Chapter Two: Materials and Methods

Mouse Breeding Strategies

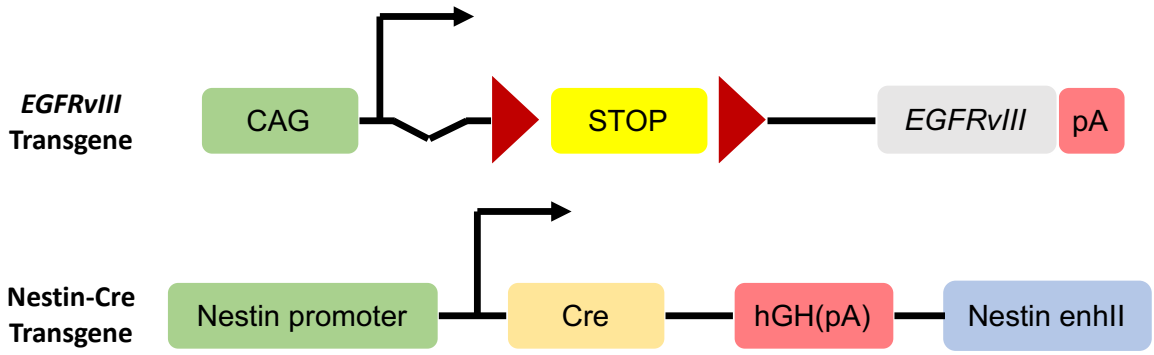
The breeding strategies used in my PhD typically involved multiple crosses between mice, in particular for the *piggyBac* transposon screens, with carefully considered breeding strategies. These strategies are described here.

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

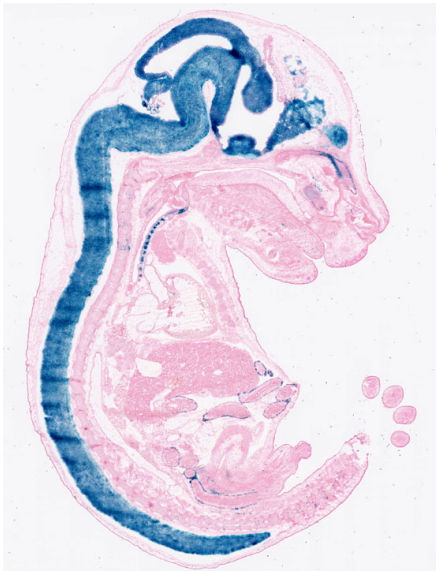


Figure 2.1. *EGFRvIII* was conditionally expressed in the central nervous system using *nes*-cre. A. Structures of *EGFRvIII [96]* and *nes*-cre [87] alleles. Notation: CAG = Cytomegalovirus (CMV) early enhancer, chicken β-actin promoter; pA = poly-adenylation signal; red triangle = *loxP* site; hGH(pA) = 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.

