	1		
Smarca2	19	26712545	26722852	4	4	249	1.4439E-05	0		1
Smoc1	12	81134247	81147872	4	4	16193	2.2767E-08	0		1
Snd1	6	28837016	28848762	4	4	523	1.0384E-05	0	1	
Stox2	8	47236736	47246490	4	4	546	4.7938E-06	1		
Tenm2	11	36046293	36054084	4	4	1156	9.9283E-06	1		
Tm4sf1	3	57282307	57290144	4	4	1379	2.3639E-06	0		1

Tmc1	19	20962010	20975129	4	4	872	7.2815E-06	0	1	
Ttc39b	4	83272917	83302311	4	4	899	9.4283E-06	1		
Usp1	4	98918836	98923731	4	4	1140	2.6525E-05	1		
Vmn2r-ps3	3	64036667	64042545	4	4	1013	1.2578E-05	0		1
Vps53	11	76034485	76043250	4	4	429	4.1726E-06	0		1
Wdr36	18	32838374	32848032	4	4	1350	1.7072E-05	0	1	
Whsc1l1	8	25687432	25697186	4	4	102	1.2495E-05	1		
Ybx3	6	131380342	131389152	4	4	2746	1.6807E-05	1		
Yy1	12	108788286	108798991	4	4	410	9.319E-07	1		
Ankrd40	11	94323900	94329743	3	3	512	4.4229E-05	0		1
Btbd9	17	30519503	30534022	3	3	949	4.5949E-05	1		
Cacul1	19	60529850	60540158	3	3	554	4.247E-05	1		
Dusp16	6	134756395	134761289	3	3	1791	2.5904E-05	0		1
Dync2li1	17	84644856	84655489	3	3	625	1.8176E-05	0		1
Fat3	9	16408032	16410958	3	3	1896	3.8904E-05	0	1	
Galnt16	12	80526006	80529899	3	3	1069	3.3152E-05	1		
Gria3	X	41398480	41415800	3	3	22863	3.1738E-05	0	1	
Hpf1	8	60879883	60889636	3	3	636	1.1064E-05	0		1
Il1rapl1	X	87436221	87449687	3	3	183	0	0		1
Itga1	13	115036892	115044681	3	3	124	2.7668E-05	1		
Jdp2	12	85616740	85620632	3	3	1137	2.4645E-05	1		
Kcnj16	11	110991895	110998712	3	3	187	1.8284E-05	1		
Lyzl1	18	4125086	4132813	3	3	213	2.615E-05	1		
Magi1	6	93692667	93696582	3	3	720	2.7975E-05	0	1	
Mbp	18	82534932	82538795	3	3	262	5.1967E-05	1		
Pabpc6	17	9760726	9770393	3	3	443	2.4046E-05	1		
Pappa	4	65239372	65244267	3	3	330	2.0161E-05	1		
Parp8	13	116970654	116974549	3	3	626	3.4426E-05	1		
Pebp4	14	69966628	69970430	3	3	847	4.4883E-05	0	1	
Pisd-ps2	17	3080063	3086830	3	3	149	2.6555E-05	1		
Ppfibp2	7	107694551	107699432	3	3	1125	4.3106E-05	0	1	
Ppp2r2b	18	42803182	42805114	3	3	845	5.181E-05	0		1
Rps18-ps1	4	71478884	71488675	3	3	1214	1.1936E-05	0	1	
Ruvbl2	7	45428511	45438273	3	3	728	2.41E-05	0		1
Slc48a1	15	97786263	97790117	3	3	162	4.3113E-05	0		1
Trim27	13	21158858	21167622	3	3	266	2.0745E-05	0	1	
Zc3hav1l	6	38281937	38294663	3	3	19447	3.934E-07	0		1
Zfp473	7	44748088	44750040	3	3	760	5.495E-05	1		

1700018G05Rik	X	102931119	102934005	2	2	80	7.0959E-06	0		1
Basp1	15	25389106	25423940	2	2	61	0.0002068	0		1
Erdr1	Y	90706026	90854162	2	2	60	0	1		
Mtcl1	17	66393658	66432503	2	2	217	0.00014725	0		1
Myt1	2	181772347	181813638	2	2	68	1.3798E-05	0		1
Plxna1	6	89341680	89345595	2	2	605	8.4903E-07	0	1	
Acp7	7	28595604	28600485	1	1	17	3.287E-05	0		1
App	16	85050861	85094602	1	1	11	0.00014838	0	1	
Arhgef28	13	97895928	97899823	1	1	70	2.992E-05	0	1	
Htr2a	14	74642139	74644991	1	1	21	5.4403E-05	0		1
Ptk2	15	73227545	73256488	1	1	255	2.7657E-05	0	1	
Cox11	11	90635666	90642504	11	10	0	0.00015471	0	1	
Gmnc	16	26950849	26952783	4	4	0	2.0535E-05	0	1	
Tom1l1	11	90633751	90651324	9	8	0	6.0095E-05		1	
Grm3	5	9563886	9568771	4	4	0	3.4468E-05	0	1	
Pik3c3	18	29756604	29758536	4	4	0	5.6515E-05	0		1
Xylt1	7	117422552	117424504	3	3	0	5.4756E-05	0	1	

Supplementary Table 6. All *piggyBac* common integration sites (CIS) across 96 brain and spinal gliomas.

