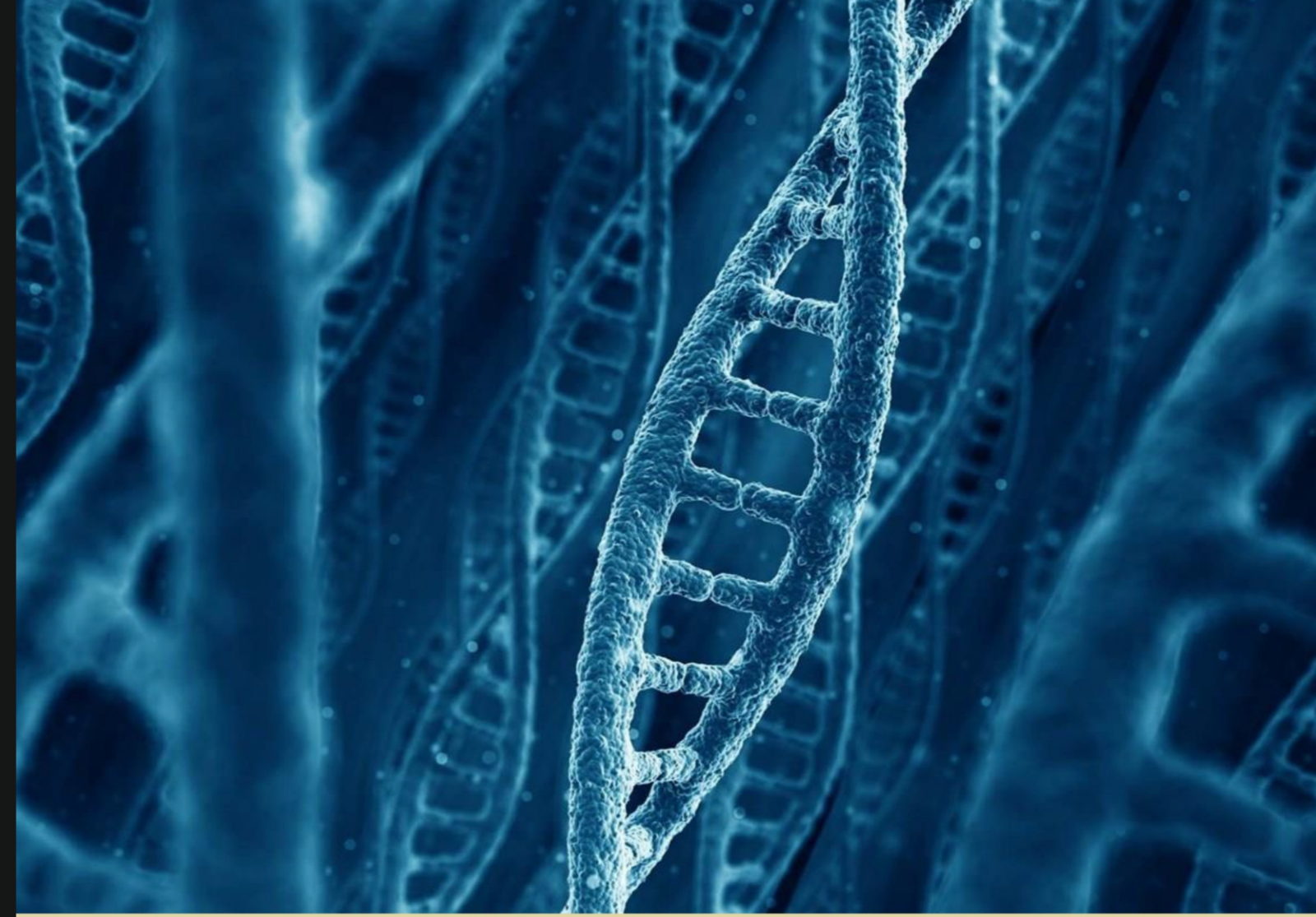


Diana Margarida Martins Carvalho

Identification of intragenic copy number alterations
and fusion genes in paediatric high grade glioma

UNIVERSIDADE DE COIMBRA



Diana Margarida Martins Carvalho

Identification of intragenic copy number alterations and fusion genes in paediatric high grade glioma

Tese de Doutoramento em Ciências e Tecnologias da Saúde (Pré-Bolonha), na especialidade de Biologia Celular e Molecular, orientada pelo Professor Doutor Chris Jones do Institute of Cancer Research, Londres, Reino Unido, co-orientada pelo Professor Doutor Rui Manuel Vieira Reis da Escola de Ciências da Saúde da Universidade do Minho (Braga, Portugal) e pela Professora Doutora Maria Celeste Lopes da Faculdade de Farmácia da Universidade de Coimbra e entregue à Faculdade de Farmácia da Universidade de Coimbra

Julho de 2014



UNIVERSIDADE DE COIMBRA



Diana Margarida Martins Carvalho

Identification of intragenic copy number alterations and fusion genes in paediatric high grade glioma

Identificação de alterações do número de cópias intragénicas e genes de fusão em gliomas pediátricos de alto grau

Tese de Doutoramento em Ciências e Tecnologia da Saúde
Especialidade de Biologia Celular e Molecular
(Pré-Bolonha)

Trabalho efectuado sob a orientação do

Professor Doutor Chris Jones

Team leader of the Glioma Team |
Divisions of Molecular Pathology and Cancer Therapeutics |
Institute of Cancer Research | London | United Kingdom

E co-orientação do

Doutor Rui Manuel Vieira Reis

Professor Auxiliar da Escola de Ciências da Saúde | Universidade
do Minho | Braga | Portugal

e da

Professora Doutora Maria Celeste Lopes

Faculdade de Farmácia |
Universidade de Coimbra | Coimbra | Portugal

Nome: Diana Margarida Martins Carvalho

Endereço electrónico: carvalho.dmm@gmail.com

Telefone: (+351) 968305649

Número do cartão de cidadão: 12797031

Título da Tese de Doutoramento:

Identificação de alterações do número de cópias intragénicas e genes de fusão em gliomas pediátricos de alto grau

Identification of intragenic copy number alterations and fusion genes in paediatric high grade glioma

Orientadores:

Professor Doutor Chris Jones

Professor Rui Manuel Vieira Reis

Professora Doutora Maria Celeste Lopes

Ano de conclusão: 2014

Designação do Ramo do Conhecimento e Especialidade

Ciências e Tecnologia da Saúde (Pré-Bolonha) – Biologia Celular e Molecular

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA TESE/TRABALHO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE

Universidade de Coimbra, ___/___/_____

Assinatura:



UNIVERSIDADE DE COIMBRA

The work presented in this thesis was developed during my PhD research project. The work was performed in the Institute of Cancer Research, London, United Kingdom, under the supervision of Doctor Chris Jones, Institute of Cancer Research, London, United Kingdom, and co-supervised by Doctor Rui Manuel Vieira Reis, Life and Health Science Research Institute, University of Minho, Braga, Portugal and by Doctor Maria Celeste Lopes, Faculty of Pharmacy, University of Coimbra, Portugal. This thesis has been supported by the grant SFRH / BD / 33473 / 2008 from Fundação para a Ciência e a Tecnologia, Lisboa, Portugal and by a Rosetrees Trust PhD Top Up fellowship, London, United Kingdom.

O trabalho apresentado nesta dissertação foi desenvolvido durante o meu projecto de doutoramento. O trabalho foi desenvolvido no Institute of Cancer Research, Londres, Reino Unido, orientado pelo Professor Doutor Chris Jones do Institute of Cancer Research em Londres, Reino Unido, e co-orientado pelo Professor Doutor Rui Manuel Vieira Reis, da Escola de Ciências da Saúde da Universidade do Minho e pela Professora Doutora Maria Celeste Lopes da Faculdade de Farmácia da Universidade de Coimbra. Este projecto foi financiado pela bolsa de doutoramento SFRH / BD / 33473 / 2008 da Fundação para a Ciência e a Tecnologia, Lisboa, Portugal e por uma bolsa de PhD Top Up da Rosetrees Trust, Londres, Reino Unido.

The following original articles have been published in peer-scientific journals, within the scope of the present PhD Thesis:

D. Carvalho, A. Mackay, L. Bjerke, R.G. Grundy, C. Lopes, R. M. Reis and C. Jones, The prognostic role of intragenic copy number breakpoints and identification of novel fusion genes in pediatric high grade glioma. *Acta Neuropathologica Communications*, 2014, 2(1): p23

S. Puget, C. Philippe, D. A. Bax, B. Job, P. Varlet, M. P. Junier, F. Andreiuolo, **D. Carvalho**, R. Reis, L. Guerrini-Rousseau, T. Roujeau, P. Dessen, C. Richon, V. Lazar, G. Le Teuff, C. Sainte-Rose, B. Georger, G. Vassal, C. Jones, and J. Grill, Mesenchymal Transition and *PDGFRA* Amplification/Mutation Are Key Distinct Oncogenic Events in Pediatric Diffuse Intrinsic Pontine Gliomas, *PLoS One*, 2012, 7(2): p e303

B.S. Paugh, X. Zhu, C. Qu, R. Endersby, J. Zhang, D. A. Bax, **D. Carvalho**, R.M. Reis, A. Onar-Thomas, A. Broniscer, C. Wetmore, C. Jones, J. Zhang, D. W. Ellison, S. J. Baker, Novel Oncogenic *PDGFRA* Mutations in Pediatric High-Grade Gliomas, *Cancer Research*, 2013, 73(20): p. 6219-29

Additional publications in the PhD time course:

D. A. Bax, A. Mackay, S. E. Little, **D. Carvalho**, M. Viana-Pereira, N. Tamber, A. E. Grigoriadis, A. Ashworth, R. M. Reis, D. W. Ellison, S. Al-Sarraj, D. Hargrave, and C. Jones, A Distinct Spectrum of Copy Number Aberrations in Pediatric High-Grade Gliomas, *Clinical Cancer Research*, 2010, 16(13), p. 3368-77.

L. Bjerke, A. Mackay, M. Nandhabalan, A. Burford, A. Jury, S. Popov, D. A. Bax, **D. Carvalho**, K.R. Taylor, M. Vinci, I. Bajrami, I. M. McGonnell, C.J. Lord, R.M. Reis, D. Hargrave, A. Ashworth, P. Workman, C. Jones, Histone H3.3 Mutations Drive Pediatric Glioblastoma through Upregulation of MYCN, *Cancer Discovery*, 2013, 3: p. 512-519

K. R. Taylor, A. Mackay, N. Truffaux, Y. Butterfield, O. Morozova, C. Philippe, C.S. Grasso, M. Vinci, **D. Carvalho**, A.M. Carvaboso, C. de Torres, O. Cruz, J. Mora, N. Entz-Werle N, W.J. Ingram, M. Monje, D. Hargrave, A.N. Bullock, S Puget, S Yip, C. Jones and J. Grill, Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma, *Nature Genetics*, 2014, 46(5): p. 457-61

Agradecimentos / Acknowledgements

The work described in this thesis would not be possible without the invaluable help and support of:

My primary supervisor Dr Chris Jones. I would like to thank you for encouraging my research and for allowing me to grow as a scientist, for your enthusiasm, motivation and immense knowledge, for believing in me (even when I didn't), and especially for always being available when I needed. Thank you!

