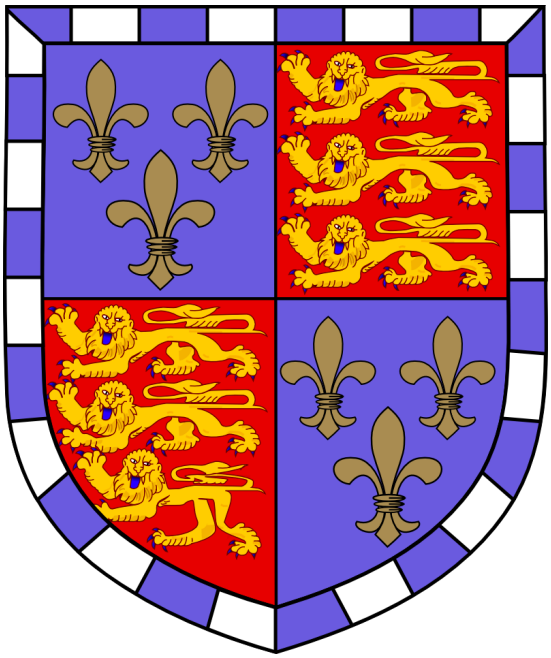
Genetic dissection of *EGFRvIII* brain and spinal mouse gliomas through whole-exome sequencing and *in vivo piggyBac* mutagenesis forward genetic screening



Imran Noorani

Christ's College, University of Cambridge

December 2018

This dissertation is submitted for the degree of Doctor of Philosophy

Preface

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

Publications

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

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

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

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

Acknowledgements

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

List of main corrections in this revised thesis:

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

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

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

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

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

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

21. Abbreviations have been explained at first usage.

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

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

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

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

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

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

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

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

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

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

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

33. Referencing format is now consistent across the Thesis.

Table of Contents

Summary of Work

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

Imran Noorani

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