name	Chromosome	position	feature_name	gene_name	Total read counts
Apc2_sense-10:80302495	10	80302495	CbASA	Apc2	356
Cdkn2a_sense-4:89294333	4	89294333	CbASA	Cdkn2a	325
Dmd_sense-X:84848795	X	84848795	CbASA	Dmd	261
Pten_sense-19:32818061	19	32818061	CbASA	Pten	230
Afap1l2_sense-19:56944704	19	56944704	CbASA	Afap1l2	119
Pcdh15_sense-10:74230381	10	74230381	CbASA	Pcdh15	112
Ubap1_sense-4:41371850	4	41371850	CbASA	Ubap1	93
Rad51b_sense-12:79300541	12	79300541	SD	Rad51b	72
Cdkn2a_sense-4:89281985	4	89281985	CbASA	Cdkn2a	69
Ebf1_sense-11:44618193	11	44618193	SD	Ebf1	59
Ebf1_sense-11:44620413	11	44620413	SD	Ebf1	56
App_sense-16:84978063	16	84978063	CbASA	App	54
Ptpro_sense-6:137441144	6	137441144	CbASA	Ptpro	52
noGene_-6:113076212	6	113076212	En2SA	noGene	43
noGene_-6:113076491	6	113076491	En2SA	noGene	43
Prdm5_sense-6:65857818	6	65857818	CbASA	Prdm5	41
Golph3_sense-15:12321957	15	12321957	CbASA	Golph3	40
Nf1_sense-11:79340060	11	79340060	CbASA	Nf1	34
Exosc9_sense-3:36553130	3	36553130	CbASA	Exosc9	33
Qk_sense-17:10273946	17	10273946	CbASA	Qk	33
Sox5_sense-6:144116443	6	144116443	CbASA	Sox5	32
Zfhx3_sense-8:108715331	8	108715331	CbASA	Zfhx3	32
Myt1_sense-2:181807661	2	181807661	CbASA	Myt1	31
Gpm6a_sense-8:54955316	8	54955316	CbASA	Gpm6a	26
noGene_-8:69716767	8	69716767	CbASA	noGene	18
Epha4_sense-1:77444864	1	77444864	CbASA	Epha4	17
Xylt1_sense-7:117381521	7	117381521	CbASA	Xylt1	17
Myt1_sense-2:181815614	2	181815614	CbASA	Myt1	16
Zfhx3_sense-8:108704236	8	108704236	CbASA	Zfhx3	14
Ctnnd2_sense-15:30634817	15	30634817	CbASA	Ctnnd2	13
Lhfpl2_sense-13:94120836	13	94120836	CbASA	Lhfpl2	13
St7_sense-6:17694363	6	17694363	CbASA	St7	13
Trip12_sense-1:84814822	1	84814822	CbASA	Trip12	13
Asap1_sense-15:64312348	15	64312348	CbASA	Asap1	12
Slc8a1_sense-17:81647808	17	81647808	CbASA	Slc8a1	12
3110035E14Rik_sense-1:9607030	1	9607030	CbASA	3110035E14Rik	11
Eif2ak3_sense-6:70845022	6	70845022	CbASA	Eif2ak3	11

Lin52_sense-12:84462453	12	84462453	CbASA	Lin52	11
Adam12_sense-7:133929892	7	133929892	CbASA	Adam12	10
Cask_sense-X:13846007	X	13846007	CbASA	Cask	10
Csmd1_sense-8:16033503	8	16033503	CbASA	Csmd1	10
Dmxl1_sense-18:49939610	18	49939610	CbASA	Dmxl1	10
Thsd7a_sense-6:12748796	6	12748796	CbASA	Thsd7a	10
Mmp16_sense-4:18112115	4	18112115	CbASA	Mmp16	9
Asrgl1_sense-19:9118528	19	9118528	CbASA	Asrgl1	8
Ehbp1_sense-11:22285487	11	22285487	CbASA	Ehbp1	8
Kit_sense-5:75575119	5	75575119	CbASA	Kit	8
Lncpint_sense-6:31139728	6	31139728	CbASA	Lncpint	8
noGene_-15:51652158	15	51652158	SD	noGene	8