My co-supervisors Dr Rui Manuel Reis and Dr Celeste Lopes that accepted me as PhD student, and gave me the opportunity to do this project. For all your words of motivation and encouragement throughout these 5 years. Thank you!

Do programa doutoral BEB08 por me ter dado a oportunidade de fazer doutoramento, e também aos meus colegas BEB do primeiro ano de doutoramento.

To my labmates, present and past.

In particular, I would like to thank Lara, Ola, Sergey, Lynley, Alexa, Ryan, Jane, and Suzie, who were always patient to teach me and always so nice to me. A special regard to Dorine for introducing me in the lab and helping me in my first year, to Lynn for all our science discussions which taught me so much, to Anna for all the help and knowledge (I couldn't do it without you!), to Alan that came to the lab in my last year and with his knowledge, infinite patience and passion for science taught me so much (I couldn't do it without you either!!!), and to Mara, the positive energy in this lab, thank you for unstoppable enthusiasm and dedication and for spreading it to everyone surrounding you. To my Portuguese mate, Marta which introduce me to all the people in the lab, helped me so much in my first year and since then became my dear friend. Obrigado. Thank you for all your friendship and support!!! To my dear friends and PhD mates, Katy and Meera, what would I do without you guys, thank your for all the chats of support/gossip, coffees, dances, for always being there when I needed, for being amazing friends 😊

Aos meus pais, António e Margarida, e a minha irmã Carla....tantas palavras de apoio e incentivo, este amor incondicional que não me deixou desistir, mesmo quando tudo parecia impossível! Um muito obrigado!

Ao Helder, a pessoa que me apoiou incondicionalmente, que nunca me deixou desistir, nem mesmo nos piores momentos, que sempre me deu forças para continuar esta jornada e sem

a qual seria impossível ter conseguido. Amo-te. Não há palavras para exprimir o quão importante és para mim ☺ Obrigado.

Às minhas Marias, TiJoana e Marlicas, as minhas irmãs de coração, minhas amigas de tantos anos. Espero que a nossa amizade continue para sempre assim. Obrigado por todas as conversas via skype, via gmail, por todas as palavras de apoio, e por mesmo com a distância que nos separa, continuarmos a ser amigas todos os dias ☺

*“You can, you should, and if you are brave
enough to start, you will”*

Stephen King

***Para os meus papás,
António Carvalho e
Margarida Martins***

Summary/Resumo

Summary

Paediatric brain tumours are the second most frequent malignancy in children, and the most common cause of cancer-related deaths in both the 0-14-year and the 15-24-year age group. In this thesis we focused our studies in paediatric malignant brain tumours of glial origin (grade III and IV): astrocytomas, oligodendrogliomas and glioblastomas. Although, paediatric high grade glioma (pHGG) is a histologically similar tumour to that arising in adults, these are distinct biological diseases, differing in copy number profiles and driver genetic alterations. Recent sequencing initiatives have conclusively shown the existence of subgroups of HGG marked by distinct driver mutations, which are significantly enriched in young children (*H3F3A* K27M), teenagers and young adults (*H3F3A* G34R/V), and middle-aged adults (*IDH1/2*). Structural rearrangements resulting in novel fusion genes are strongly associated with cancer, and numerous examples exist in both adult and childhood malignancies. Although, only few have been described in pHGG. Structural variants (SV) frequently result in chimeric proteins targetable by novel therapeutic approaches, an outcome desperately needed in pHGG. The work summarized in this thesis aims to explore novel structural rearrangements in childhood malignant gliomas, contributing to uncover the molecular mechanisms underpinning pHGG, to further contribute to distinguish the adult and childhood disease.

PDGFRA is a receptor tyrosine kinase (RTK) that triggers essential cellular responses such as proliferation, migration, and survival. This gene is commonly amplified and overexpressed in paediatric malignant gliomas, playing an important role in this disease. To determine whether *PDGFRA* was also targeted by more subtle mutations missed by copy number analysis, we screened a large series of cases for single base changes and small *indels*, as well as the *PDGFRA* Δ 8,9 deletion and *KDR:PDGFRA* (*KP*) fusion gene, previously reported in adult high grade glioma (aHGG). Somatic-activating mutations were identified in nonbrainstem pHGG and Diffuse Intrinsic Pontine Gliomas (DIPG), including missense mutations and in-frame deletions and insertions not previously described. In our studies *PDGFRA* alterations were more common in pHGG arising outside the brainstem (14% vs 6%), and 8/18 (44%) cases had concomitant amplification and mutation of *PDGFRA*. The adult rearrangements (*KP* and *PDGFRA* Δ 8,9) were only found in adult cases and one case of pHGG. Thus, a distinct spectrum of *PDGFRA* alterations is present in pHGG.

As functionally important intragenic copy number aberrations (iCNA) and fusion genes begin to be identified in aHGG, the same has not yet been done in the childhood setting. We applied an iCNA algorithm to our previously published dataset of DNA copy number profiling

in pHGG with a view to identify novel intragenic breakpoints. We reported a series of 288 iCNA events in pHGG, with the presence of intragenic breakpoints itself a negative prognostic factor. We identified an increased number of iCNA in older children compared to infants, and increased iCNA in *H3F3A* K27M mutant tumours compared to G34R/V and wild-type. We observed numerous gene disruptions by iCNA due to both deletions and amplifications, targeting known HGG-associated genes such as *RB1* and *NF1*, putative tumour suppressors such as *FAF1* and *KIDINS220*, and novel candidates such as *PTPRE* and *KCND2*. We further identified two novel fusion genes in pHGG – *CSGALNACT2:RET* and the complex fusion *DHX57:TMEM178:MAP4K3*.

To further examine whether SVs that lead to the production of oncogenic fusion proteins were present in pHGG, we used whole genome and transcriptome paired end sequencing to detect novel gene fusions in pHGG model cell lines – KNS42, SF188 and UW479. Our study showed for the first time that glioma cell lines are highly rearranged and that they are characterized by several fusion genes. We discovered three extra-chromosomal structural rearrangement structures in SF188, which involve chromosomes 4q12 (*SCFD2*, *FIP1L1*), 8q24 (*MYC*), 11p11, 11q13 (*CCND1*), 11p14, 11q23 (*MLL*) and 12q14 (*CDK4*). The amplified loci represented three very complex structures, which were present in the form of three double minutes (DM). We characterized a fusion gene in each cell line: *GORASP2:CDADC1* in KNS42, *NUBPL:AKAP6* in UW479 and *RPTOR:TULP4* in SF188. The *RPTOR:TULP4* fusion gene described in SF188 involved *RPTOR*, an important component of mTOR signalling. This fusion leads to the expression of a truncated form of *RPTOR* in SF188. We discovered that *RPTOR* was disrupted 2.2% of pHGG cases and that patient samples harboured similar truncations to the SF188 cell line. As the roles of these fusion genes are still unclear, further studies need to be performed in order to better understand their role in pHGG.

In summary these data expand upon our understanding of the genomic events driving these tumours and represent novel targets for therapeutic intervention in these poor prognosis cancers of childhood.

Resumo

Os tumores cerebrais pediátricos são a segundo tumor maligno mais comum em crianças. Estes tumores são a principal causa de morte por cancro nos grupos dos 0-14 e dos 15-24 anos. Nesta tese, centramos os nossos estudos em tumores pediátricos malignos de origem glial (grau III e IV): astrocitomas, oligodendrogliomas, e glioblastomas (GBM). Apesar dos gliomas pediátricos de alto grau serem histologicamente semelhantes aos gliomas adultos malignos, estes são doenças biologicamente diferentes, possuindo perfis de número de cópias de ADN e alterações genéticas diferentes. As recentes iniciativas de sequenciação de última geração demonstraram a existência de sub-grupos de gliomas de alto grau que são caracterizados por mutações distintas: nas crianças (*H3F3A* K27M), nos adolescentes e jovens adultos (*H3F3A* G34R/V) e nos adultos de meia-idade (*IDH1/2*). As aberrações estruturais dos cromossomas dão origem à formação de genes de fusão que estão normalmente associados com cancro, existindo vários casos descritos em cancros de crianças e adultos. No entanto, nos gliomas pediátricos de alto grau apenas alguns genes de fusão foram descritos. Estes rearranjos estruturais normalmente dão origem a proteínas quiméricas que potencialmente podem ser importantes na resposta tumoral à terapia dirigida, algo extremamente necessário para o tratamento desta doença. O trabalho resumido nesta tese pretende explorar os mecanismos moleculares que estão por detrás dos gliomas pediátricos de alto grau, contribuindo desta forma para a distinção entre os gliomas malignos pediátricos e adultos.

O *PDGFRA* é um receptor de tirosina cinase (RTK) que activa várias respostas celulares, tais como a proliferação, a migração e a sobrevivência das células. Este gene está normalmente amplificado e é sobre-expresso em gliomas malignos pediátricos, desempenhando um papel muito importante nesta doença. De forma a determinar se o *PDGFRA* é alvo de mutações que não são detectadas através da análise do número de cópias de ADN, decidimos estudar a presença de alterações moleculares do *PDGFRA* (mutações pontuais e pequenas inserções/deleções), assim como a deleção $\Delta 8,9$ no *PDGFRA* e o gene de fusão *KDR:PDGFRA* (*KP*), anteriormente descritos em gliomas malignos em adultos. Os nossos resultados demonstraram a presença de mutações somáticas de ganho de função, incluindo mutações *missense* e deleções/inserções *in-frame* que não tinham sido descritas anteriormente na literatura. Estas alterações estão presentes em ambos gliomas pediátricos supratentoriais de alto grau e em gliomas pontinos difusos (DIPG). Os nossos estudos demonstraram que as alterações no *PDGFRA* são mais comuns em gliomas supratentoriais malignos do que em DIPG, e 8/18 (44%) dos casos possuem simultaneamente amplificação e mutação no *PDGFRA*. Os rearranjos estruturais

normalmente presentes em adultos (*KP* e *PDGFRA* Δ 8,9) foram apenas encontrados num caso de glioma pediátrico de alto grau. Estes resultados indicam que as alterações presentes no *PDGFRA* em crianças são diferentes daquelas presentes em adultos.

