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.
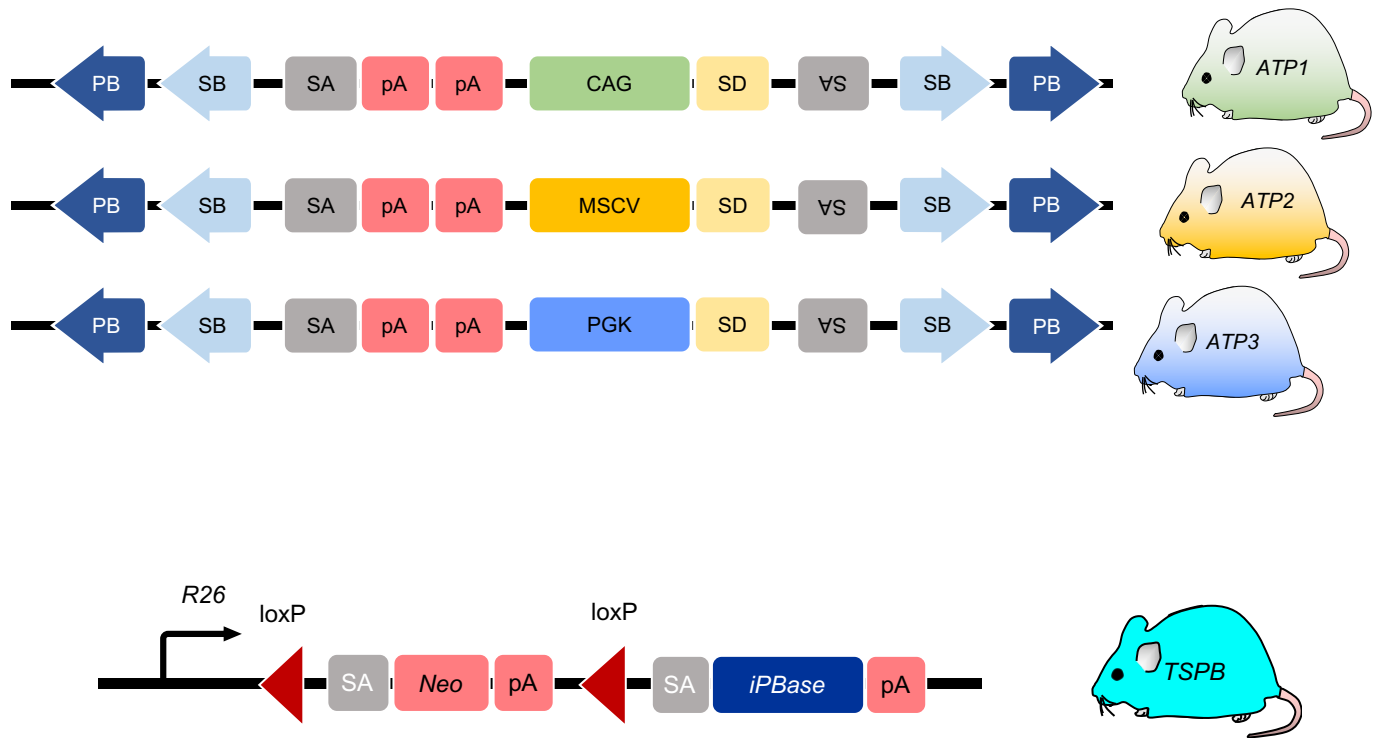
Figure 4.1. A conditional *piggyBac* system for *in vivo* forward genetic screening. Top panel shows the different ATP *piggyBac* transposon mouse lines generated by Rad et al [70]. These lines differ in their promoters for driving expression of the transposon; ATP1 contains a CAG promoter, and this 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.
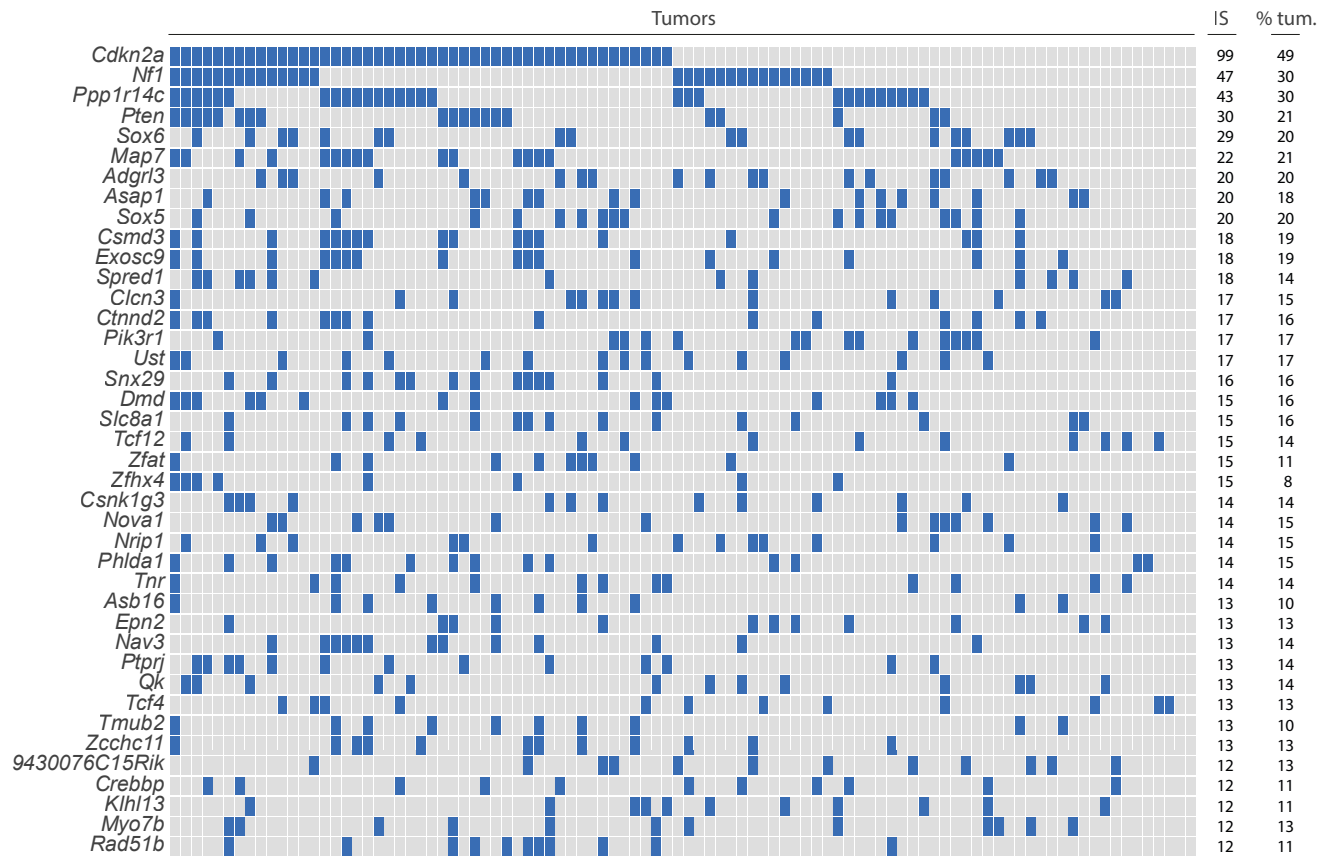

Gene	IS	% tum.
Cdkn2a	99	49
Nf1	47	30
Ppp1r14c	43	30
Pten	30	21
Sox6	29	20
Map7	22	21
Adgrl3	20	20
Asap1	20	18
Sox5	20	20
Csmd3	18	19
Exosc9	18	19
Spred1	18	14
Clcn3	17	15
Ctnnd2	17	16
Pik3r1	17	17
Ust	17	17
Snx29	16	16
Dmd	15	16
Slc8a1	15	16
Tcf12	15	14
Zfat	15	11
Zfhx4	15	8
Csnk1g3	14	14
Nova1	14	15
Nrip1	14	15
Phlda1	14	15
Tnr	14	14
Asb16	13	10
Epn2	13	13
Nav3	13	14
Ptprj	13	14
Qk	13	14
Tcf4	13	13
Tmub2	13	10
Zcchc11	13	13
9430076C15Rik	12	13
Crebbp	12	11
Klhl13	12	11
Myo7b	12	13
Rad51b	12	11