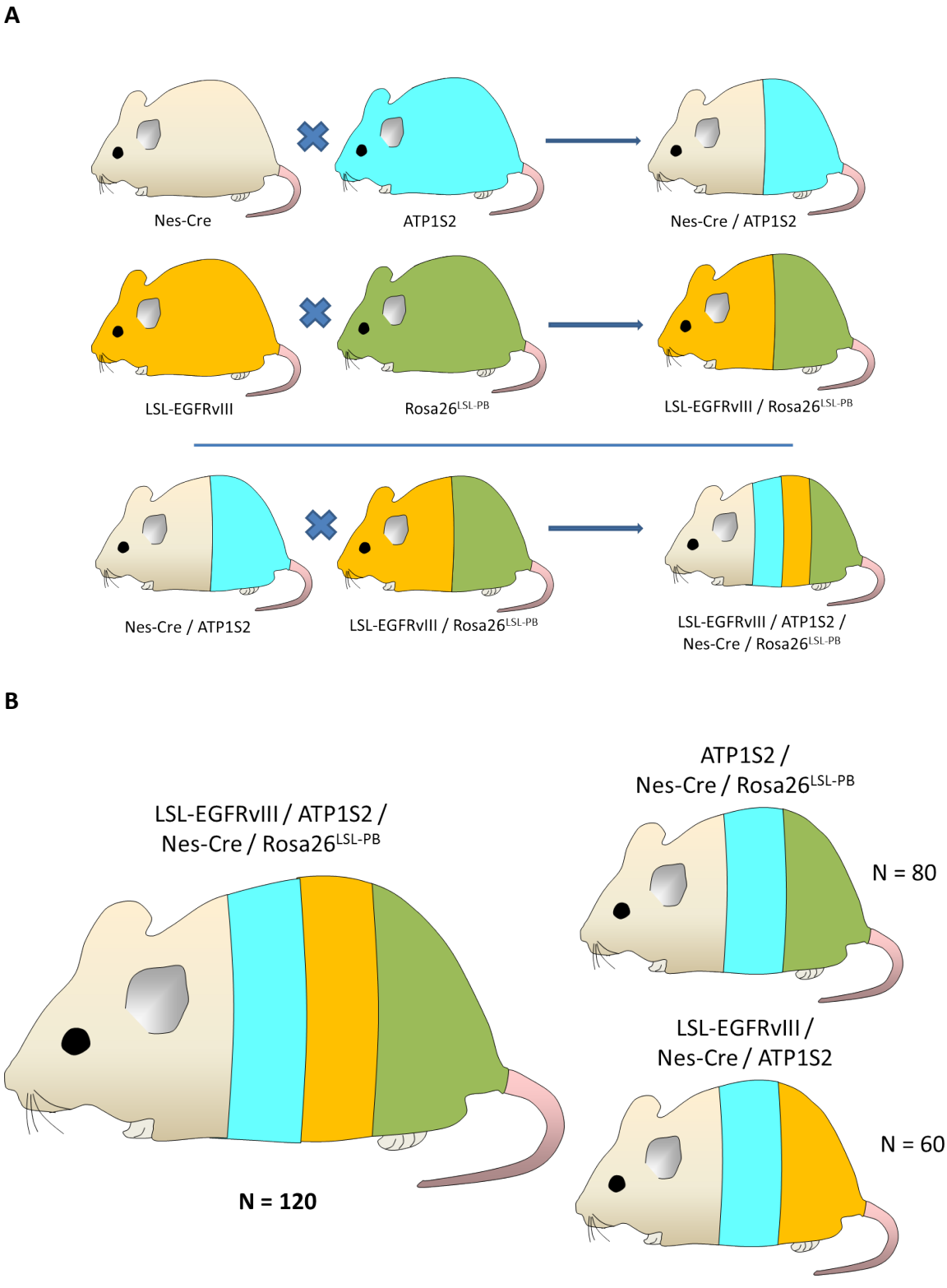


Figure 2.2. Breeding strategy for generating *EGFRvIII*-transposition mice. A. The mouse crosses required in this strategy. B. Final cohorts, with the number of mice with each genotype generated.

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.

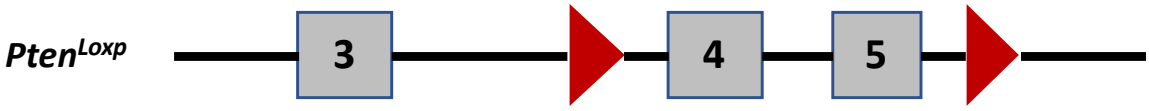


Figure 2.3. *Pten*^{LoxP} allele structure, with exons 4 and 5 flanked by LoxP sites (red arrows) and subsequently deleted with cre expression. Adapted from [42].

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.