Enquanto que as alterações do número de cópias intragénicas (iCNA) e de genes de fusão funcionalmente importantes começam a ser identificados em gliomas de alto grau em adultos, o mesmo ainda não ocorreu nos gliomas pediátricos. Ao aplicar o algoritmo iCNA a um conjunto de dados de perfis do número de cópias de ADN em casos de gliomas pediátricos malignos, previamente publicado pelo nosso grupo, identificamos novas alterações intragénicas. Neste estudo, identificamos 288 eventos iCNA em gliomas pediátricos de alto grau, sendo que a presença destes *breakpoints* intragénicos é por si só um factor de prognóstico negativo. Quando comparadas com crianças mais jovens, os adolescentes possuem maior número de iCNA. Os tumores que têm a mutação *H3F3A* K27M também apresentam maior número de iCNA quando comparados com os tumores que tem a mutação *H3F3A* G34R/V ou com os casos normais. No nosso estudo também observamos inúmeras disrupções de genes por iCNA devido a deleções e amplificações. Os genes alvo destas disrupções estão associados a gliomas malignos, tais como *RB1* e *NF1*, a supressores tumorais putativos, como por exemplo o *FAF1* e o *KIDINS220*, e também a novos candidatos tumorais, como o *PTPRE* e *KCND2*. Identificamos também dois novos genes de fusão, *CSGALNACT2:RET* e *DHX57:TMEM178:MAP4K3*.

De forma a explorar a ocorrência de alterações cromossómicas estruturais que dão origem a proteínas de fusão oncogénicas, utilizamos a sequenciação de ultima geração do ADN e ARN de linhas celulares de gliomas pediátricos de alto grau – KNS42, SF188 e UW479. O nosso estudo mostrou pela primeira vez que estas linhas celulares são caracterizadas por vários genes de fusão. Descobrimos ainda a presença de três estruturas extra-cromossomais presentes na linha celular SF188. Estas estruturas envolvem os cromossomas 4q12 (*SCFD2*, *FIP1L1*), 8q24 (*MYC*), 11p11, 11q13 (*CCND1*), 11p14, 11q23 (*MLL*) e o 12q14 (*CDK4*). Estas regiões estão amplificadas e estão presentes em três estruturas muito complexas sob a forma de *double minutes* (DM). Caracterizamos ainda um gene de fusão por cada linha celular: o *GORASP2:CDADC1* na KNS42, o *NUBPL:AKAP6* na UW479 e o *RPTOR:TULP4* na SF188. O gene de fusão *RPTOR:TULP4* descrito na SF188 envolve o gene *RPTOR*, um importante componente da via de sinalização do mTOR, e leva a expressão de uma forma truncada do RPTOR. Esta quimera foi encontrada em 2.2% dos gliomas pediátricos de alto grau

Concluindo, o trabalho resumido nesta tese amplia o nosso conhecimento sobre os eventos genómicos característicos destes tumores, e representa novos e potenciais *targets* para a terapia dirigida, numa doença que ainda não tem cura.

Contents

Contents

ABBREVIATIONS	xxiii
THESIS LAYOUT.....	xxvii
1. INTRODUCTION	3
1.1. Incidence and general overview.....	3
1.1.1. Histological classification and clinical features of central nervous system tumours	3
1.1.2. Malignant glioma epidemiology	6
1.1.3. Current therapy for paediatric high grade glioma	8
1.2. Molecular biology of high grade glioma.....	11
1.2.1. Molecular profile of paediatric high grade glioma.....	11
1.3. Somatic structural variation in cancer	18
1.3.1. Mechanism of structural variation	20
1.3.2. Biological impact of structural variants	21
1.3.3. Technologies for measurement of structural variants	21
1.3.4. General overview of fusion genes in cancer	23
1.3.5. Gene fusions in brain tumours	25
1.4. References	27
2. AIMS	37
3. <i>PDGFRA</i> ALTERATIONS IN PAEDIATRIC HIGH GRADE GLIOMA.....	42
3.1. Introduction.	42
3.2. Materials and Methods.....	45
3.3. Results	48
3.4. Discussion.....	53
3.5. References.....	57

4. INTRAGENIC COPY NUMBER BREAKPOINT ANALYSIS IDENTIFIES A NOVEL COMPLEX FUSION <i>DHX57:TMEM178:MAP4K3</i> IN PAEDIATRIC HIGH GRADE GLIOMA	64
4.1. Introduction	64
4.2. Material and Methods.....	65
4.3. Results	67
4.4. Discussion.....	76
4.5. References.....	79
5. INTEGRATED WHOLE GENOME AND RNA SEQUENCING IDENTIFIES NOVEL EXPRESSED FUSION TRANSCRIPTS IN PAEDIATRIC HIGH GRADE GLIOMA	83
5.1. Introduction	83
5.2. Material and Methods.....	84
5.3. Results	92
5.4. Discussion.....	113
5.5. References.....	116
6. <i>RAPTOR</i> REARRANGEMENTS IN PAEDIATRIC HIGH GRADE GLIOMA.....	123
6.1. Introduction	123
6.2. Material and Methods.....	125
6.3. Results	127
6.4. Discussion.....	134
6.5. References.....	136
7. CONCLUSIONS AND FUTURE PERSPECTIVES.....	141
7.1. PDGFRA alterations in paediatric high grade glioma.....	141
7.2. Intragenic copy number aberrations and structural variants in paediatric high grade glioma.....	143
7.3. References.....	146

8. APPENDIX.....	151
Appendix I	151
Appendix II	163
Appendix III	181
Appendix IV	203
Appendix V	219
Appendix VI	231
9. ADDITIONAL PUBLICATIONS IN THE PHD TIME COURSE.....	235
Paper I - D. A. Bax, A. Mackay, S. E. Little, <u>D. Carvalho</u> , M. Viana-Pereira, N. Tamber, A. E. Grigoriadis, A. Ashworth, R. M. Reis, D. W. Ellison, S. Al-Sarraj, D. Hargrave, and C. Jones, A Distinct Spectrum of Copy Number Aberrations in Pediatric High-Grade Gliomas, Clinical Cancer Research, 2010, 16(13), p. 3368-77.	
Paper II - L. Bjerke, A. Mackay, M. Nandhabalan, A. Burford, A. Jury, S. Popov, D. A. Bax, <u>D. Carvalho</u> , K.R. Taylor, M. Vinci, I. Bajrami, I. M. McGonnell, C.J. Lord, R.M. Reis, D. Hargrave, A. Ashworth, P. Workman, C. Jones, Histone H3.3 Mutations Drive Pediatric Glioblastoma through Upregulation of MYCN, Cancer Discovery, 2013, 3: p. 512-519	
Paper III - K. R. Taylor, A. Mackay, N. Truffaux, Y. Butterfield, O. Morozova, C. Philippe, C.S. Grasso, M. Vinci, <u>D. Carvalho</u> , A.M. Carvaboso, C. de Torres, O. Cruz, J. Mora, N. Entz-Werle N, W.J. Ingram, M. Monje, D. Hargrave, A.N. Bullock, S Puget, S Yip, C. Jones and J. Grill, Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma, Nature Genetics, 2014, 46(5): p. 457-61	

Abbreviations

aa – amino acid

aCGH array comparative genomic hybridization

ACVR1 – activin A receptor, type I

aHGG – adult high grade glioma

bp – base pair

BRAF - b-raf proto-oncogene, serine/threonine kinase

CCND1/ CCND2/CCND3 – cyclin D1/ D2 / D3

CDK4/CDK6 – cyclin dependent kinase 4 / 6

CDKN1C/CDKN2A/CDKN2B/CDKN2C - cyclin-dependent kinase inhibitor 1C / 2A / 2B / 2C

Chr - chromosome

CNV – Copy Number Variant

CNS – central nervous system

CTX – inter-chromosomal translocation

DEL – deletion

DIPG – diffuse intrinsic pontine glioma

DNA - deoxyribonucleic acid

DSB – double strand break

EGFR – epidermal growth factor receptor

FFPE – formalin fixed paraffin embedded

FGFR1 – fibroblast growth factor receptor 1

FISH – fluorescent *in situ* hybridization

G-banding – Giemsa-banding

GBM – glioblastoma

G-CIMP – glioma-CpG island methylator phenotype

GIST – gastrointestinal stromal tumours

HGG – high grade glioma

iCNA – intragenic copy number aberration

Ig-like – immunoglobulin-like domain

Indels – insertions and deletions

Ins – insertion

Inv – inversion

ITX – intra-chromosomal translocation

Kb – kilobase

KDR – kinase insert domain receptor (a type III receptor tyrosine kinase)

KP – *KDR:PDGFRA*

LCR – low copy repeat
LGG – low grade glioma
MAPK – mitogen activated kinase-like protein
Mb – megabase
MET – MET proto-oncogene, receptor tyrosine kinase
MYC – v-myc avian myelocytomatosis viral oncogene homolog
MYCN – v-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog
NF1 – neurofibromin 1
NHEJ – non-homologous end joining
NSCLC – non-small cell lung cancer
NTRK – neurotrophic tyrosine receptor kinase
PDGFA/PDGFB – platelet-derived growth factor alpha / beta
PDGFRA/PDGFRB – platelet-derived growth factor receptor alpha / beta
pHGG – paediatric high grade glioma
PI3K – phosphatidylinositol-4,5-bisphosphate 3-kinase
PTEN – phosphatase and tensin homolog
Rb/Rb1 – retinoblastoma / retinoblastoma 1
RTK – receptor tyrosine kinase
SNP array – single nucleotide polymorphism array
SNV – single nucleotide variant
STAT3 – signal transducer and activator of transcription 3
SV – structural variant
TACC3 – transforming, acidic coiled-coil containing protein 3
TCGA – The Cancer Genome Atlas
TERT – telomerase reverse transcriptase
TK – tyrosine kinase
TMZ – temozolomide
TP53 – tumour protein p53
VEGFA – vascular endothelial growth factor alpha
VEGFR – vascular endothelial growth factor receptor
WGA – whole genome amplification
WGS – whole genome sequencing

Thesis Layout

Thesis Layout

Chapter 1 is a general introduction to paediatric high grade gliomas (pHGG), the disease explored during my PhD, and the occurrence of structural variants (SVs) in cancer and pHGG, describing the most up-to-date relevant literature for both topics.

Chapter 2 describes the general aims of this thesis.

Chapter 3 presents two studies which report the characterization of *PDGFRA* in a series of pHGG, looking at *PDGFRA* mutation status and structural rearrangements in the context of paediatric glioblastoma.

Chapter 4 comprises a study which identified numerous potentially functional gene disruptions and a novel validated complex fusion, *DHX57:TMEM178:MAP4K3* in a pHGG cohort, using an algorithm designed to identify intragenic copy number breakpoints.

Chapter 5 describes the genomic characterization of the few pHGG models commonly used for the study of this disease. The structural rearrangements present in KNS42, SF188 and UW479 are described and validated using different techniques.

Chapter 6 reports a rearrangement found in SF188, *RPTOR:TULP4*. This fusion gene is described, validated and further screened in a series of paediatric and adult HGG.

Chapter 7, the final chapter, discusses the major findings reported in this thesis, supported by the most up-to-date literature in pHGG and structural rearrangements.