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.
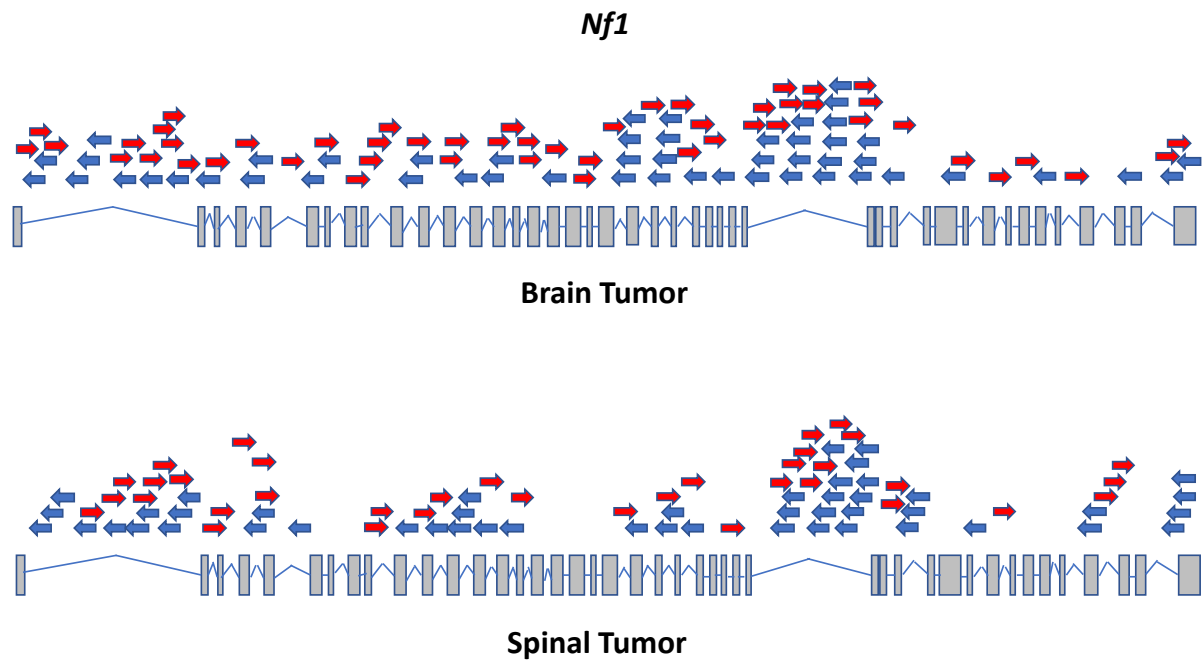
Fig 4.3. The position of all transposon insertions across *Nf1* (a known brain tumor driver) in *EGFRvIII*-PB brain and spinal gliomas, showing a gene-disruption insertion pattern. Blue arrow = antisense orientation; red arrow = sense orientation with respect to gene direction.
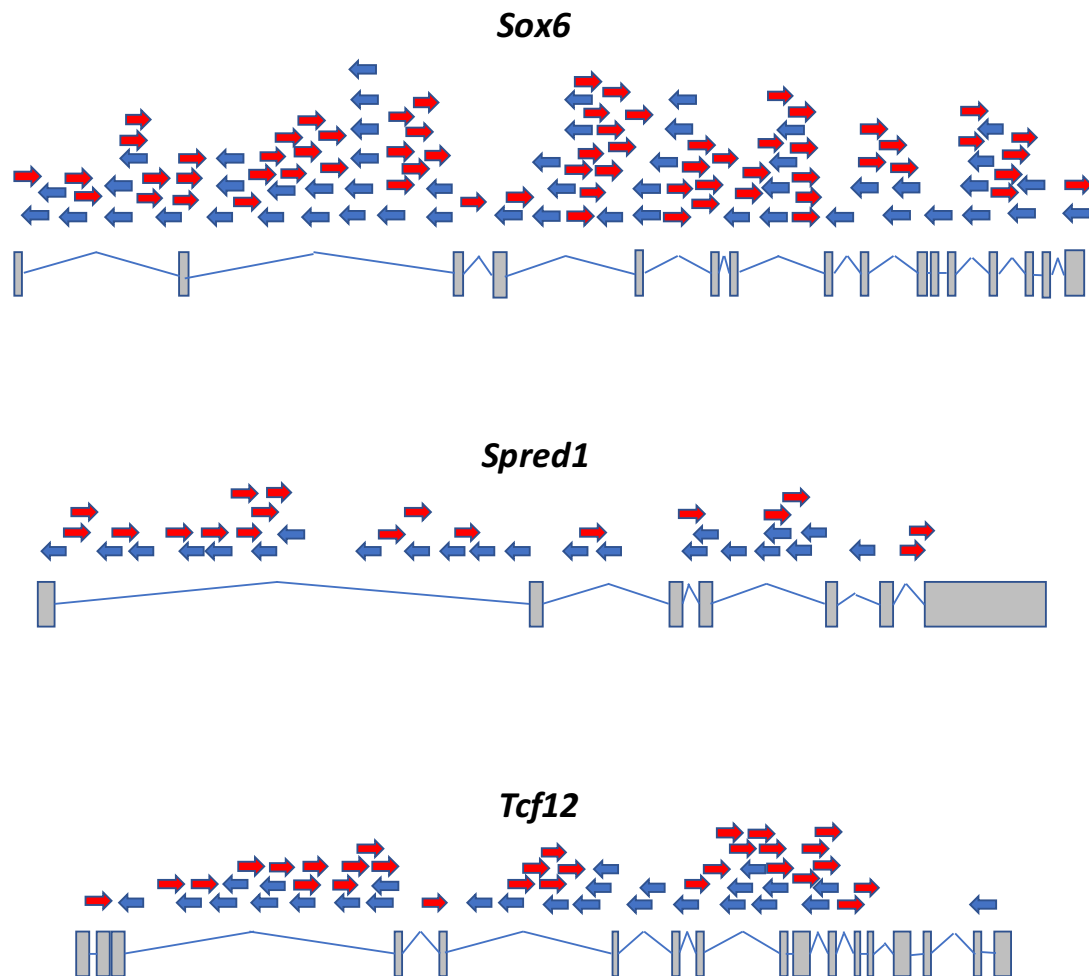
Sox6

Spred1

Tcf12

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 x 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.
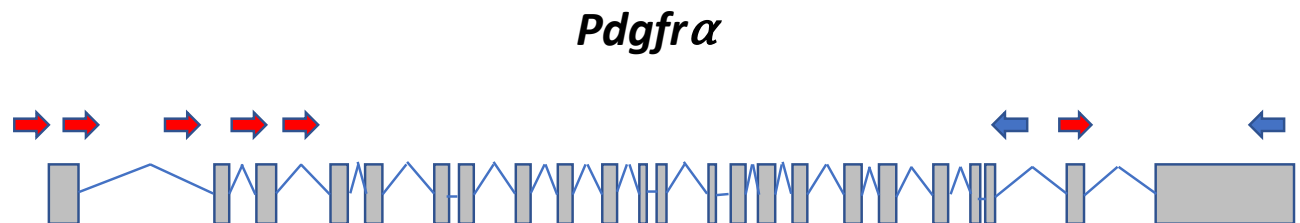
*Pdgfr*α