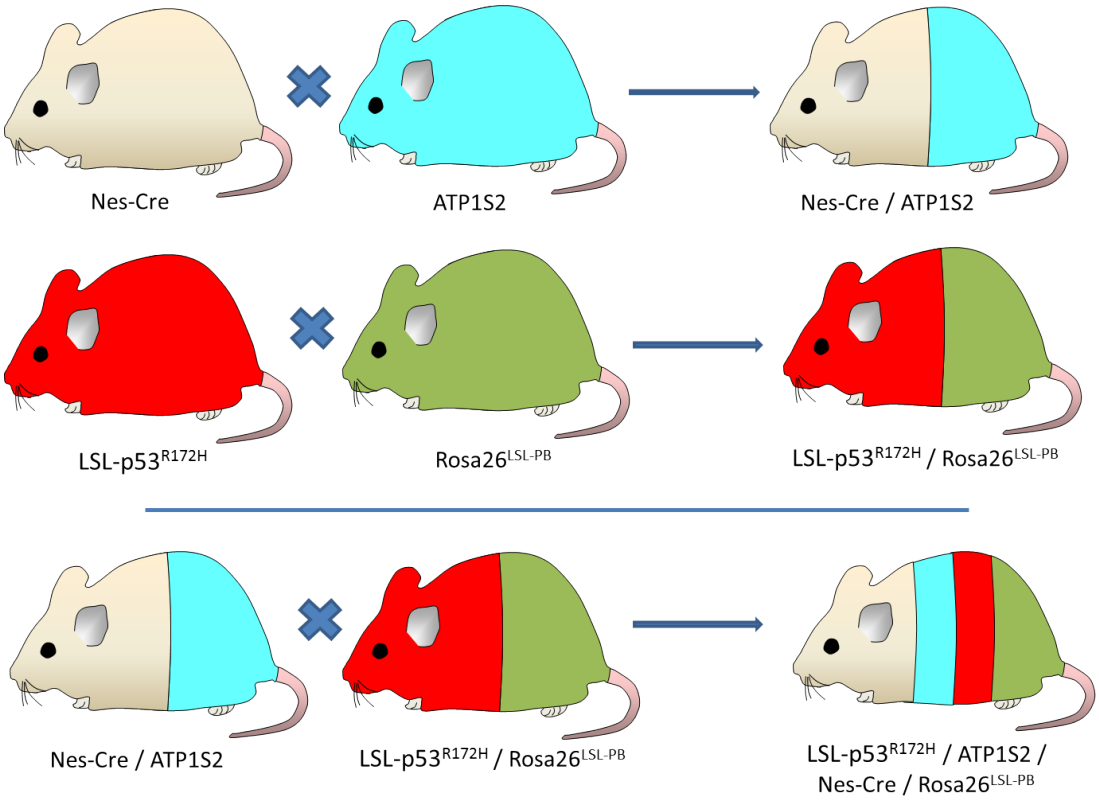


Figure 2.4. Breeding strategy for generating experimental *Trp53*^{R172H} – PB mice.