1. Introduction

1. INTRODUCTION

1.1. Incidence and general overview

Cancer is the third leading cause of death in the world and the first in economically developed countries [1, 2]. Worldwide, based on GLOBOCAN 2012, approximately 14.1 million new cases of cancer and nearly 8.2 million cancer-related deaths are estimated to have occurred within the space of 12 months (excluding non-melanoma skin cancer) [2-4]. Regarding the paediatric population, cancer is the leading cause of death by disease in the United States [5]. Approximately 14,276 new cases of paediatric cancers were diagnosed in children between 0–19 years in 2010, with nearly 2000 deaths estimated to occur [6]. In the UK, between 2006-2008, approximately 850 children and young adults (0-24 years) were newly diagnosed with cancer [7]. Despite an overall improvement in survival of both paediatric and adult tumours, some cancers have had no changes in decades.

1.1.1. Histological classification and clinical features of central nervous system tumours

Brain tumours are classified according to The World Health Organization (WHO) criteria, according to the morphological resemblance to the differentiated cells in the brain, cytoarchitecture and immunohistological marker profile [8, 9]. In addition to a morphological grouping, the WHO classification system also clusters tumours according to its grade – I, II, III and IV, in ascending order of malignancy. Higher grade tumours (grade III and IV) have a worst clinical outcome and are histologically characterized by the presence of nuclear atypia, increased proliferation, microvascular proliferation and necrosis. Lower grade tumours (grade I and II) have low proliferative potential and the possibility of cure after surgical resection [8, 9]. This thesis will focus on high grade tumours of glial origin, astrocytomas (grade III), oligodendroglioma (grade III and IV) and glioblastoma (GBM) (grade IV) in the paediatric setting (Figure 1.1).

High grade gliomas (HGG) are histologically characterized by microvascular proliferation, cellular atypia, elevated mitotic activity and high cell density (Figure 1.2). The grade IV glioma, also known as GBM (previously *multiforme*, GBM), is the most aggressive variant of diffuse glioma with a very poor prognosis of approximately 1 year, which may be extended by a few months using surgery and radiochemotherapy [10, 11]. These lesions show increased vascular endothelial proliferation, necrosis as well as a very high cell density and morphological atypia (Figure 1.2C). Mostly it occurs *de novo* – primary GBM – but it can also evolve from a preceding low-grade lesion, described as secondary GBM.

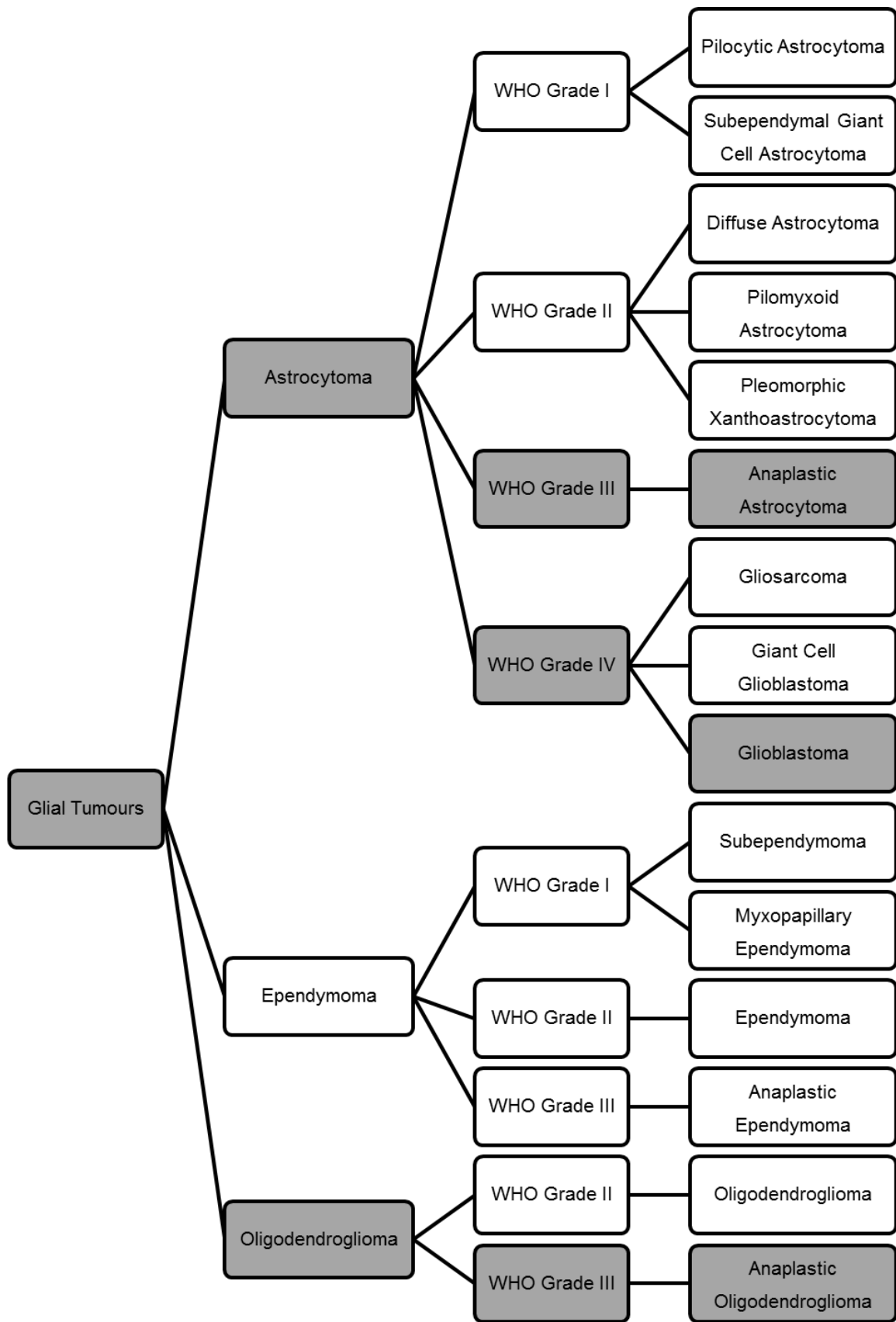


Figure 1.1. Overview of histological subtypes of glial brain tumours and WHO grading based on the 2007 WHO classification [8, 9]. Highlighted in grey are the tumours studied in this thesis.

Despite their distinct clinical histories, they are morphologically and clinically indistinguishable [12, 13]. Secondary GBMs occur less frequently (~5% of GBMs) and predominantly in younger patients (median age ~45 years vs ~60 years for primary GBMs). Despite the many differences between the adult and paediatric population, the WHO classification does not make a distinction for a specific diagnosis based on the patient's age.

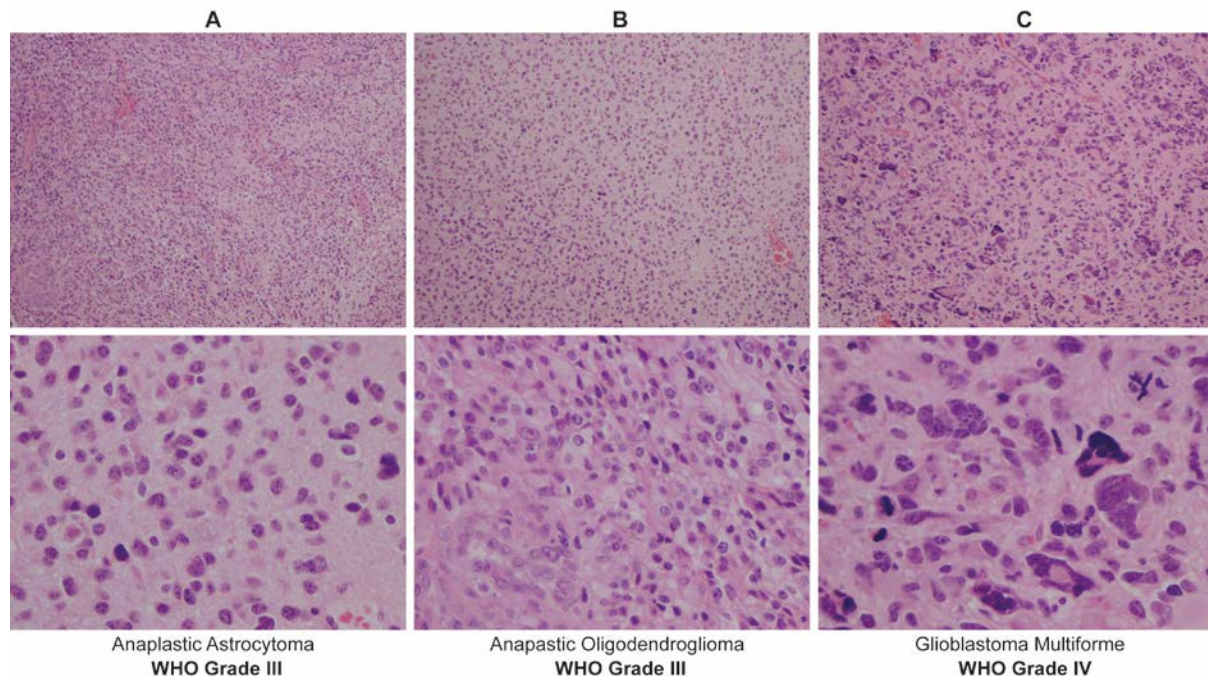


Figure 1.2. Schematic showing the classification of diffuse gliomas of astrocytic and oligodendroglial lineages. Representative micrographs for each tumour (anaplastic astrocytoma **(A)**), anaplastic oligodendroglioma **(B)** and glioblastoma **(C)** class studied in this thesis are given.

In the central nervous system (CNS) it is a common idea that a sequential differentiation takes place, starting with stem cells that give rise to progenitor cells that finally progress to fully differentiated cells. In the normal brain the CNS stem cell gives rise to neuronal and glial progenitors, which subsequently give rise to the mature cell types found in the brain, including neurons, oligodendrocytes and astrocytes [14]. Brain tumours may have its origin from either these CNS stem cells or their immediate downstream more proliferative progenitors (glial or neuronal), but the precise cell of origin remains largely unknown [15, 16]. It is generally believed that astrocytomas arise from astrocytic precursors, whereas oligodendrogliomas arise from oligodendrocytic, and that mixed gliomas arise from both progenitors (Figure 1.3). Interestingly, all of these cell types can give rise to glioma if genetically modified and even mature neurons and astrocytes can also generate malignant gliomas [17].

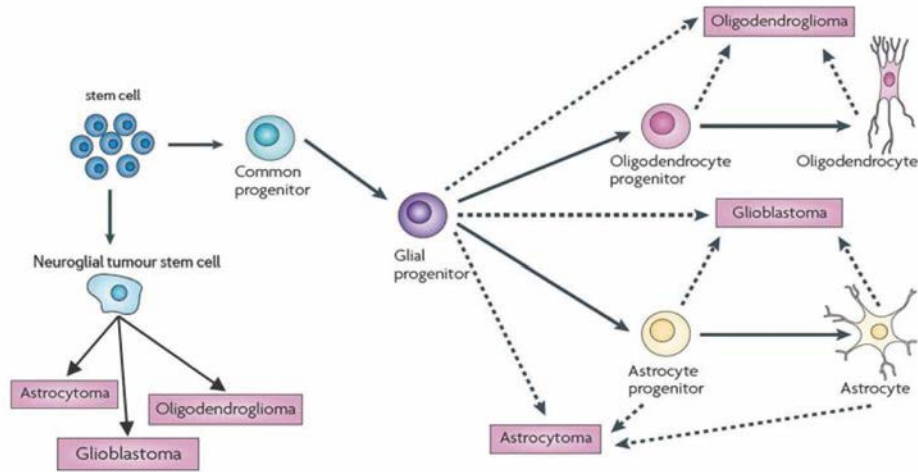


Figure 1.3. The neuroglial lineage tree. Committed glial progenitors, previously produced by self-renewing, common progenitors eventually differentiate into mature astrocytes and oligodendrocytes. Dashed arrows indicate a selection of likely candidates as cell of origin for each glioma variant [15].