Figure 4.5. Insertional pattern consistent with *Pdgfr*α 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.
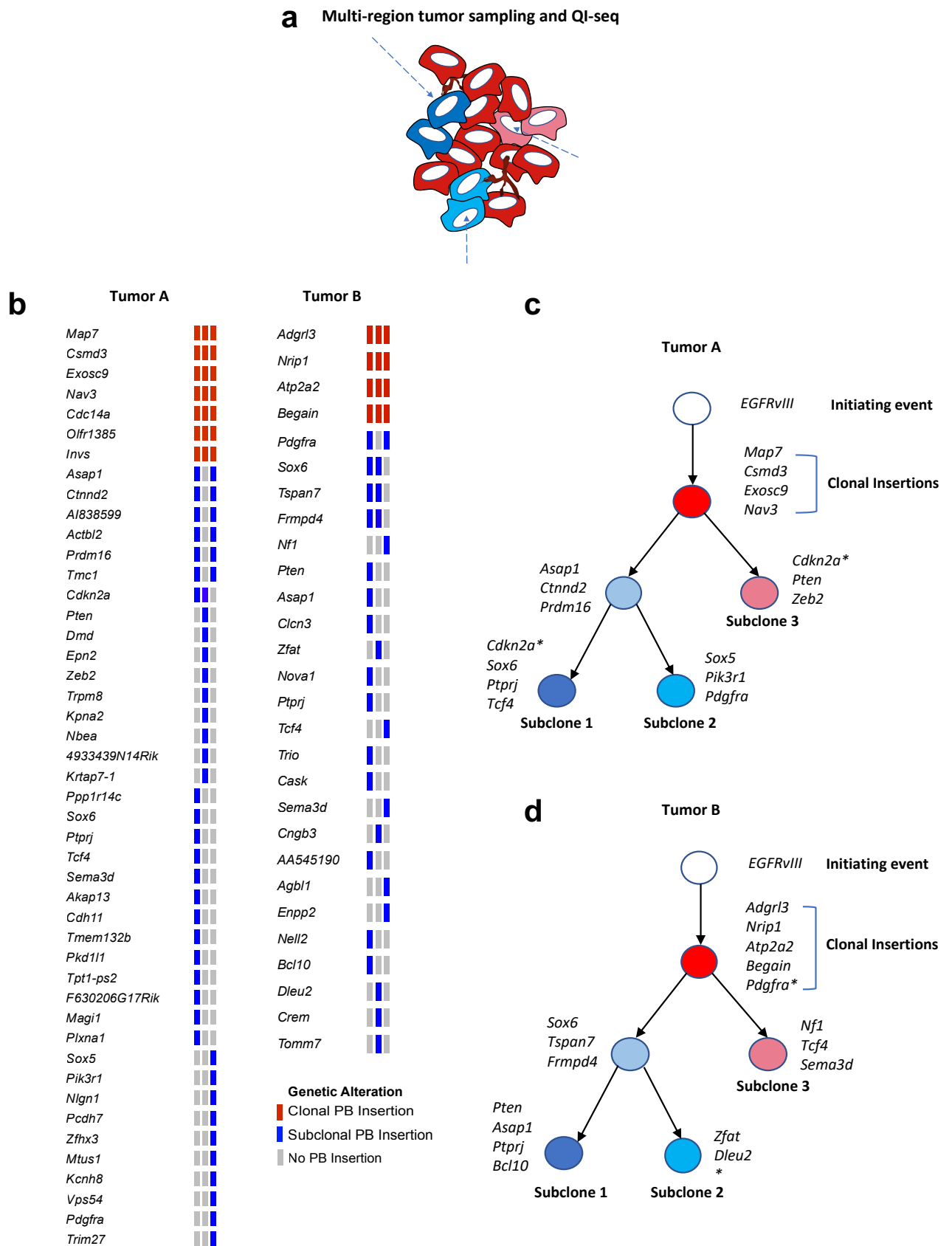
a   **Multi-region tumor sampling and QI-seq**

b

Tumor A	Tumor B
Map7	Adgrl3
Csmd3	Nrip1
Exosc9	Atp2a2
Nav3	Begain
Cdc14a	Pdgfra
Olfr1385	Sox6
Invs	Tspan7
Asap1	Frmpd4
Ctnnd2	Nf1
AI838599	Pten
Actbl2	Asap1
Prdm16	Clcn3
Tmc1	Zfat
Cdkn2a	Nova1
Pten	Ptprj
Dmd	Tcf4
Epn2	Trio
Zeb2	Cask
Trpm8	Sema3d
Kpna2	Cngb3
Nbea	AA545190
4933439N14Rik	Agbl1
Krtap7-1	Enpp2
Ppp1r14c	Nell2
Sox6	Bcl10
Ptprj	Dleu2
Tcf4	Crem
Sema3d	Tomm7
Akap13	
Cdh11	
Tmem132b	
Pkd1l1	
Tpt1-ps2	
F630206G17Rik	
Magi1	
Plxna1	
Sox5	
Pik3r1	
Nlgn1	
Pcdh7	
Zfhx3	
Mtus1	
Kcnh8	
Vps54	
Pdgfra	
Trim27	

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 — **Subclone 1**

Sox5
Pik3r1
Pdgfra — **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 — **Subclone 1**

Zfat
Dleu2
* — **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.
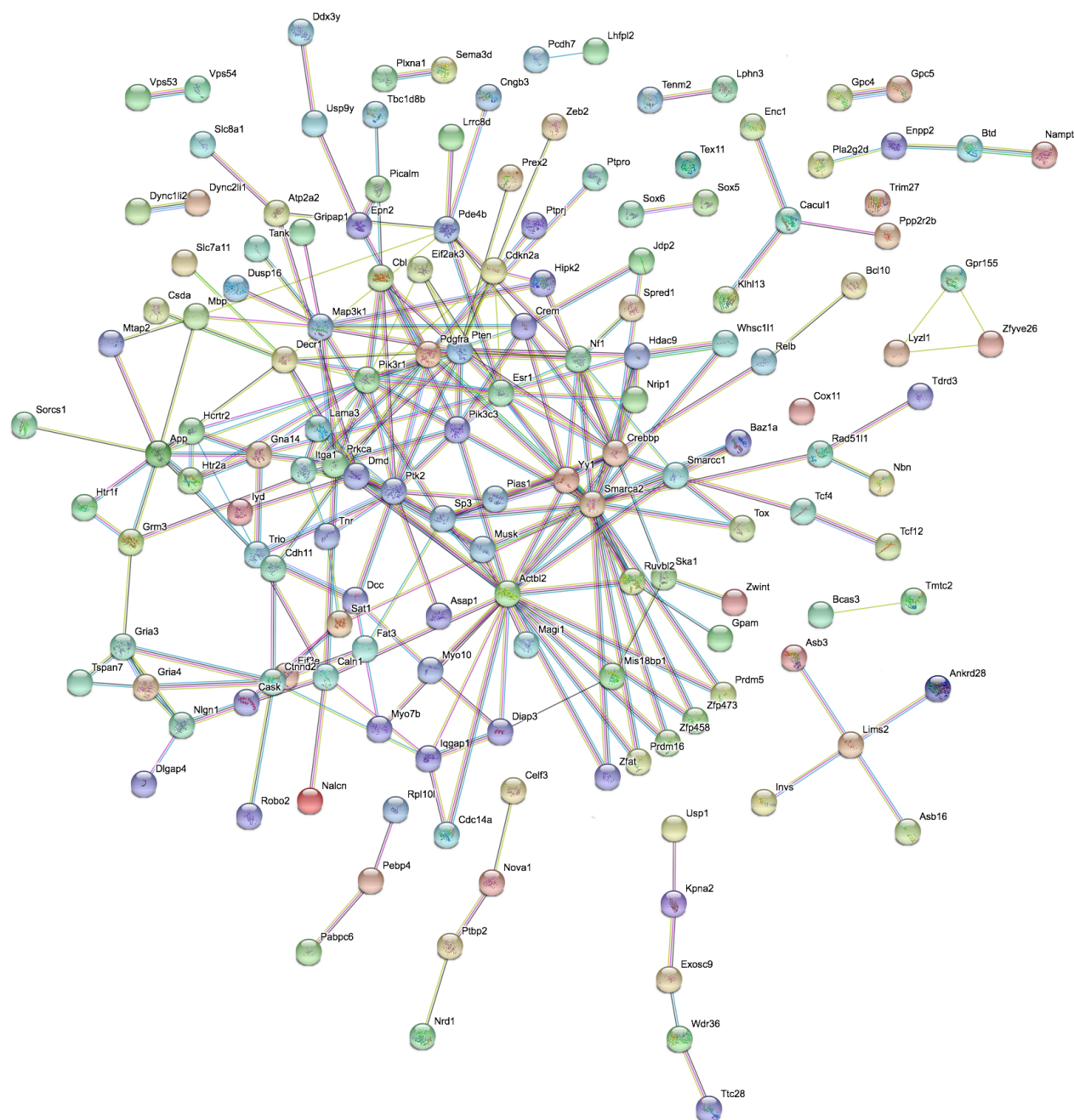
Fig 4.7. Network analysis of all interacting CIS transposon genes. An analysis, performed using STRING, to determine the functional connectivity between CIS genes demonstrates there are 253 interactions between their proteins, showing *PiggyBac* mutagenesis has identified mutations in functionally interacting 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.

Pathway ID	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

Pathway ID	Biological Process (GO)		
	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)	Count in Gene Set	False Discovery Rate
Pathway ID	Pathway Description		
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/.