noGene_-2:50636953	2	50636953	SD	noGene	8
Qk_sense-17:10238892	17	10238892	CbASA	Qk	8
Unc5c_sense-3:141465838	3	141465838	CbASA	Unc5c	8
Inpp5f_sense-7:128664306	7	128664306	CbASA	Inpp5f	7
noGene_-8:123041948	8	123041948	CbASA	noGene	7
Slc25a38_sense-9:120110573	9	120110573	CbASA	Slc25a38	7
Utrn_sense-10:12475270	10	12475270	CbASA	Utrn	7
Erbb4_sense-1:68250579	1	68250579	CbASA	Erbb4	6
Nf1_sense-11:79475884	11	79475884	CbASA	Nf1	6
noGene_-15:95668776	15	95668776	SD	noGene	6
Plppr1_sense-4:49206496	4	49206496	CbASA	Plppr1	6
Spag9_sense-11:94013744	11	94013744	CbASA	Spag9	6
Ust_sense-10:8390887	10	8390887	CbASA	Ust	6
App_sense-16:84971407	16	84971407	CbASA	App	5
Col11a1_sense-3:114124353	3	114124353	CbASA	Col11a1	5
Fndc4_sense-5:31295663	5	31295663	SD	Fndc4	5
Gria4_sense-9:4664767	9	4664767	CbASA	Gria4	5
Kif1b_sense-4:149307428	4	149307428	CbASA	Kif1b	5
Mmp16_sense-4:17987721	4	17987721	CbASA	Mmp16	5
Ndufs4_sense-13:114316856	13	114316856	CbASA	Ndufs4	5
noGene_-1:139474960	1	139474960	CbASA	noGene	5
noGene_-8:54779651	8	54779651	CbASA	noGene	5
noGene_-8:69716283	8	69716283	CbASA	noGene	5
noGene_-X:102950132	X	102950132	CbASA	noGene	5
Pde4b_sense-4:102570879	4	102570879	CbASA	Pde4b	5
Ptprt_sense-2:162238006	2	162238006	CbASA	Ptprt	5

Rgcc_sense-14:79300714	14	79300714	CbASA	Rgcc	5
Usp24_sense-4:106371096	4	106371096	CbASA	Usp24	5
Xylt1_sense-7:117549098	7	117549098	CbASA	Xylt1	5
Zeb2_sense-2:45110015	2	45110015	CbASA	Zeb2	5
1110038B12Rik_sense-17:34951884	17	34951884	CbASA	1110038B12Rik	4
Aig1_sense-10:13690523	10	13690523	CbASA	Aig1	4
Arap1_sense-7:101404426	7	101404426	CbASA	Arap1	4
Col11a1_sense-3:114090258	3	114090258	CbASA	Col11a1	4
Derl1_sense-15:57875496	15	57875496	CbASA	Derl1	4
Dscam_sense-16:97038897	16	97038897	CbASA	Dscam	4
Gm10184_antisense-17:89910340	17	89910340	En2SA	Gm10184	4
Gmds_sense-13:32225134	13	32225134	CbASA	Gmds	4
Itpr1_sense-6:108391970	6	108391970	CbASA	Itpr1	4
Lcorl_sense-5:45795292	5	45795292	CbASA	Lcorl	4
Man1a2_sense-3:100578475	3	100578475	CbASA	Man1a2	4
Map3k4_sense-17:12318497	17	12318497	CbASA	Map3k4	4
Myt1_sense-2:181822760	2	181822760	CbASA	Myt1	4
Nav1_sense-1:135532348	1	135532348	CbASA	Nav1	4
Nav3_sense-10:109714232	10	109714232	CbASA	Nav3	4
Ncam1_sense-9:49798677	9	49798677	CbASA	Ncam1	4
noGene_-15:51663808	15	51663808	SD	noGene	4
noGene_-3:113600371	3	113600371	CbASA	noGene	4
noGene_-6:113076759	6	113076759	En2SA	noGene	4
Pcdhgb4_sense-18:37722918	18	37722918	CbASA	Pcdhgb4	4
Pde4b_sense-4:102255457	4	102255457	CbASA	Pde4b	4