1.1.2. Malignant glioma epidemiology

The annual age-adjusted worldwide incidence of primary brain and CNS tumours is 3.4 per 100,000 people with a mortality rate of more than two-thirds (2.5 per 100,000 people per year) [3, 4]. The incidence and mortality rates are different between men and women, with both rates being higher in men (3.9 vs 3.0 per 100,000 people and 3.0 vs 2.1 per 100,000 people, respectively) [3, 4]. In the paediatric population, there is a slight incidence difference between male and female, with a higher rate in male (1.2 vs 1.0 per 100 000 people per year). The mortality rate for both males and females is approximately 0.5-0.6 per 100 000 per year [3, 4].

In the paediatric population the world annual incidence of brain and CNS neoplasia is lower than in adults with a rate of 1.1 per 100,000 people per year [3, 4]. Nevertheless brain and CNS tumours are the second most frequent malignancy in children, just after leukaemia. These tumours account for 24.5% of all tumours in those 0–14 years of age and 8.9% in those 15–24 years of age [18]. More importantly, CNS tumours are the most common cause of cancer-related deaths in both the 0-14 year and the 15-24 year age group.

The majority of CNS tumours in all age groups are gliomas. There is an increased incidence of malignant glioma with increasing age (Table 1.1), with low grade glioma (LGG) being more common in children, whereas HGG is more common in adults [19]. Tumour location is one of the key differences between adult HGG (aHGG) and paediatric HGG (pHGG). While

Table 1.1. Rate (95%CI) of malignant glioma per 100.000 person/years.

Age range (years)	AA	AO	GBM	Glioma Malignant, NOS
0-4	0.05 (0.04-0.07)	NR	0.09 (0.07-0.11)	0.86 (0.81-0.92)
5-9	0.08 (0.07-0.10)	NR	0.14 (0.12-0.17)	0.8 (0.75-0.87)
10-14	0.09 (0.07-0.11)	NR	0.16 (0.14-0.19)	0.43 (0.39-0.47)
15-19	0.09 (0.08-0.11)	0.02 (0.02-0.04)	0.17 (0.14-0.19)	0.24 (0.22-0.28)
0-14	0.08 (0.07-0.09)	0.01 (0.01-0.01)	0.13 (0.12-0.15)	0.70 (0.67-0.73)
0-19	0.08 (0.07-0.09)	0.01 (0.01-0.01)	0.14 (0.13-0.15)	0.58 (0.56-0.61)
20-34	0.26 (0.24-0.28)	0.09 (0.08-0.10)	0.39 (0.37-0.41)	0.23 (0.22-0.25)
35-44	0.35 (0.33-0.38)	0.17 (0.15-0.19)	1.21 (1.16-1.25)	0.24 (0.22-0.26)
45-54	0.43 (0.41-0.46)	0.18 (0.17-0.20)	3.66 (3.58-3.74)	0.28 (0.25-0.30)
55-64	0.67 (0.63-0.71)	0.22 (0.20-0.25)	8.16 (8.02-8.30)	0.39 (0.36-0.42)
65-74	0.92 (0.86-0.98)	0.19 (0.17-0.22)	13.21 (12.98-13.45)	0.68 (0.63-0.73)
75-84	0.96 (0.88-1.04)	0.14 (0.11-0.17)	14.64 (14.34-14.94)	1.22 (1.14-1.31)
85+	0.40 (0.33-0.48)	NR	8.96 (8.59-9.33)	1.55 (1.40-1.71)

Data from Therese A. Dolecek et al, Neuro-Oncology, 2012 CBRTU, age-adjusted to the 2000 US standard population. Abbreviations: NOS, not otherwise specified; NR, not reported; AA: Anaplastic Astrocytoma; AO: Anaplastic Oligodendroglioma; GBM: Glioblastoma

in adults tumours occur mainly in the supratentorial regions of the brain (cerebral hemispheres and midline structures above the tentorium), in children the infratentorial region (brainstem and cerebellum) is more commonly affected (Figure 1.4) [20, 21]. Infratentorial malignant gliomas of the brainstem represent 10% of pHGG and 80% of these are characterized as diffuse intrinsic pontine gliomas (DIPG) [22]. Supratentorial HGG occur more frequently in young children than in adolescents or adults. The malignant gliomas that arise in the thalamus represent 12% of pHGG and are more common in children than adults [23, 24]. Less commonly, malignant glioma arises in the spine with an incidence of 3% in both adults and children [25].

The majority of HGG symptoms are common to the paediatric and adult populations. The patients normally present focal neurological deficits, cognitive decline, seizures, and symptoms of increased intracranial pressure, such as headaches, drowsiness, confusion, nausea and vomiting [26]. These symptoms can occur at any stage of the disease and they

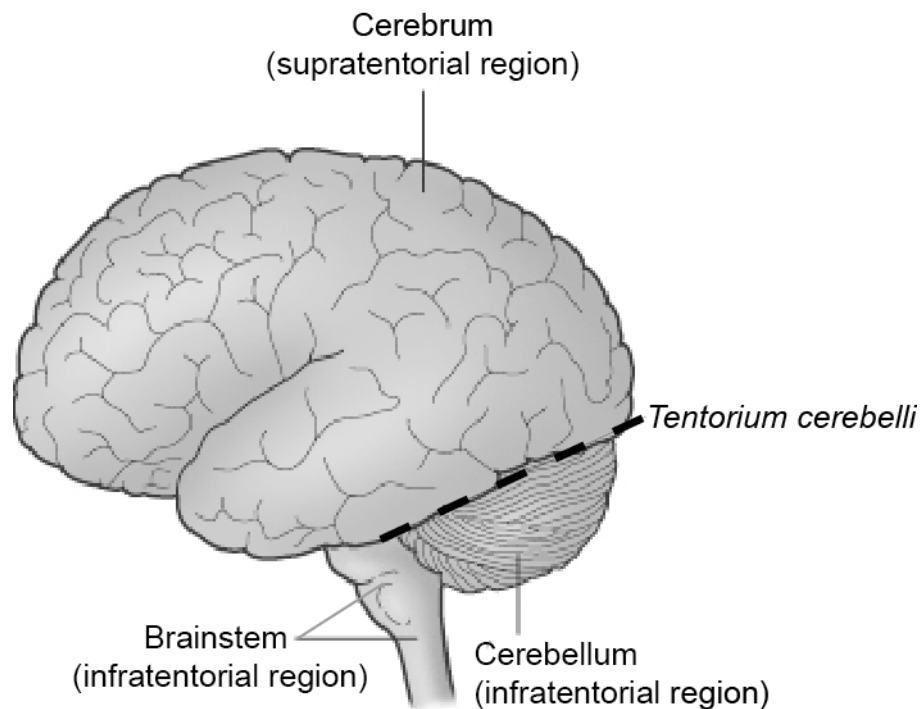


Figure 1.4. Representation of different anatomical regions of the brain. The cerebrum corresponds to the supratentorial region (above the *tentorium cerebelli*) of the brain and comprises the cerebral cortex (the left and right cerebral hemispheres), as well as several subcortical structures, including the hippocampus, basal ganglia, and olfactory bulb. The brainstem is part of the infratentorial region (below the *tentorium cerebelli*) of the brain and it includes the medulla oblongata, pons, and midbrain.

depend on the location and growth rate of the tumour. A rapidly growing and/or more aggressive tumour normally present symptoms that arise suddenly. On the other hand, a more prolonged clinical history possibly represents a slow growing, low grade tumour. Moreover, infants (>3 years) may also present with head enlargement, due to an incomplete fused skull [26, 27].

pHGG has always been a biologically understudied cancer when compared with aHGG, despite being one of the leading causes of cancer-related deaths in children, representing 15-20% of all childhood tumours of the CNS. Like in adults, even with aggressive multimodal therapy, children with GBM have a poor clinical outcome, with a current two-year survival rate of 10-15% [28].

1.1.3. Current therapy for paediatric high grade glioma

The initial treatment for a child with a newly diagnosed HGG is to attempt a maximal safe surgical resection [29, 30]. The ultimate goal of surgery is to obtain a complete resection without postoperative complications, however, even when a complete clear radiographic

scan is achieved, it is understood that microscopic tumour cells are still present due to the infiltrative nature of these lesions [30]. Sometimes a complete resection is not feasible due to tumour location (eg. brainstem HGG), in which case a biopsy might be performed. Additional treatments are normally required, including chemotherapy, radiation therapy or a combination of both in order to minimise the high likelihood of local recurrence [29, 30].

Radiotherapy was the first adjuvant therapy that improved survival in aHGG, and it also became the mainstay of therapy for children, especially for children older than 3 years old with newly diagnosed HGG [30]. Normally, the conventional dosing for a child is 50-60 Gy delivered in daily fractions of approximately 180-200 cGy over a 6 week period [31]. Based on strong research evidence [32], it has been shown that a combination of chemotherapy and radiotherapy brings benefits to children with newly diagnosed HGG. Since the CCG-943 trial, the addition of chemotherapy to radiotherapy has been adopted by many as the accepted standard of care for these patients. In this trial, patients were randomized to receive either radiotherapy alone *versus* radiotherapy followed by prednisone, lomustine and vincristine (PCV) chemotherapy. Five-year event free survival was 46% in the chemotherapy treated group *versus* 18% in the radiation alone group, showing a significant survival advantage [33]. The immediate successor to CCG-943 was the CCC-945 trial, in which an additional chemotherapy arm was added to PCV (PCV in combination with eight agents: prednisone, lomustine, vincristine, hydroxyurea, cisplatin, cytarabine, dacarbazine, and procarbazine). There was no statistical difference between the two arms and the outcomes were worse as compared to the previous CCG-943 study [34], although a subsequent central pathology review revealed that many patients enrolled in the CCG-943 and CCG-945 included LGG [35].

In 1995, Stupp and colleagues demonstrated that the addition of temozolomide (TMZ), an oral alkylating agent, to GBM treatment in adult patients gave meaningful significant survival benefit with minimal additional toxicity as compared to radiation alone [36]. The TMZ clinical trial in children with HGG by contrast showed no survival benefits when compared with control [37]. The Stupp protocol is nevertheless, due to TMZ tolerability and ease of administration, still used by many clinicians to treat newly diagnosed patients who are not enrolled on a clinical trial [30].