CIS Genes GO Enrichment Pathways
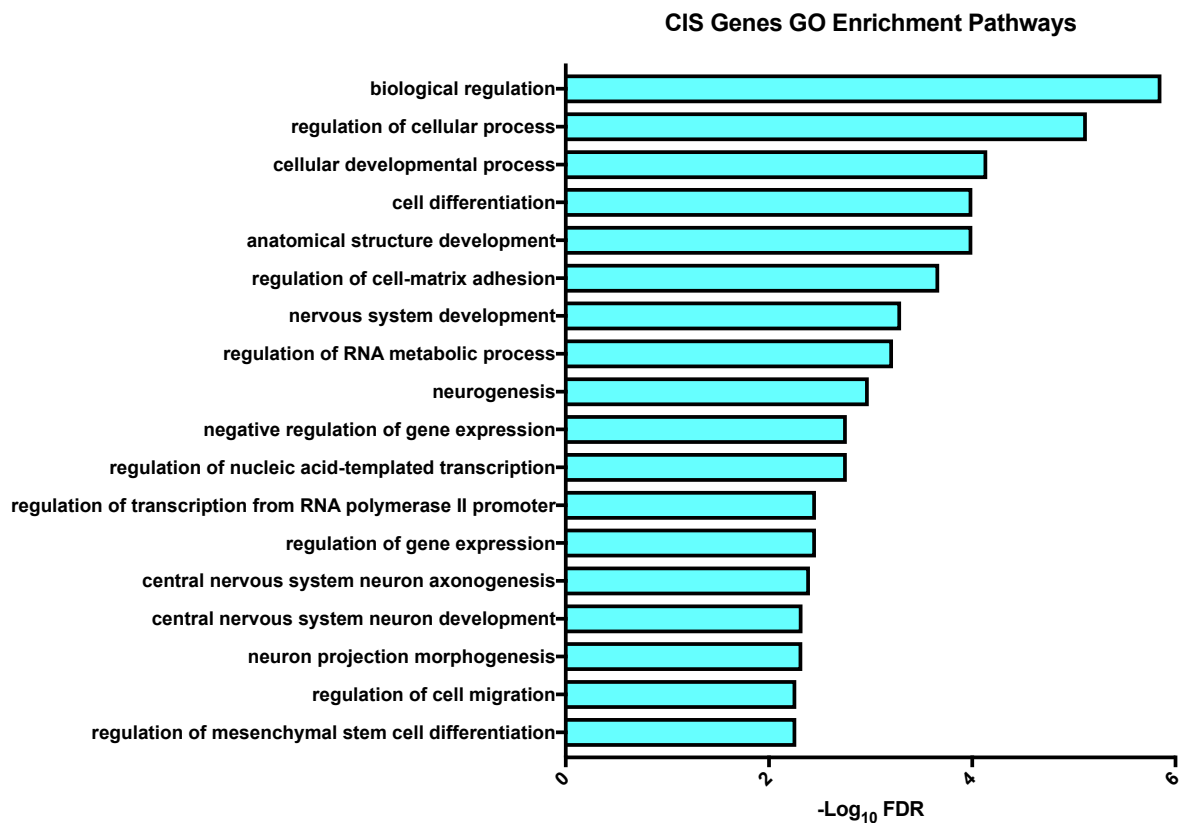
Fig 4.8. DAVID gene ontology (GO) analysis of all 281 glioma CIS genes shows significant enrichment for pathways including neurogenesis and mesenchymal stem cell differentiation, suggesting these pathways are important in driving *EGFR*-mutant gliomagenesis (FDR = false discovery rate). Fisher's exact 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.
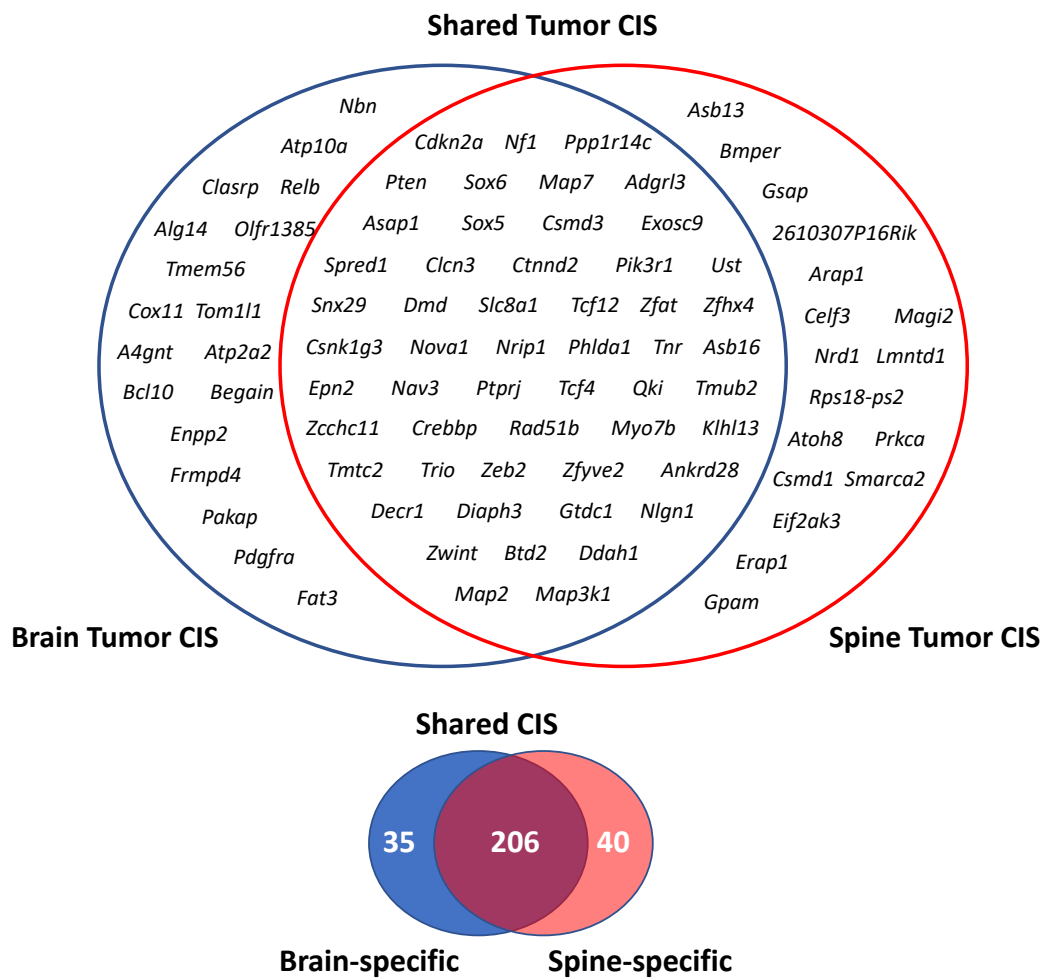
Fig 4.9. Brain and spinal gliomas share a core set of drivers. Upper Venn diagram shows the top genes from each tumor cohort, with core drivers including genes such as *Cdkn2a*, *Pten* and *Sox6.* Lower Venn diagram shows amongst all transposon CIS genes, brain and spinal cord tumors share 206 genes (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.