Sept7_sense-9:25265497	9	25265497	CbASA	43350	4
Sox6_sense-7:115701695	7	115701695	CbASA	Sox6	4
Tcf4_sense-18:69564685	18	69564685	CbASA	Tcf4	4
Tnr_sense-1:159524216	1	159524216	CbASA	Tnr	4
Ube3c_sense-5:29619788	5	29619788	CbASA	Ube3c	4
Xylt1_sense-7:117667619	7	117667619	CbASA	Xylt1	4
Abca1_sense-4:53127596	4	53127596	CbASA	Abca1	3
Abhd2_sense-7:79297208	7	79297208	CbASA	Abhd2	3
Agap1_sense-1:89789325	1	89789325	CbASA	Agap1	3
Amy1_sense-3:113605571	3	113605571	CbASA	Amy1	3
Anks1b_sense-10:90680920	10	90680920	CbASA	Anks1b	3
Arhgap31_sense-16:38712573	16	38712573	CbASA	Arhgap31	3
Arid2_sense-15:96289153	15	96289153	CbASA	Arid2	3

Brinp1_sense-4:68954013	4	68954013	CbASA	Brinp1	3
Btd_sense-14:31662326	14	31662326	CbASA	Btd	3
Cdc5l_sense-17:45427938	17	45427938	CbASA	Cdc5l	3
Clk1_sense-1:58423919	1	58423919	CbASA	Clk1	3
Col9a1_sense-1:24230955	1	24230955	CbASA	Col9a1	3
Diaph2_sense-X:130109382	X	130109382	CbASA	Diaph2	3
Dpp10_sense-1:123854165	1	123854165	CbASA	Dpp10	3
Dync1h1_sense-12:110601855	12	110601855	CbASA	Dync1h1	3
Fam126b_sense-1:58556508	1	58556508	CbASA	Fam126b	3
Gm10259_antisense-3:25212629	3	25212629	En2SA	Gm10259	3
Gria3_sense-X:41478800	X	41478800	CbASA	Gria3	3
Hnrnpa2b1_sense-6:51463410	6	51463410	CbASA	Hnrnpa2b1	3
Itga9_sense-9:118672905	9	118672905	CbASA	Itga9	3
Kcnh5_sense-12:75087404	12	75087404	CbASA	Kcnh5	3
Kiz_sense-2:146861183	2	146861183	CbASA	Kiz	3
Klhl9_sense-4:88722161	4	88722161	CbASA	Klhl9	3
Lbh_sense-17:72921291	17	72921291	CbASA	Lbh	3
Lclat1_sense-17:73196796	17	73196796	CbASA	Lclat1	3
Lyrm4_sense-13:36092805	13	36092805	CbASA	Lyrm4	3
Map4_sense-9:110063124	9	110063124	CbASA	Map4	3
Map4k3_sense-17:80649526	17	80649526	CbASA	Map4k3	3
Megf11_sense-9:64544687	9	64544687	CbASA	Megf11	3
Mex3c_sense-18:73573777	18	73573777	CbASA	Mex3c	3
Mmp16_sense-4:18054579	4	18054579	CbASA	Mmp16	3
Nlgn1_sense-3:26153264	3	26153264	CbASA	Nlgn1	3
noGene_-15:47586613	15	47586613	CbASA	noGene	3
noGene_-19:22040987	19	22040987	SD	noGene	3
noGene_-2:44565811	2	44565811	CbASA	noGene	3
noGene_-4:32321264	4	32321264	SD	noGene	3
noGene_-6:113076539	6	113076539	En2SA	noGene	3
noGene_-6:31126448	6	31126448	CbASA	noGene	3
noGene_-9:68427694	9	68427694	SD	noGene	3
Nxph1_sense-6:8950308	6	8950308	CbASA	Nxph1	3
Pcdh15_sense-10:74626843	10	74626843	CbASA	Pcdh15	3
Pde7a_sense-3:19256777	3	19256777	CbASA	Pde7a	3
Plekha5_sense-6:140426556	6	140426556	CbASA	Plekha5	3
Ptprj_sense-2:90452059	2	90452059	CbASA	Ptprj	3
Qk_sense-17:10318731	17	10318731	CbASA	Qk	3

Rnpc3_sense-3:113605571	3	113605571	CbASA	Rnpc3	3