The increased understanding of signalling pathways involved in the initiation of malignant gliomas led to the development of novel targeted therapies, such as compounds directed against a variety of receptor tyrosine kinases (RTKs) (eg. erlotinib, gefinitinib, cetuximab) and anti-angiogenic drugs (bevacizumab and cediranib).

Table 1.2. A summary of targeted therapies in adult and paediatric malignant glioma, including pontine glioma. Adapted from [38].

Agent	Tumour	OR (%)	Median PFS (months)	6-month PFS (%)	Comments
EGFR					
Erlotinib	aHGG	0–25	1.8–5.2	0–19	Recurrent or relapsed
	pHGG	0	1.5	34	Recurrent or relapsed
	DIPG	17	8	90	Newly diagnosed with radiotherapy
Gefitinib	aHGG	0–18	2.0–6.8	9–62	Recurrent or relapsed
	pHGG	0	NR	15*	Recurrent or relapsed
	DIPG	14	7.4	88	Newly diagnosed with radiotherapy
Nimotuzumab	aHGG	NS	NS	NS	NS
	pHGG	4	1.8	NR	Recurrent or relapsed
	DIPG	10	5.5	NR	Newly diagnosed with radiotherapy
PDGFR					
Imatinib	aHGG	0–18	1.8–9.5	3–52	Recurrent or relapsed
	pHGG	0	NR	18	Recurrent or relapsed
	DIPG	3	NR	70	Newly diagnosed with radiotherapy
VEGFR/EGFR					
Vandetanib	aHGG	16	NR	NR	Recurrent or relapsed
	pHGG	NS	NS	NS	NS
	DIPG	NR	NR	88	Newly diagnosed with radiotherapy
VEGF					
Bevacizumab	aHGG	20–26	4–4.2	20–36	Recurrent or relapsed
	pHGG	0	4.5	42	Recurrent or relapsed with irinotecan
	DIPG	0	2.5	10	Recurrent or relapsed with irinotecan
mTOR					
Temsirolimus	aHGG	0–7	2.3	2–8	Recurrent or relapsed
	pHGG	0	1.9	NR	Recurrent or relapsed
	DIPG	0	2.5	NR	Recurrent or relapsed
Farnesyltransferase					
Tipifarnib	aHGG	6	2	11	Recurrent or relapsed
	pHGG	NS	NS	NS	NS
	DIPG	NR	NR	44	Newly diagnosed with radiotherapy

*1-year PFS. Abbreviations: aHGG, adult high-grade glioma; DIPG, diffuse intrinsic pontine glioma; pHGG, paediatric high-grade glioma; NR, not reported; NS, not studied; OR, objective response; PFS, progression-free survival.

Unfortunately to date, these have not been shown to provide significant survival benefit when used alone or in combination with other therapies (Table 1.2) [39]. Despite combined surgery, chemotherapy, and radiation treatments, a significant proportion of children will still have progressive or recurrent disease.

1.2. Molecular biology of high grade glioma

The biology of HGG in children and young adolescents was for a long time woefully understudied, in part due to limited number of paediatric samples and the belief that this disease was likely identical to aHGG. In the last few years large scale genomic profiling studies have helped to elucidate a clearer biological picture of pHGG and provided a deeper insight into gliomagenesis across all age groups. With the help of exome and whole genome sequencing (WGS) the genetic alterations underpinning pHGG are starting to be revealed.

1.2.1. Molecular profile of pHGG

Gene expression

In terms of gene expression signatures, in 2006 Philips and colleagues proposed for the first time the clustering of adult GBMs into three groups, designated Proneural, Mesenchymal and Proliferative. Both Mesenchymal and Proliferative types showed the most pronounced expression of stem-cell-related genes, similar chromosomal losses on chromosome 10 and gains on chromosome 7, with worse survival compared with the Proneural subtype.

Later, Verhaak and colleagues studying The Cancer Genome Atlas (TCGA) dataset proposed the existence of four subtypes instead of three [40, 41], described as Proneural, Neural, Classical and Mesenchymal. In the Classical subtype the vast majority of GBMs present chromosome 7 amplification paired with chromosome 10 loss, along with high rates of *EGFR* (epidermal growth factor receptor) alterations (amplification and mutation) and Rb (retinoblastoma protein) pathway disruption, mediated in part by *CDKN2A* (cyclin-dependent kinase inhibitor 2A) homozygous deletion and other alterations on *RB1*, *CDK4* (cyclin-dependent kinase 4) and *CCND2* (cyclin D2) genes, but no *TP53* (tumour protein p53) mutations were found [41, 42]. The Neural subclass included tumours that contain mutations and deoxyribonucleic acid (DNA) copy number aberrations (CNA) but surprisingly their expression patterns are the most similar to normal brain tumours. These tumours express neuronal markers, such as NEFL (neurofilament light polypeptide), GABRA1 (gamma-aminobutyric acid A receptor, alpha 1) and SLC12A5 (solute carrier family 12 member 5) [41]. In the Mesenchymal subtype focal *NF1* (neurofibromin 1) hemizygous deletions were the most common event. CNAs were described for *EGFR* and *MET* (mesenchymal epithelial

transition factor receptor), together with a high expression of CH13L1 (chitinase-3-like-1) and MET. The Proneural subset is associated with a younger age and expression of genes typical of an oligodendrocytic lineage, including *PDGFRA* (platelet-derived growth factor receptor), *NKX2-2* (NK2 homeobox 2) and *OLIG2* (oligodendrocyte lineage transcription factor 2), and also presented point mutations in *IDH1* (isocitrate dehydrogenase 1) gene, which have been previously associated with secondary GBM [43-45]. The Proneural group was further subdivided into glioma-CpG island methylator phenotype (G-CIMP)-positive and G-CIMP negative [46], with a subset of samples showing robust hypermethylation at a large number of loci, which correlated strongly with the presence of *IDH1* mutations. When comparing the Philips *versus* the Nounshmer classification schemes, it appears that the Proneural and the Mesenchymal signatures are the most robustly defined. Putative drivers of mesenchymal transformation have been recently identified, including *CTNND2* (encoding catenin- δ) as a negative driver and *RHPN2* (Rho GTPase binding protein 2) as a positive driver [47, 48]. Moreover, three transcription factors, STAT3 (signal transducer and activator of transcription 3), C/EBP β (CCTA/enhancer-binding protein- β) and TAZ (transcriptional co-activator with PDZ-binding motif), have recently emerged as epigenetic master regulators of the mesenchymal subgroup and can be used to predict poor clinical outcome [49, 50]. In addition, a study by Brennan and colleagues [51] measured the relative protein levels of signalling molecules within pathways crucial to glioma biology [42], a targeted proteomic analysis which showed some overlap with the genomic GBM subclasses previously described. Three distinct patterns were each associated with EGFR activation, PDGF activation or NF1 silencing, which correspond to the Classical-Neural, Proneural and Mesenchymal groups. The EGFR group has elevated expression of Notch ligands, leading to Notch pathway activation. The high levels of PDGFB ligand and PDGFRB phosphorylation characterize the PDGF subclass. The NF1 group is associated with low levels of NF1 and suppressed levels of PI3K (phosphatidylinositol-4, 5-bisphosphate 3-kinase) and MAPK (mitogen-activated protein kinase) pathways proteins [51].

Physiologic miRNA activity has been shown to also play an important role in gliomagenesis, through the regulation of genes involved in proliferation, apoptosis, migration, angiogenesis and aspects of stem cell biology. A recent study by Sumazin and colleagues showed that drivers of tumour initiation and different GBM subtypes, like *PTEN* (phosphatase and tensin homolog), *PDGFRA*, *RB1* and *VEGFA* (vascular endothelial growth factor A), are regulated by a distinct set of miRNAs [52].

The first gene expression profiling studies of pHGG underlined both the similarities and differences of these tumours when compared with aHGG. A study by Paugh *et al* used gene

expression profiles to distinguish different subgroups of pHGG, identifying three main groups – HC1, HC2 and HC3 [53]. The HC1 subgroup showed overexpression of genes associated with cell cycle regulation; overexpression of neuronal differentiation genes was seen in HC2 group and finally the HC3 group overexpressed genes related to cellular matrix-receptor interactions and cell adhesion. Using a gene set enrichment analysis, they showed that these three paediatric subgroups seem to overlap with the subgroups previously identified in aHGG - Proliferative, Proneural and Mesenchymal [54]. Significantly, 88% of PDGFR (platelet-derived growth factor receptor) driven tumours clustered into the Proliferative/HC1 subclass, showing that this pathway is a strong driver of proliferation in pHGG, and distinct to the strong PDGFRA/IDH1/Proneural associations observed in aHGG. Gain of 1q was found at high frequency both in Proliferative/HC1 and Proneural/HC2 subgroup, but highly underrepresented in the Mesenchymal/HC3 subclass. In addition, the Verhaak *et al* [41] Proneural subgroup, as well as being associated with PDGFRA, is closely correlated with *IDH1* mutations, a genotype largely absent from the Paugh *et al* data . Furthermore, the PDGFRA-driven gene expression signature in pHGG is different from that seen in aHGG, which is associated with cell-cycle regulators and genes associated with proliferation . Hence, pHGG are preferentially and differentially driven by *PDGFRA* amplification, in contrast to *EGFR* in adult tumours.

Copy Number aberrations

Some of the most common chromosomal aberrations found in aHGG are also described in pHGG but at significantly lower frequency. Chromosome 7 gain (74-83%) together with 10q loss (80-86%) is by far the most commonly described abnormality in adult GBM [55, 56]. In the paediatric setting these are far less common (13-19%, 16-38%, respectively) (Table 1.3) [55, 56]. Tumours with 7+/10- are highly enriched in the Classic adult subclass but are less common in the Proneural subtype, and completely absent in the IDH-mutant G-CIMP positive tumours [41, 46, 54, 57]. In addition, aHGG have also consistently, though at lower frequency, revealed gains of chromosomes 19 and 20, and losses of 9p, 22q, 13q, 14q and 6q [55, 56] (Table 1.3). Paediatric tumours normally display gain of 1q (19-20%) and, to a lesser extent, losses of 16q (7-18%) and 4q (2-15%) [38]. On average, childhood HGG harbour considerably fewer DNA copy number alterations than their adult counterparts, with ~15% of these tumours lacking any detectable CNAs [38, 57-59]

Table 1.3. Location-specific and age specific genetic differences in malignant glioma. *Adapted from [38, 60].*