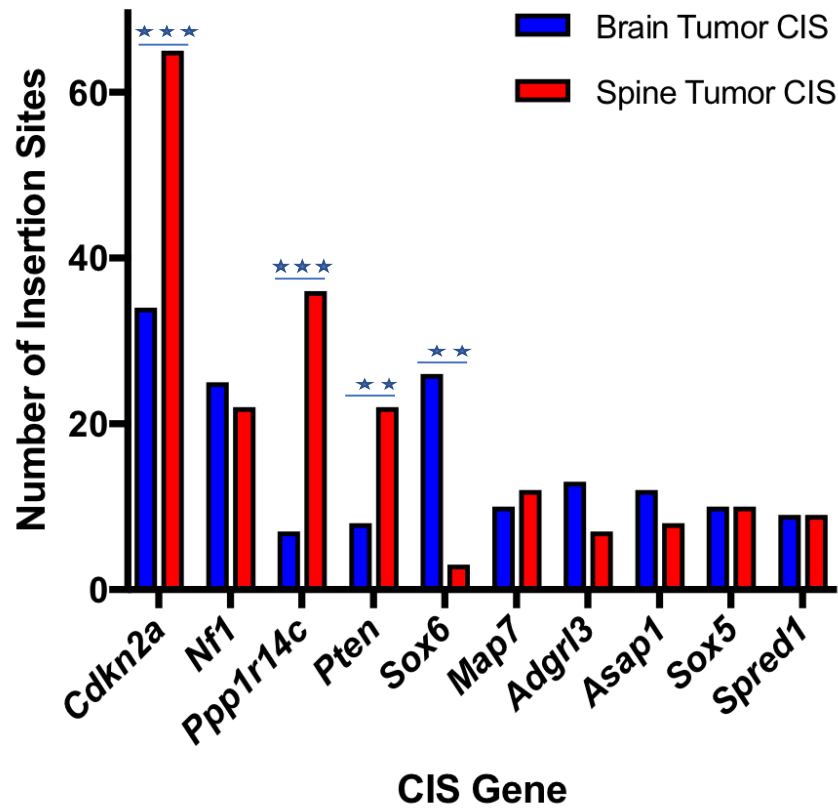
Fig 4.10. Bar plot comparing number of insertions between brain and spinal tumors for the top 10 CIS genes. *Cdkn2a*, *Ppp1r14c* and *Pten* have significantly more insertions (normalized for number of tumors analyzed) in spinal than brain tumors, and *Sox6* has more insertions in brain tumors (Fisher's 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
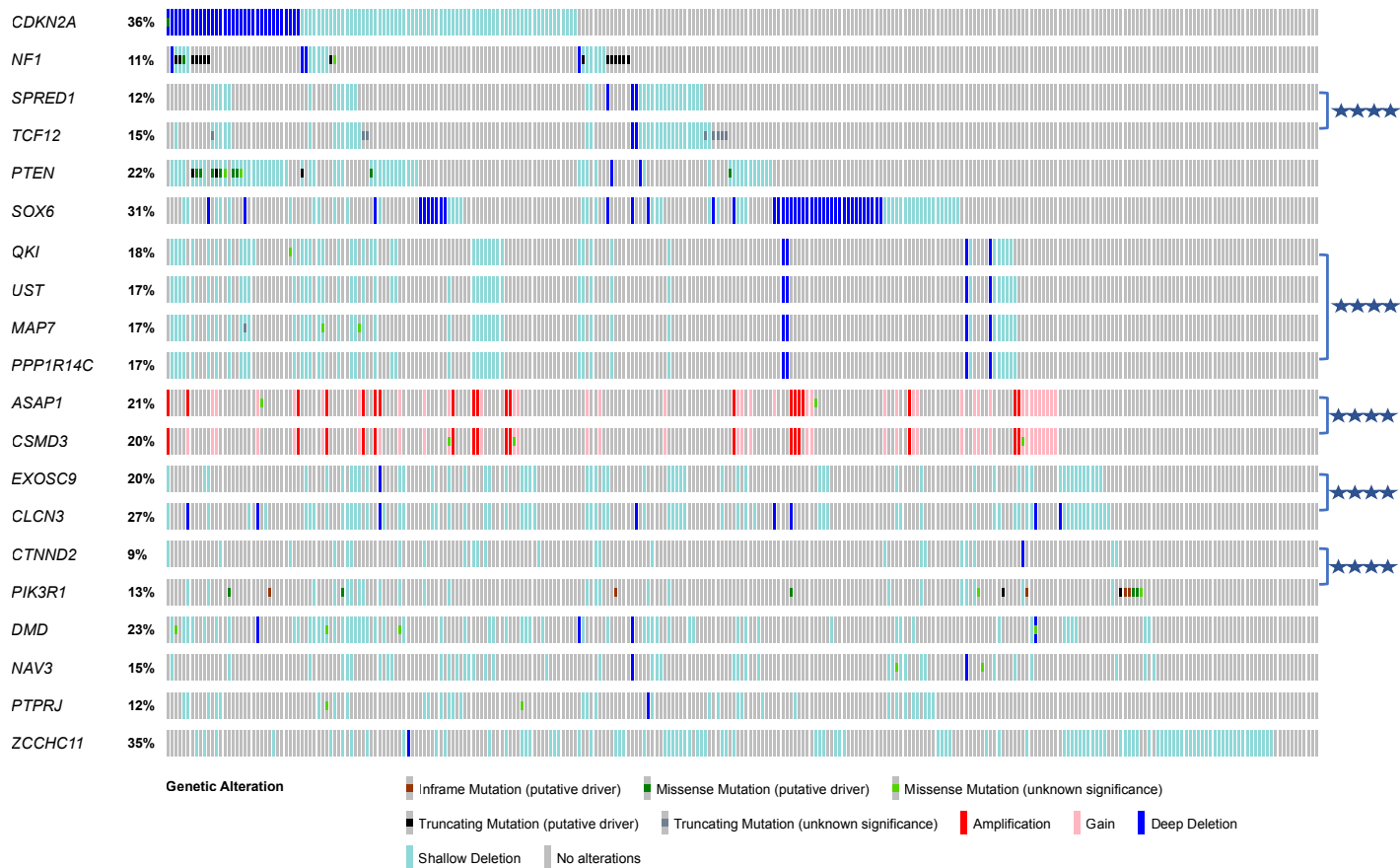
Fig 4.12. Top *PiggyBac* CIS genes are recurrently altered in human low-grade brain gliomas. Patient data was analyzed from The Cancer Genome Atlas (TCGA) dataset (n=283), for cross-comparison of the main CIS genes in mouse brain and spinal tumors. The frequency of alterations of CIS genes observed in 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