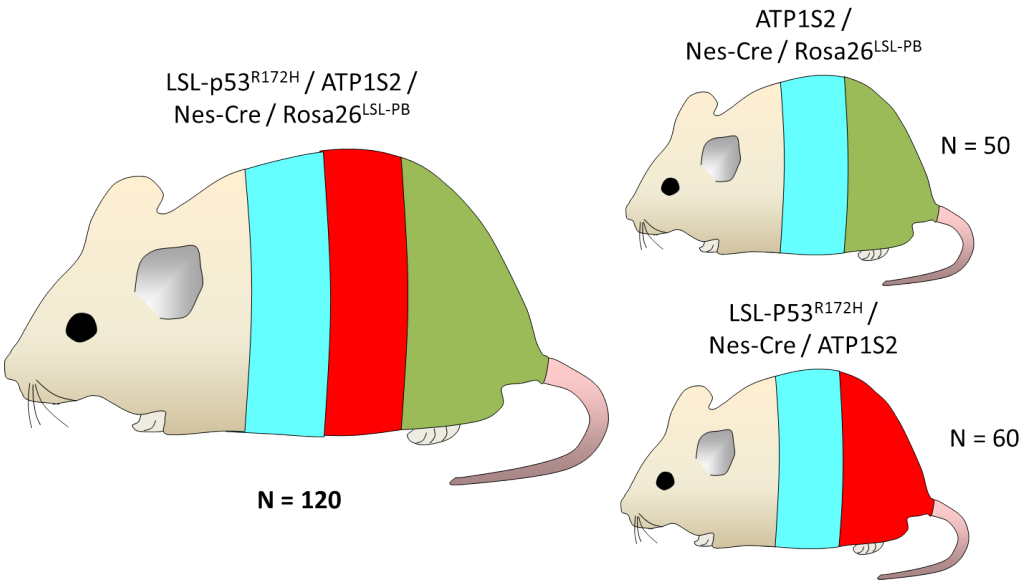


Figure 2.5. Cohorts of experimental mice (left) and control mice (right).

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.

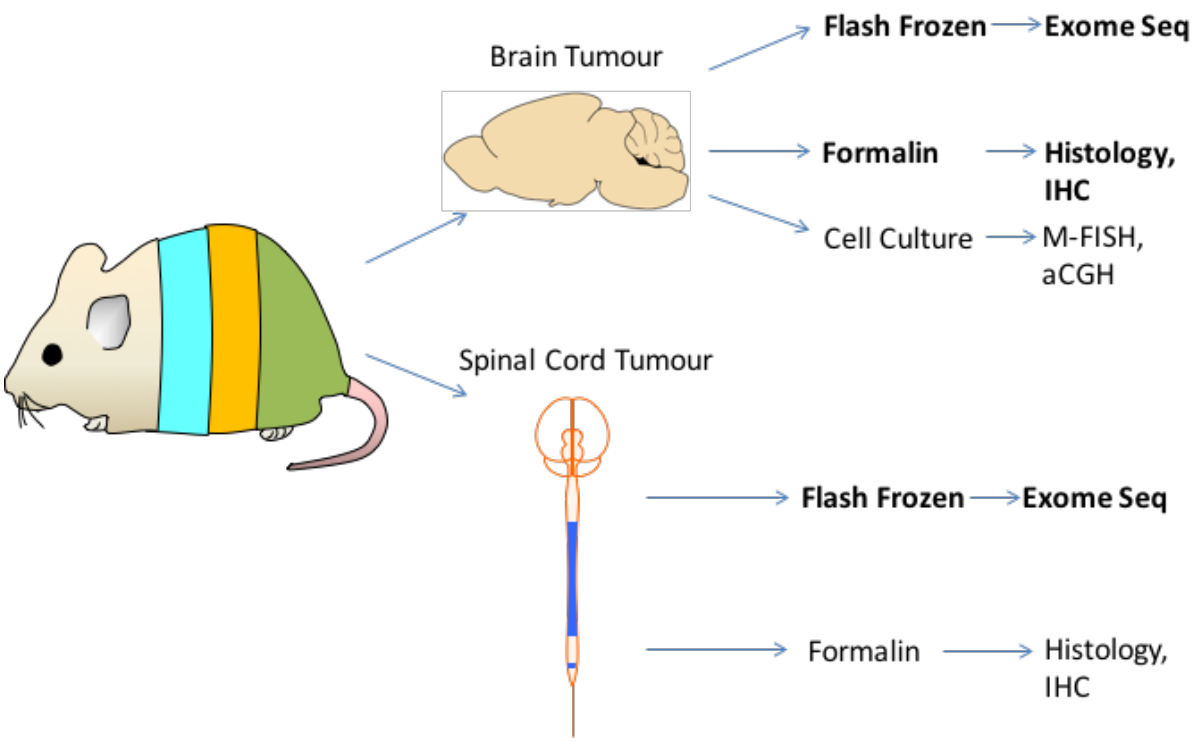


Figure 2.6. Processing strategy for tumors induced in mice. Brain and spinal tumors were collected into different storage reagents dependent on downstream sequencing and experimental requirements.

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.