Genetic Abnormality	DIPG*	HGG#				
		Infant	Child	Adolescent	Young adults	Adults
		<3ys	3-14ys	14-21ys	21-44ys	>45ys
Gain of 1 q	++	++	++	++	+	-
Loss of 16q	+	++	++	++	-	-
Stable Genome	-	++	++	++	-	-
Gain of 7	+	-	-	-	++	+++
Loss of 10q	++	+	+	+	++	+++
<i>EGFR</i> Amp	+	-	+	+	++	+++
<i>PDGFRA</i> Amp	+++	-	++	++	++	+
<i>CDKN2A/2B</i> Del	-	+	++	++	+++	+++
p53 pathway Alt	+++	+++	++	++	++	++
PI3K pathway Alt	++	+	++	++	++	+++
Rb pathway Alt	++	+	+	+	++	+++
<i>BRAF</i> V600E	-	-	+	++	+	-
<i>IDH1</i> R132X	-	-	-	+	+++	+
<i>H3F3A</i> K27M	+++	NR	+++	++	+	-
<i>H3F3A</i> G34R/V	-	NR	+	++	+	-
<i>HIST1H3B</i> K27M	++	NR	-	-	-	-
<i>ACVR1</i> mutation	++	-	-	-	-	-
<i>NTRK</i> SV	+	++	+	NR	NR	NR

*Peak age 4-9years. Grade not specified; infratentorially located. #Supratentorially located. Abbreviations: - low; + moderate; ++ high; +++ very high; DIPG, diffuse intrinsic pontine glioma; HGG, high grade glioma; NR, not reported; ys, years; Amp, amplification; Del, deletion; Alt, alterations

A study by Bax *et al* distinguished four different subclasses based on the pattern of different genomic profiles of paediatric GBM, described as Stable, Aneuploid, Rearranged or Amplifier genomes [59]. The Stable genome group has few, low level, focal changes, it occurs in 20.6% of cases and it is independent of histological grade or type, but seems to convey an improved survival in patients with HGG. This group appears to be unique to the paediatric setting. In contrast, the Aneuploid group comprises tumours with large, single copy alterations involving whole chromosomes or chromosomal arms, occurring in 34.9% of cases. The Rearranged group includes tumours with numerous, low level, intra-chromosomal breaks resulting in multiple gains and losses. Finally, tumours with single or multiple high level amplification were part of the Amplifier group, that had a significantly worse clinical outcome.

In 2008, high-throughput molecular profiling studies of aHGG lead to the identification of three major intracellular signalling pathways which are commonly targeted by genetic alterations. Later in 2013, TCGA initiative expanded the scope and depth of molecular data on GBM, incorporating next generation datasets available [56]. These studies showed that GBM has deregulation of growth factor signalling via amplifications and mutations of the RTK genes and the activation of the PI3K pathway; inactivation of the p53 pathway, often via direct mutation; and dysregulation of Rb tumour suppressor pathways via activation/inactivation of enhancers/repressors (Figure 1.5) [42, 44, 61]. Combining copy number and sequencing data, 79%, 86% and 90% of 251 samples harboured somatic alterations in core components of the RB, TP53 and RTK pathways [42, 56]. Contrary to adult GBM, alterations of the three core signalling pathways of HGG is seen in significantly lower frequency in children GBM: 25% RTK/PI3K, 19% p53 and 22% RB [59], at least at the copy number level. Genomic amplification of the *EGFR* gene is the most frequent genetic alteration in aHGG (>40%) but it is rarely seen in children HGG (0-11%) [38, 44, 55]. On the other hand, focal amplification at 4q12 encompassing the *PDGFRA* gene is the one focal alteration which is more commonly observed in children with HGG [58, 59]. *PDGFRA* is the most common event in paediatric setting and it leads to PDGFR signalling activation which is augmented by rare findings of *PDGFA*, *PDGFB*, and *PDGFRB* gain or amplification [38, 58, 59, 62]. Several studies have also suggested that *PDGFRA* is the key oncogenic event in DIPG and that this alteration is more commonly observed in radiation-induced gliomas [58]. Additionally, aHGG also harbour *CDKN2A* and *CDKN2B* deletions in 53% of cases, yet these events are rarely seen in pHGG [56]. The G1 checkpoint regulators *CCND1*, *CCND2*, *CCND3*, *CDK4* and *CDK6* are predominantly amplified in DIPGs, whereas *CDKN2A* homozygous deletions are restricted to supratentorial pHGG [63]. Other genes that are less frequently targeted by homozygous deletion in adult GBM (1-10%) include *PTEN*, *RB1*, *CDKN2C*, *FAF1* (FAS-associated factor 1), *QKI* (QKI KH domain containing), *NF1*, *NPAS3* (neuronal PAS domain protein 3) and *TP53* [60]. Focal amplification of *PDGFRA*, *MDM2* (mouse double minute 2 homolog), *MDM4*, *MET* often co-occur in various combinations in aHGG (1.5-13%) [56]. Nevertheless, amplification of genes such as *MYC* (v-myc myelocytomatosis viral oncogene homolog), *MYCN* (v-myc myelocytomatosis viral oncogene homolog, neuroblastoma derived) and *KRAS* (vi-Ki-ras Kirsten rat sarcoma viral oncogene homolog) and deletion of *CDKN2C* occur at low frequencies in pHGG [38, 58, 59, 64].

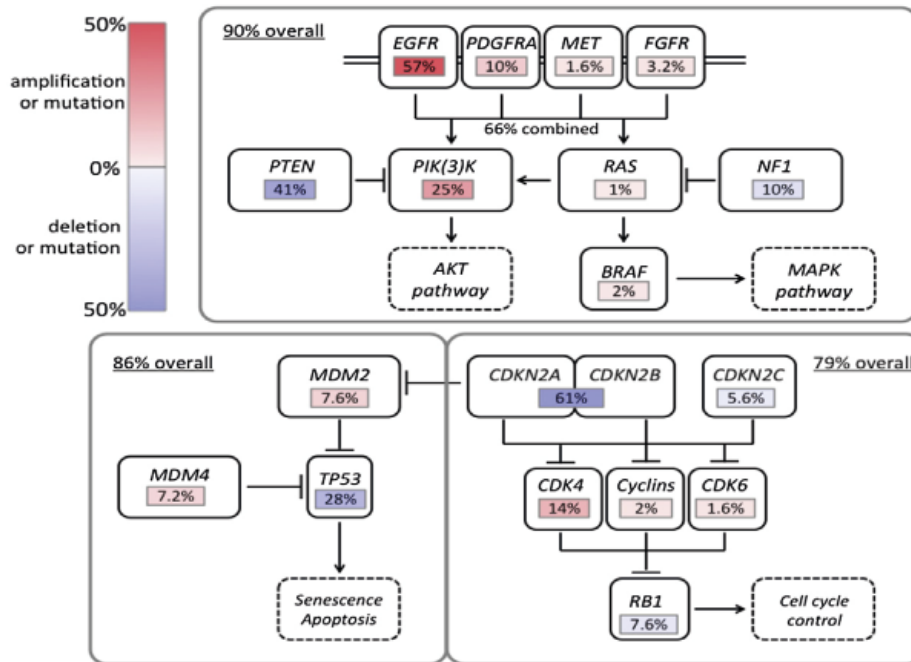


Figure 1.5. Landscape of core signalling pathway alterations in glioblastoma. *Adapted from [56]*

Gene Mutations

The most recent studies in both paediatric and aHGG aimed at full genome-wide characterization have led to the identification of novel recurrent mutations in both diseases. The availability of increasing number of samples for genome sequencing has now facilitated the detection of lower frequency events in well-known cancer genes and genes that were previously not associated with cancer. The number of coding mutations per tumour sample is highly variable in the adult population (3-179) and considerably lower in the paediatric setting (15 per tumour sample) [56, 60, 65].

pHGG present similar frequencies of *TP53* mutations as aHGG (34-37% vs 20-29%, respectively) [60]. However, HGGs arising in infants (<3 years) have a better prognosis and a lower frequency (9%) of *TP53* alterations [66], and in contrast, DIPGs present a higher frequency of these mutations (42-50%) [63, 67, 68]. Frequent somatic mutations are also found in *NF1* (25%), similar to that of aHGG, and both *PDGFRA* (8%) and *EGFR* (4%) are also mutated in pHGG [63]. The frequent mutations observed in *IDH1* and *IDH2* in secondary adult GBM, can only be found in 8.6% of pHGG cases, and are restricted to patients >13 years of age [38].

The single most important mutations found in pHGG which clearly distinguish these tumours from their adult counterparts are the recently discovered histone H3 mutations [65, 69].

These mutations occur on the histone tail, at or near important modification sites, are restricted to ~38% of childhood and young adults (<30 years) HGGs [60] and are mutually exclusive with recurrent point mutations in *IDH1* [69]. Heterozygous mutations in *H3F3A* (H3 histone, family 3A) result in amino acid substitutions at positions K27 and G34. Additionally, K27 is also found to be mutated in the H3.1 histone genes *HIST1H3B* (histone cluster 1, H3b) and *HIST1H3C* [65, 70]. The *H3F3A* p.27 lysine-to-methionine substitution occurs in midline GBM and DIPG cases in younger children (70-88% cases [65, 67]), though the same substitution in *HIST1H3B* is specific to DIPG and it occurs at lower frequency (11-31%). Moreover, recurrent activating mutations in the *ACVR1* (activin A receptor, type I) gene were exclusively found in DIPG, and were strongly associated with younger age, longer survival time and the presence of *HIST1H3B* p.Lys27Met as well as wild-type *TP53* [63, 67]. The K27M mutation confers a reduction of the global trimethylation at this site (downregulation of the repressive mark), which dramatically alters gene expression leading to overexpression of PRC2 (polycomb repressive complex 2) target genes [71]. *H3F3A* p.Glycine-to-Arginine/Valine at residue 34 occurs in adolescents which present supratentorial HGG. This mutation interferes with the regulation of H3K36me3 active mark, but contrary to H3K27M does not lead to a global downregulation of the trimethylation at this residue [72]. A study by Bjerke *et al* [72] showed that in a G34V mutant paediatric GBM cell line (KNS42) *MYCN* is the gene that is most strongly enriched for H3K36me3 marks, and is also associated with transcriptional upregulation of this locus. Furthermore, the H3K36 trimethyltransferase *SETD2* (SET domain containing 2) is also mutated in children and young adults, and these mutations are mutually exclusive with *H3F3A* G34 mutations [73], which may be an alternative mechanism to deregulate the active mark. Moreover, it has been shown by Li and colleagues that *SETD2* downregulation, followed by depletion of H3K36me3 leads to an increased spontaneous mutation frequency and chromosomal depletion [74]. In addition, the G34 tumours also harbour mutations in H3.3 chaperone complex *ATRX/DAXX* (alpha thalassemia/mental retardation syndrome X-linked / death-domain associated protein), which is responsible for depositing the histone H3 at telomeres, pericentric heterochromatin and actively transcribed regions. The mutations in *ATRX* and *DAXX* contribute to the ALT (alternative lengthening telomeres) phenotype commonly seen in pGBM, which maintains or increases telomere length [69] in the absence of TERT promoter mutations. In addition, these tumours normally present a pattern of hypomethylation [57, 69].

In addition to recurrent histone H3, *ATRX/DAXX* and *SETD2* mutations, frequent mutations in other histone writers and erasers and in chromatin-remodelling genes have also been recently described (eg. *MLL*, *KDM5C*, *KDM3A*, *JMJD1C*) [63]. These mutations are often

concurrent with the mutations in H3 genes, with 91% of DIPG and 48% of hemispheric HGG containing mutations in histone genes and/or this group of epigenetic regulators [63].