201

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/.
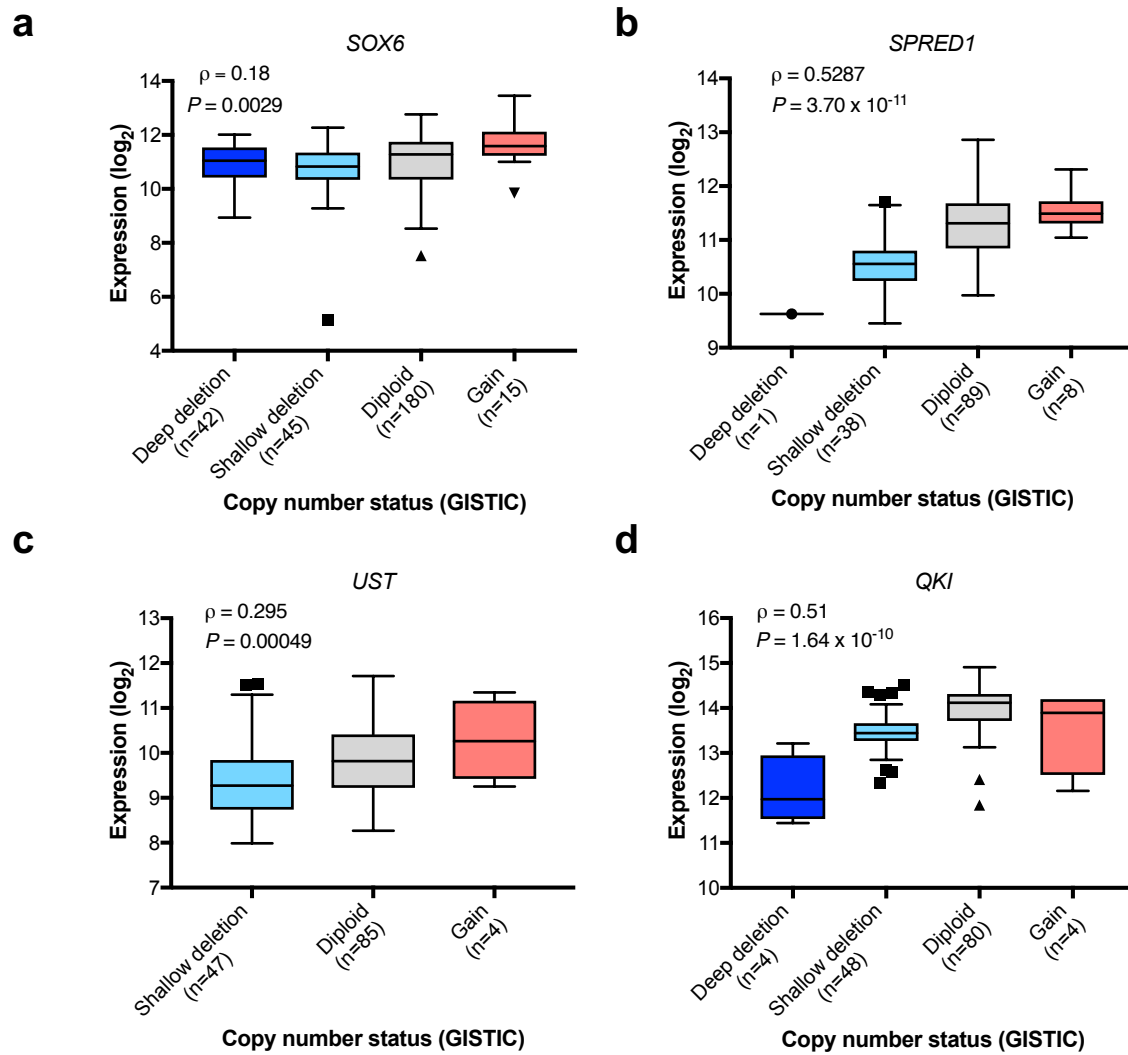
Fig 4.13. Deletions in putative tumor suppressors are associated with reduced gene expression.
A – D. Correlation of expression levels of *SOX6* in LGGs (A), *SPRED1* (B), *UST* (C) and *QKI* (D) in human patient GBMs with their respective copy number levels using the entire TCGA human datasets (RNA-seq 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.
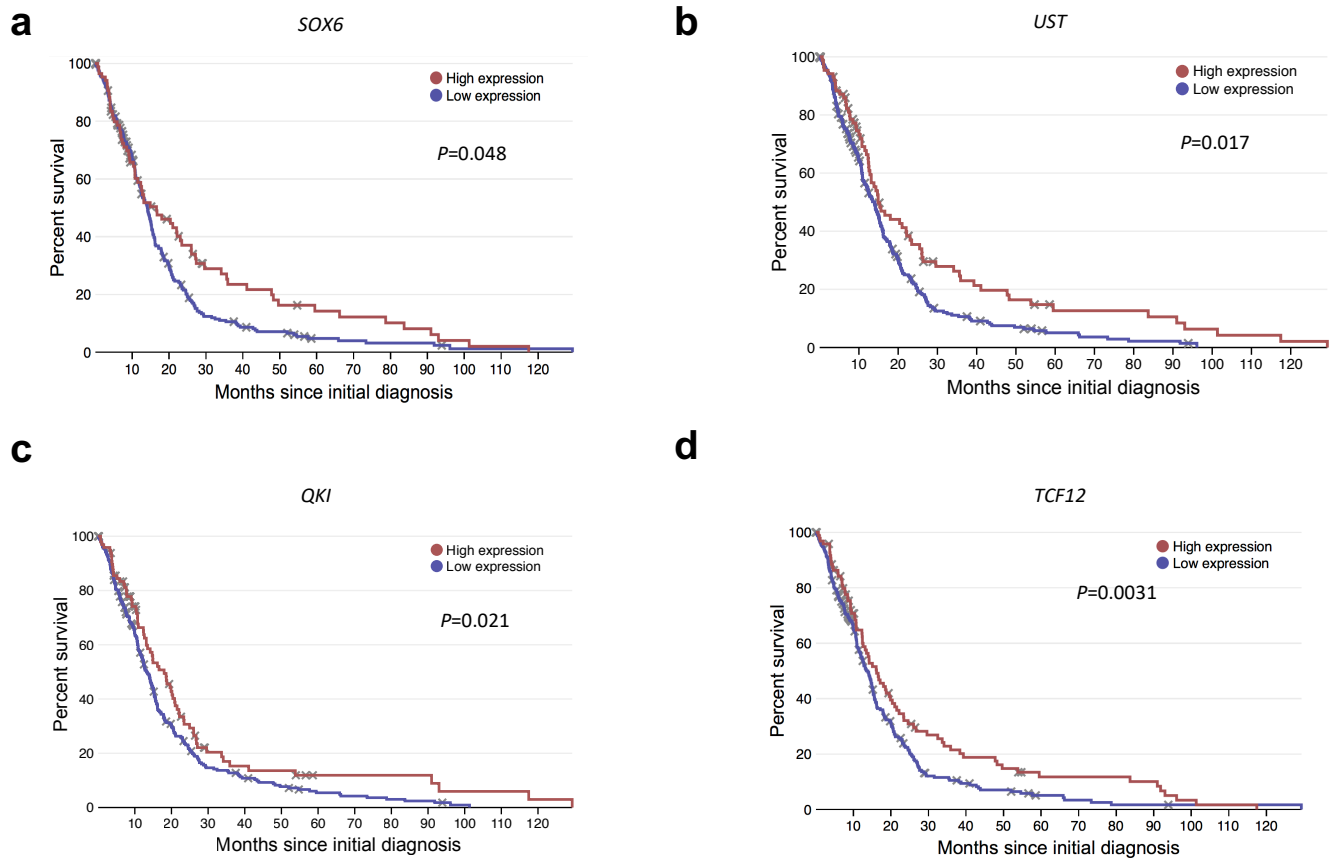
Fig 4.14. A-D. Kaplan-Meier plots of GBM patient survival in relation to expression levels of key CIS genes *SOX6* (a), *UST* (b), *QKI* (c) and *TCF12* (d). *P* values were calculated using the log-rank test comparing the top 30% of expression level with the lower 70% for each gene. The entire TCGA 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.
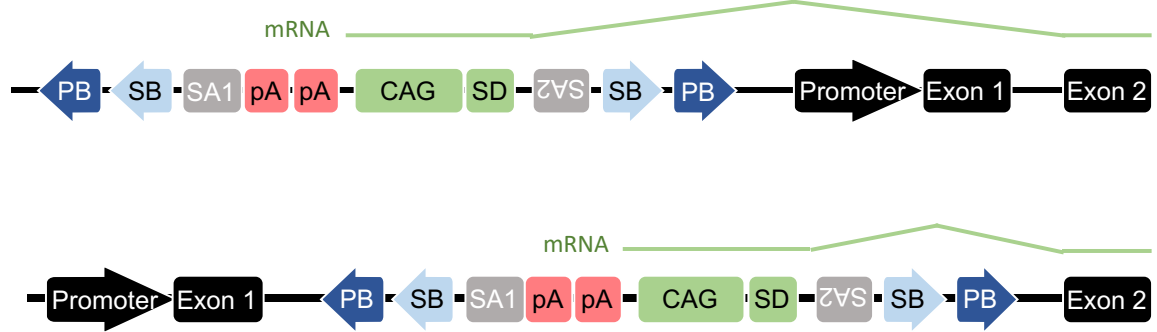

EGFRvIII; Nes-cre; ATP1S2; TSPB
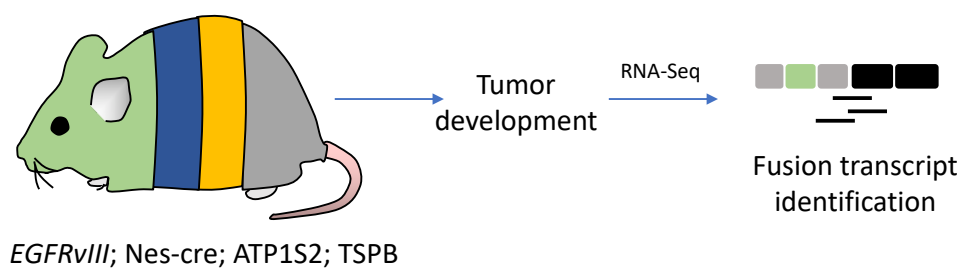
Fig 4.15. Effects of PB insertions on glioma transcriptomes. A. RNA-seq was performed on tumors from *EGFRvIII*-PB mice (n=36), with IM-Fusion[121] analysis of the data to identify fusion transcripts.

Oncogene Activation

Truncation (tumor suppressor or oncogene)
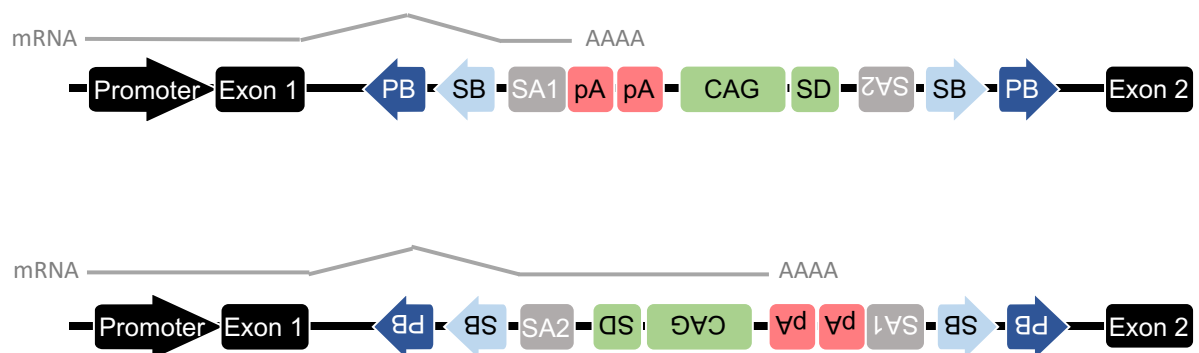
Fig 4.16. Overview of the effect ATP1-S2 transposons on the transcriptome: the transposon can insert in the sense orientation upstream of a gene's promoter or in an early intron, driving gene transcription through the transposon's promoter and splice donor (SD). Alternatively, it can cause transcript 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.