Sccpdh_sense-1:179670696	1	179670696	CbASA	Sccpdh	3
Slco3a1_sense-7:74504176	7	74504176	CbASA	Slco3a1	3
Sorcs1_sense-19:50378896	19	50378896	CbASA	Sorcs1	3
Spink8_sense-9:109816751	9	109816751	SD	Spink8	3
Tcf4_sense-18:69564164	18	69564164	CbASA	Tcf4	3
Tmem132b_sense-5:125698751	5	125698751	CbASA	Tmem132b	3
Utrn_sense-10:12455362	10	12455362	CbASA	Utrn	3
Utrn_sense-10:12478412	10	12478412	CbASA	Utrn	3
2810410L24Rik_sense-11:120188488	11	120188488	CbASA	2810410L24Rik	2
Adam23_sense-1:63545595	1	63545595	CbASA	Adam23	2
Adamts1_sense-16:85795905	16	85795905	CbASA	Adamts1	2
Adgrl3_sense-5:81197632	5	81197632	CbASA	Adgrl3	2
Adgrl3_sense-5:81560944	5	81560944	CbASA	Adgrl3	2
Afap1l2_sense-19:56943540	19	56943540	SD	Afap1l2	2
Agrn_sense-4:156195243	4	156195243	CbASA	Agrn	2
Aifm2_sense-10:61725855	10	61725855	SD	Aifm2	2
Anks1b_sense-10:90042703	10	90042703	CbASA	Anks1b	2
App_sense-16:85030366	16	85030366	SD	App	2
Arsb_sense-13:93794219	13	93794219	CbASA	Arsb	2
Aspm_sense-1:139474374	1	139474374	CbASA	Aspm	2
Astn2_sense-4:65911659	4	65911659	CbASA	Astn2	2
Atcay_sense-10:81210523	10	81210523	CbASA	Atcay	2
Atxn10_sense-15:85393507	15	85393507	CbASA	Atxn10	2
Brinp1_sense-4:68792824	4	68792824	CbASA	Brinp1	2
Brinp3_sense-1:146751992	1	146751992	CbASA	Brinp3	2
Cadm2_sense-16:66882732	16	66882732	CbASA	Cadm2	2
Cdc37l1_sense-19:28999352	19	28999352	CbASA	Cdc37l1	2
Cdh11_sense-8:102647462	8	102647462	CbASA	Cdh11	2
Cdh20_sense-1:104768899	1	104768899	CbASA	Cdh20	2
Chd7_sense-4:8690776	4	8690776	CbASA	Chd7	2
Chmp2b_sense-16:65546848	16	65546848	CbASA	Chmp2b	2
Chsy3_sense-18:59179549	18	59179549	CbASA	Chsy3	2
Chtop_sense-3:90507544	3	90507544	CbASA	Chtop	2
Clock_sense-5:76226932	5	76226932	CbASA	Clock	2
Cntn1_sense-15:92318074	15	92318074	CbASA	Cntn1	2
Cntn4_sense-6:106550517	6	106550517	CbASA	Cntn4	2
Col11a1_sense-3:114066575	3	114066575	CbASA	Col11a1	2

Col11a1_sense-3:114147824	3	114147824	CbASA	Col11a1	2
Col11a1_sense-3:114213287	3	114213287	CbASA	Col11a1	2
Cpne8_sense-15:90679167	15	90679167	CbASA	Cpne8	2
Creb5_sense-6:53610533	6	53610533	CbASA	Creb5	2
Csmd1_sense-8:16271521	8	16271521	CbASA	Csmd1	2
Csmd1_sense-8:17027224	8	17027224	CbASA	Csmd1	2
Ctnna2_sense-6:77845477	6	77845477	CbASA	Ctnna2	2
Ctnnd2_sense-15:30332147	15	30332147	CbASA	Ctnnd2	2
Cux1_sense-5:136304722	5	136304722	CbASA	Cux1	2
Dcn_sense-10:97483338	10	97483338	SD	Dcn	2
Ddx39b_sense-17:35249070	17	35249070	CbASA	Ddx39b	2
Diaph2_sense-X:130259242	X	130259242	CbASA	Diaph2	2
Dock10_sense-1:80648212	1	80648212	CbASA	Dock10	2