Shared, P = 4.43 X 10^{-11}
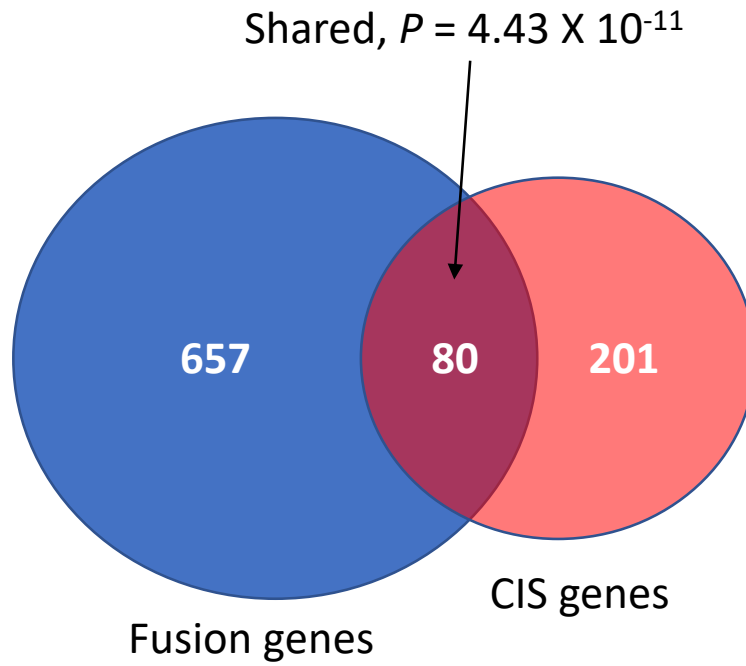
Fig 4.17. Of all genes with fusion transcripts, 80 genes overlapped with CIS genes identified by QI-seq. *P* value was calculated using a two-sided Fisher's exact test. All fusion transcripts detected with the carp-beta-actin splice acceptor (CbASA) and splice donor (SD) contained *PiggyBac* in the sense 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.
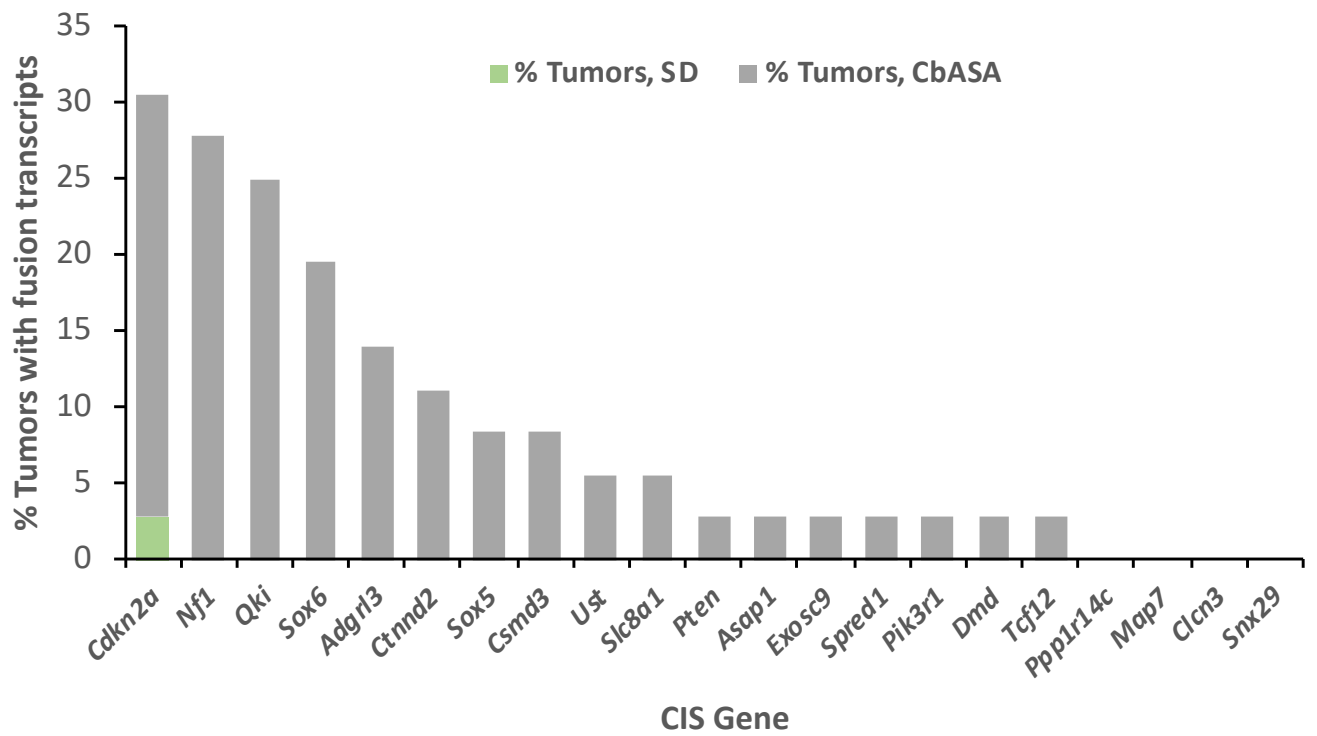
Fig 4.18. Bar plot showing percentage of gliomas with fusion transcripts amongst top 20 CIS genes (*Qki* is also included here).
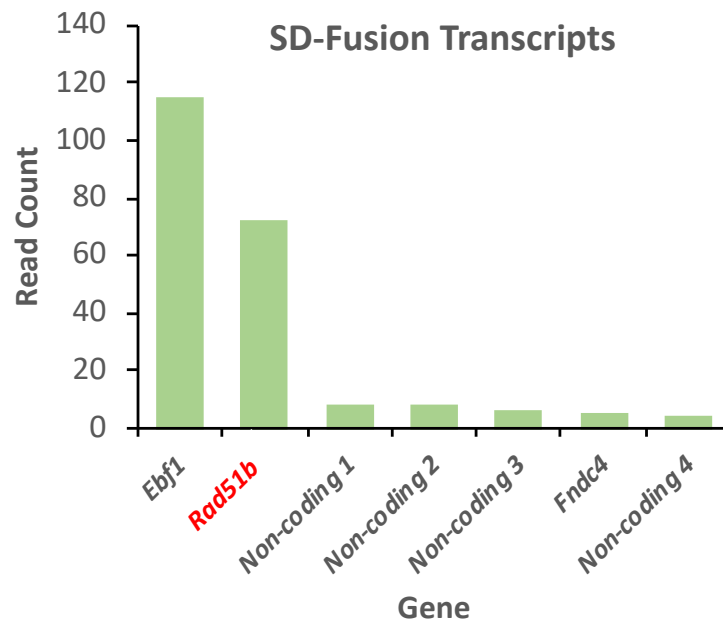
Fig 4.19. Bar plot showing the top fusion transcripts containing the PB splice donor ranked by read count; among them, only *Rad51b* was also identified as a CIS gene.

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.
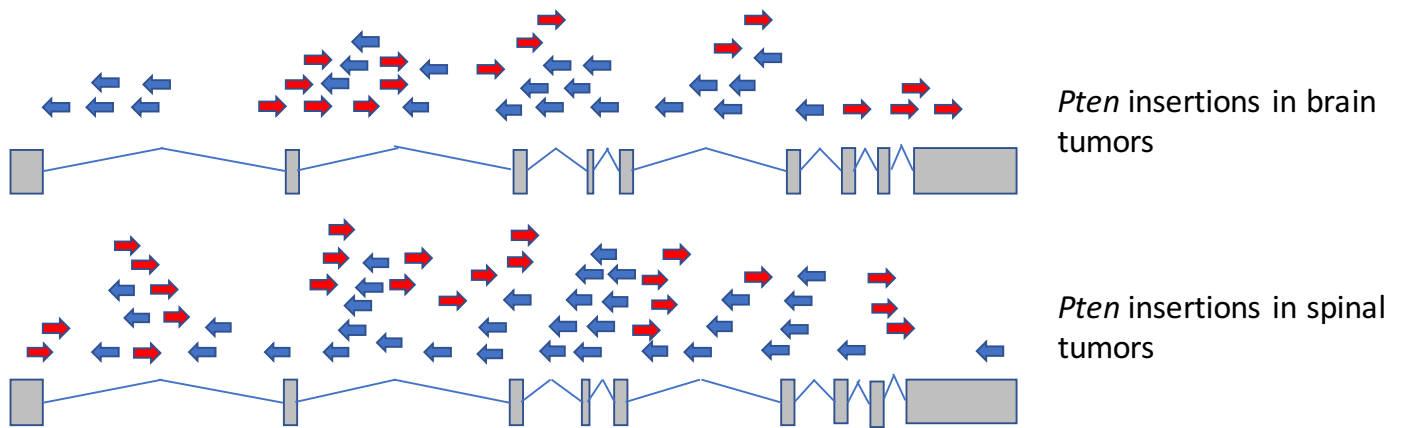
Pten insertions in brain tumors

Pten insertions in spinal tumors

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.
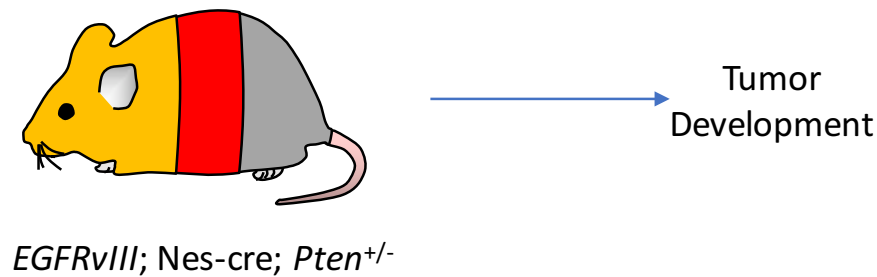
EGFRvIII/+ ; Pten⁺/⁻ Mice

EGFRvIII; Nes-cre; *Pten⁺/⁻*

Tumor
Development

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.
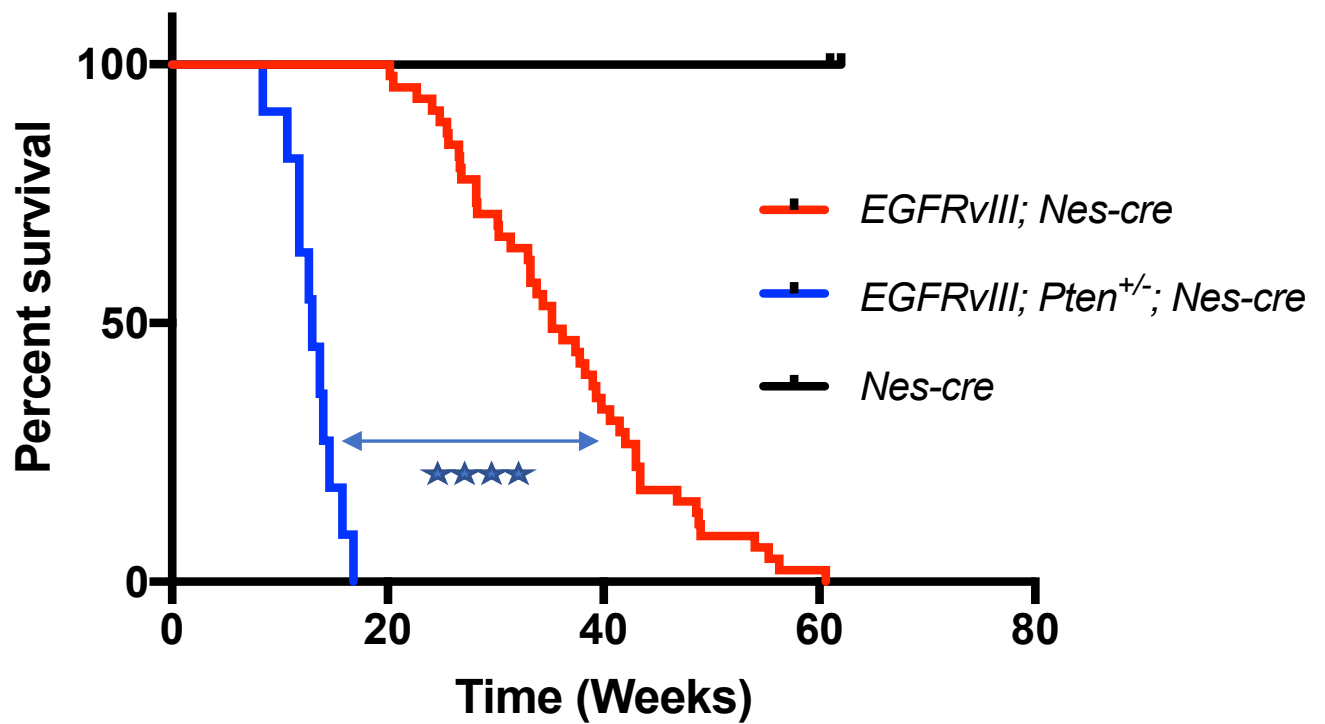
Fig 4.22. Kaplan Meier progression-free survival curves for *EGFRvIII Pten+/+* mice (red line, n=31) compared with *EGFRvIII Pten+/-* mice (blue line, n=11). There is a significantly shorter progression-free survival when there is loss of one PTEN allele because of signs of spinal cord compression due to spinal tumor development. **** denotes p < 0.0001, log-rank test.
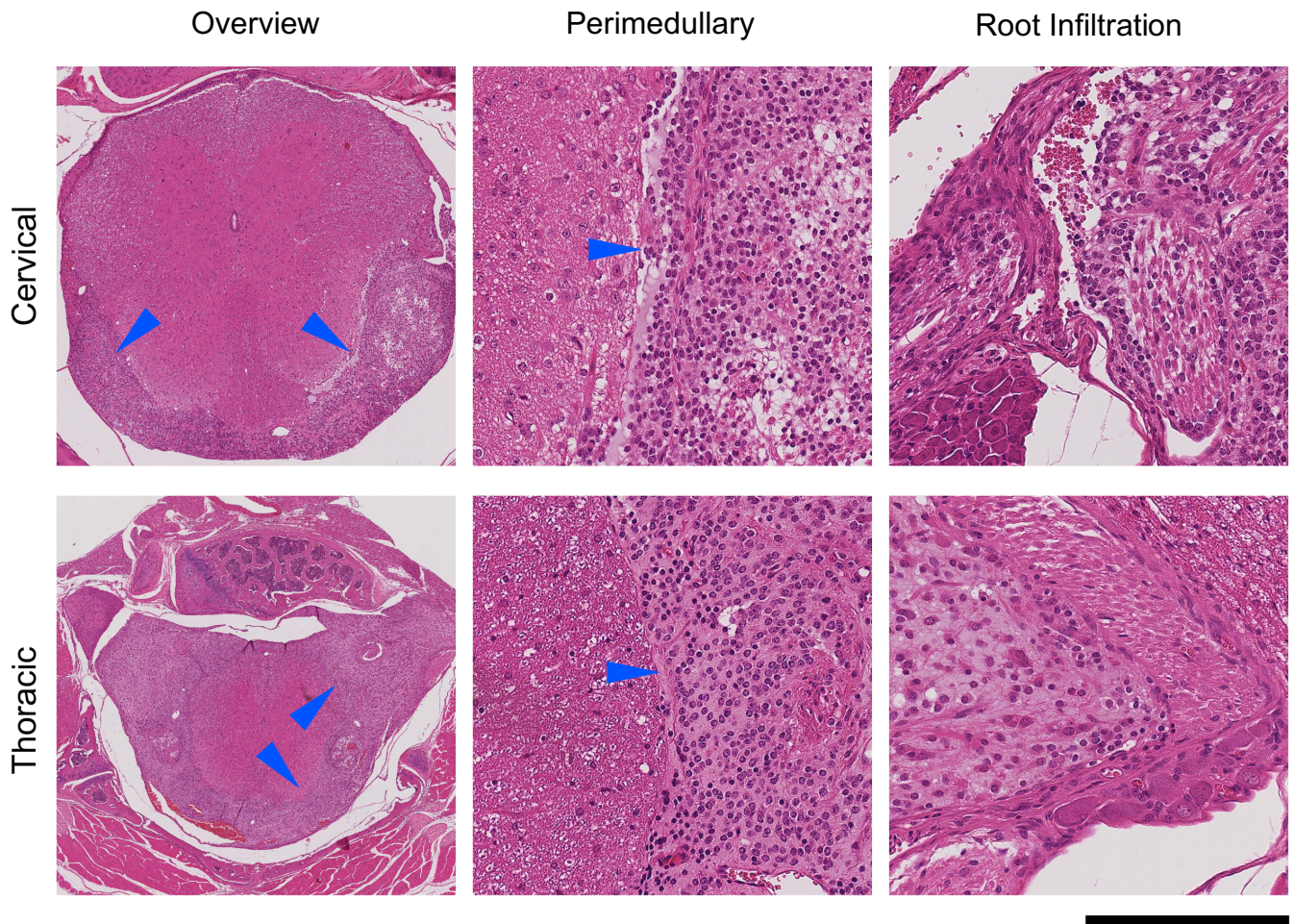
Fig 4.23. *EGFRvIII ; nes-cre; Pten^{+/-} spinal tumor growth and nerve root invasion.* Left panels show cervical and thoracic spinal cord with encasement by tumor cells growing within the subarachnoid space. Middle panels, detailed view of the spinal cord and tumor cells. Right panels, tumor cells invading 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

221

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*, *Pdgfrα* 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

223

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*.