Erbb3_sense-10:128578099	10	128578099	CbASA	Erbb3	2
Ext1_sense-15:53083084	15	53083084	CbASA	Ext1	2
Fam63a_sense-3:95283626	3	95283626	CbASA	Fam63a	2
Fchsd2_sense-7:101139197	7	101139197	CbASA	Fchsd2	2
Fgfrl1_sense-5:108695119	5	108695119	CbASA	Fgfrl1	2
Gab1_sense-8:80879531	8	80879531	CbASA	Gab1	2
Gm31266_sense-3:41562419	3	41562419	SD	Gm31266	2
Gna12_sense-5:140785396	5	140785396	CbASA	Gna12	2
Gpc6_sense-14:117624551	14	117624551	CbASA	Gpc6	2
Gphn_sense-12:78454852	12	78454852	CbASA	Gphn	2
Grid2_sense-6:64094440	6	64094440	CbASA	Grid2	2
Grik4_sense-9:42808492	9	42808492	CbASA	Grik4	2
Gys1_sense-7:45435171	7	45435171	CbASA	Gys1	2
Hivep2_sense-10:14066685	10	14066685	CbASA	Hivep2	2
Hspa4_sense-11:53284099	11	53284099	CbASA	Hspa4	2
Il1rapl1_sense-X:87419303	X	87419303	SD	Il1rapl1	2
Impad1_sense-4:4769311	4	4769311	CbASA	Impad1	2
Itga9_sense-9:118698475	9	118698475	CbASA	Itga9	2
Kat6a_sense-8:22914388	8	22914388	CbASA	Kat6a	2
Khdrbs2_sense-1:32310797	1	32310797	SD	Khdrbs2	2
Kif1b_sense-4:149247227	4	149247227	CbASA	Kif1b	2
Lama4_sense-10:39017402	10	39017402	CbASA	Lama4	2
Lhfpl3_sense-5:22746626	5	22746626	CbASA	Lhfpl3	2
Lims2_sense-18:31944557	18	31944557	CbASA	Lims2	2
Lrrtm3_sense-10:64087847	10	64087847	CbASA	Lrrtm3	2

Lsamp_sense-16:39984728	16	39984728	CbASA	Lsamp	2
Lypd1_sense-1:125910390	1	125910390	CbASA	Lypd1	2
March11_sense-15:26310970	15	26310970	SD	43170	2
Megf10_sense-18:57133731	18	57133731	CbASA	Megf10	2
Mgat5_sense-1:127320766	1	127320766	CbASA	Mgat5	2
Micu2_sense-14:57945384	14	57945384	CbASA	Micu2	2
Mnat1_sense-12:73123927	12	73123927	CbASA	Mnat1	2
Mtmr12_sense-15:12266769	15	12266769	CbASA	Mtmr12	2
Mtss1l_sense-8:110732179	8	110732179	CbASA	Mtss1l	2
Ncam2_sense-16:81517618	16	81517618	CbASA	Ncam2	2
Neto1_sense-18:86473844	18	86473844	CbASA	Neto1	2
noGene_-13:89633247	13	89633247	SD	noGene	2
noGene_-14:100482191	14	100482191	CbASA	noGene	2
noGene_-16:23257311	16	23257311	CbASA	noGene	2
noGene_-16:45413414	16	45413414	SD	noGene	2
noGene_-16:85043575	16	85043575	CbASA	noGene	2
noGene_-6:146891519	6	146891519	CbASA	noGene	2
noGene_-7:76858317	7	76858317	SD	noGene	2
noGene_-8:69716384	8	69716384	CbASA	noGene	2
noGene_-X:72770542	X	72770542	SD	noGene	2
Nova1_sense-12:46816885	12	46816885	CbASA	Nova1	2
Nup93_sense-8:94243772	8	94243772	CbASA	Nup93	2
Omg_sense-11:79503902	11	79503902	CbASA	Omg	2
P2rx7_sense-5:122664675	5	122664675	CbASA	P2rx7	2
Pcdh10_sense-3:45392875	3	45392875	CbASA	Pcdh10	2
Pcdhga6_sense-18:37709653	18	37709653	CbASA	Pcdhga6	2
Pcsk6_sense-7:65959310	7	65959310	CbASA	Pcsk6	2
Pdgfra_sense-5:75170952	5	75170952	CbASA	Pdgfra	2
Pdgfra_sense-5:75189337	5	75189337	CbASA	Pdgfra	2
Pdlim5_sense-3:142391712	3	142391712	CbASA	Pdlim5	2
Phlpp1_sense-1:106173448	1	106173448	CbASA	Phlpp1	2
Phlpp1_sense-1:106380423	1	106380423	CbASA	Phlpp1	2
Phtf2_sense-5:20801901	5	20801901	CbASA	Phtf2	2
Plxna2_sense-1:194712319	1	194712319	CbASA	Plxna2	2
Ppfibp1_sense-6:146938990	6	146938990	CbASA	Ppfibp1	2
Ppp2r2b_sense-18:43059161	18	43059161	CbASA	Ppp2r2b	2
Prdm5_sense-6:65831372	6	65831372	CbASA	Prdm5	2
Prex1_sense-2:166713514	2	166713514	CbASA	Prex1	2

Prkg1_sense-19:31302285	19	31302285	CbASA	Prkg1	2
Prkg1_sense-19:31663970	19	31663970	CbASA	Prkg1	2
Psmb7_sense-2:38613431	2	38613431	CbASA	Psmb7	2
Ptk2_sense-15:73242365	15	73242365	CbASA	Ptk2	2
Ptprk_sense-10:28569974	10	28569974	CbASA	Ptprk	2
Ptprm_sense-17:66890923	17	66890923	CbASA	Ptprm	2
Ptprt_sense-2:162253752	2	162253752	CbASA	Ptprt	2
R3hdm1_sense-1:128169058	1	128169058	CbASA	R3hdm1	2
Rab14_sense-2:35189899	2	35189899	CbASA	Rab14	2
Rabep1_sense-11:70925913	11	70925913	CbASA	Rabep1	2
Rapgef6_sense-11:54642952	11	54642952	CbASA	Rapgef6	2
Rars2_sense-4:34623514	4	34623514	CbASA	Rars2	2
Rasa3_sense-8:13631774	8	13631774	CbASA	Rasa3	2
Rev3l_sense-10:39732614	10	39732614	CbASA	Rev3l	2
Rev3l_sense-10:39859222	10	39859222	CbASA	Rev3l	2
Ripk1_sense-13:34002707	13	34002707	CbASA	Ripk1	2
S100a3_sense-3:90560558	3	90560558	SD	S100a3	2
Sash1_sense-10:8780518	10	8780518	CbASA	Sash1	2
Sash1_sense-10:8850362	10	8850362	CbASA	Sash1	2
Scn3a_sense-2:65472278	2	65472278	CbASA	Scn3a	2
Sgip1_sense-4:102760566	4	102760566	CbASA	Sgip1	2
Sh3rf1_sense-8:61226319	8	61226319	CbASA	Sh3rf1	2
Sorcs3_sense-19:48206798	19	48206798	CbASA	Sorcs3	2
Sox5_sense-6:143861383	6	143861383	CbASA	Sox5	2
Sox6_sense-7:115580564	7	115580564	CbASA	Sox6	2
Sox6_sense-7:115943893	7	115943893	CbASA	Sox6	2
Spon1_sense-7:113788553	7	113788553	CbASA	Spon1	2
Stxbp5_sense-10:9841870	10	9841870	CbASA	Stxbp5	2
Tcf4_sense-18:69349145	18	69349145	CbASA	Tcf4	2
Tmem9_sense-1:136019750	1	136019750	CbASA	Tmem9	2
Tox3_sense-8:90274631	8	90274631	CbASA	Tox3	2
Tpm1_sense-9:67047785	9	67047785	CbASA	Tpm1	2
Txndc12_sense-4:108834857	4	108834857	CbASA	Txndc12	2
Ywhae_sense-11:75733092	11	75733092	CbASA	Ywhae	2
Zbtb38_sense-9:96728981	9	96728981	CbASA	Zbtb38	2
Zc3h13_sense-14:75339473	14	75339473	CbASA	Zc3h13	2

Supplementary Table 7. Top fusion transcripts, ranked according to RNA-seq read counts from 36 brain and spinal gliomas.