1.3. Somatic structural variation in cancer

Cancer is a genetic disease that occurs at the cellular level with two dominating types of initiating events: the inactivation of genes by deletion, mutation or epigenetic mechanisms, and the activation or deregulation of genes as a consequence of point mutation, amplification or cytogenetic abnormality [75]. These events occur gradually and cumulatively in a cell, enhancing a cell's evolutionary fitness, promoting outgrowth and progression towards cancer [76, 77]. One of the most difficult forms of genetic variation to interpret with respect to functional consequences are structural variants (SV). These may involve equal exchange of material between two chromosomal regions – balanced rearrangements – or lead to loss or gain of portions of the genome – unbalanced rearrangements (leading to copy number variation). Balanced rearrangements include inversions and reciprocal translocations; unbalanced rearrangements include deletions, duplications, insertions, amplifications and non-reciprocal translocations [78] (Figure 1.6). SVs are commonly classified as sequence variants of at least 50 base pairs (bp) in size, therefore discriminating them from smaller variants, such as single-nucleotide variants (SNV) and short insertions and deletions (indels) [79, 80]. Furthermore, it has been hypothesized that the primary pathogenic changes in cancer result from balanced rearrangements, while the secondary changes that occur during cancer progression are from unbalanced changes [81]. The early chromosomal alterations give the cell a proliferative advantage, and alter the normal cellular environment by disrupting proto-oncogenes and tumour suppressor genes, leading to additional changes in the genome. The importance of chromosomal aberrations in tumour development varies substantially between and within tumour types; some undergo marked genome rearrangement, whereas others may evolve by alternative mechanisms [54, 59, 82]. The number of aberrations is normally small in pre-malignant, hyperproliferative lesions and substantially greater in more advanced tumours. Among these rearrangements, balanced chromosomal translocations are considered as the primary cause for many cancers like lymphoma, leukaemia and some solid tumours [75, 83, 84].

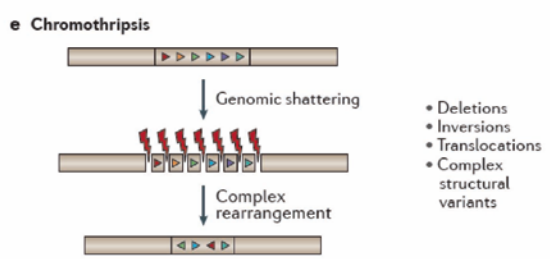
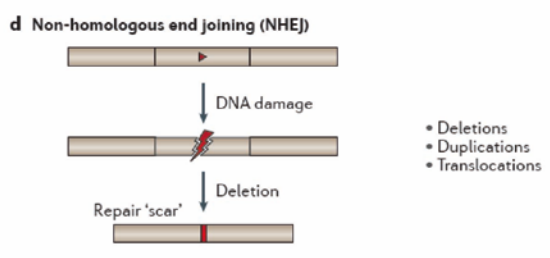
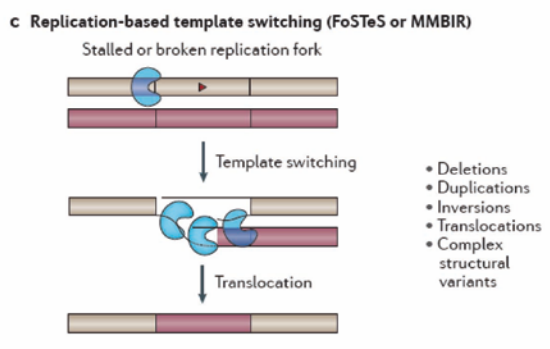
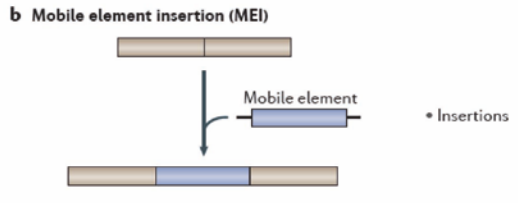
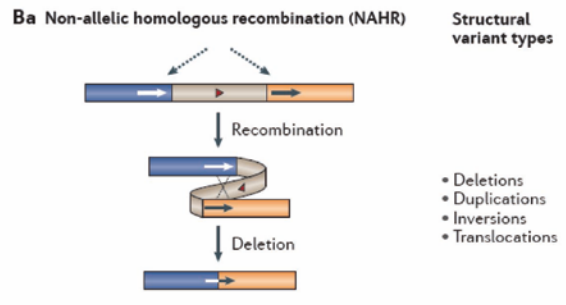
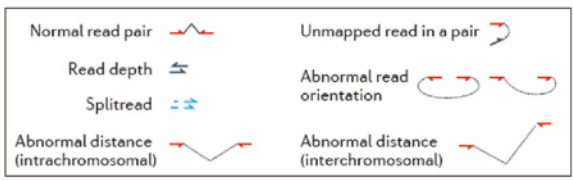
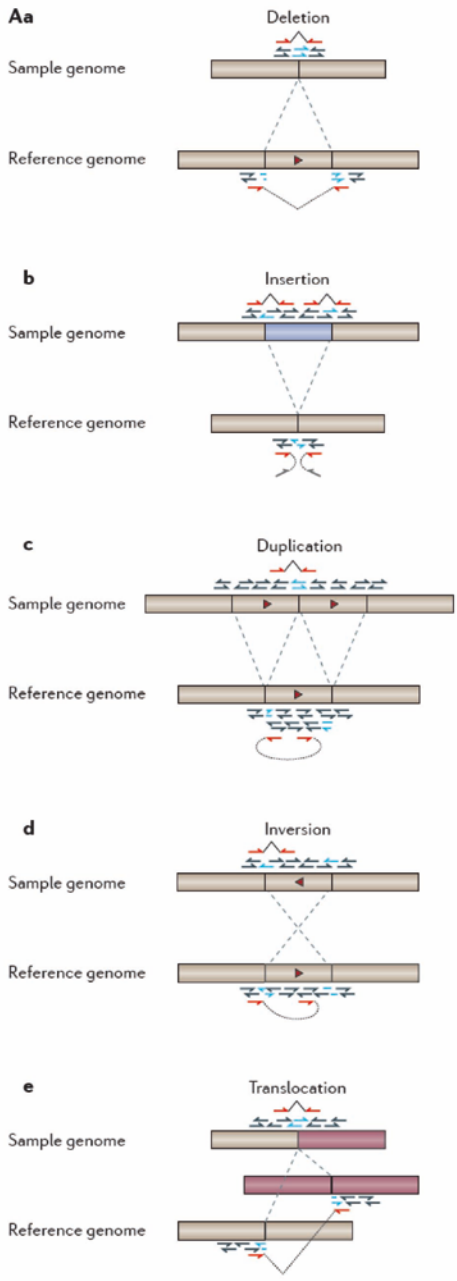


Figure 1.6. Structural variants: classes and formation mechanisms. A) Structural variants comprise unbalanced copy-number variation (a), insertions (b) and duplications (c), as well as balanced variants such as inversions (d) and translocations (e). B) Molecular mechanisms leading to structural variant formation a. Recurrent structural variants often result from non-allelic homologous recombination (NAHR) which involves recombination between long highly similar low-copy-number repeats (blue and orange segments). b. Novel genomic insertions can involve mobile element insertion of transposable elements by retrotransposition. c. DNA-replication-associated template-switching events, involving the fork-stalling and template switching (FoSTeS) and microhomology-mediated break-induced replication (MMBIR) mechanisms. d. Many structural variants in humans are attributed to non-homologous end joining (NHEJ) which is often accompanied by the addition or deletion of several nucleotides in the form of a 'repair-scar' (small red bar). e. Chromothripsis —a phenomenon that involves chromosome shattering leading to numerous breakpoints, followed by error-prone DNA repair. *Adapted from [79]*

1.3.1. Mechanisms of structural variation

For a rearrangement to occur, several criteria need to be fulfilled. First, the sequences undergoing rearrangement need to be in close spatial proximity. Secondly, there needs to be a DNA break or a switch in the template strand used during DNA replication. Finally, some sequences are more prone to rearrangements owing to their structure and epigenetic modifications [85]. Double-strand DNA breaks (DSB) are known to play the essential role in the occurrence of SVs, irrespective of the origin and cell type, and are thought to be the prerequisite for any chromosomal aberration to occur, though the exact mechanism is mostly unclear. Mutagens of both internal and external origins continuously damage normal DNA causing these DSBs. Direct exposure to harmful exogenous agents, like ionizing radiation or drugs may cause DSBs, but they can also occur as a product of a normal aerobic metabolism which creates oxygen free radicals that damage DNA. There are also several cellular processes that generate and repair DSBs as part of normal programmed genomic rearrangements: V(D)J recombination and class switch recombination, both related to antibody generation [83, 86].

Changes in the structure of chromosomes are repaired by two general mechanisms: homologous recombination and non-homologous recombination (Figure 1.6) [87]. A highly conserved non-homologous DNA repair pathway, Non-Homologous End Joining (NHEJ), repairs DSBs in the absence of extensive sequence homology and is often accompanied by the addition or deletion of several nucleotides in the form of a "repair scar" [79]. NHEJ events sometimes have some degree of microhomology (2-25 bp of similarity) at their breakpoints [88]. On the other hand, non-allelic homologous recombination (NAHR) involves

recombination of homologous sequences at regions of repetitive sequences on the genome, known as low-copy-repeats (LCRs) [87, 88]. Other mechanisms, such as fork stalling and template switching (FoSTeS) and microhomology-mediated break-induce replication (MMBIR) have also been proposed as mechanisms that lead to simple or complex SVs, frequently involving duplicative events [87, 88]. Moreover, Stephens and colleagues reported that there is another mechanism that leads to genetic instability, mutated, amplified and rearranged genes, a one-off cataclysmic event in which chromosomes are shattered and then stitched back together, a phenomenon termed chromothripsis [76]. Chromothripsis seems to be a rare event, only seen in 2-3% of all cancers and also in some congenital diseases [76, 89, 90] but it is significantly higher in adult GBM (>30%) and in pHGG (31%) [63, 91]. The DSBs caused by chromothripsis are repaired by NHEJ mechanism, and when these repair systems fail chromosomal aberrations occur [83, 92].

1.3.2. Biological impact of structural variants

The molecular consequences of SVs are relatively easy to predict when the variant involves entire genes – the expression of which is assumed to vary accordingly – or when the breakpoints of the rearrangements intersect with coding regions, hence leading to gene fusions or truncations. It is considerably more problematic when SVs occur kilobases (kb) upstream of genes and yet may still result in gene expression changes [87].

1.3.3. Technologies for measurement of structural variation

SVs have been thoroughly studied using a variety of methods. The first technique used to reliably discriminate between chromosomes and identify translocations, inversions and CNVs larger than 5Mb was the Giemsa staining of condensed chromosomes (G-banding) [93]. The discovery of the Philadelphia chromosome in chronic myeloid leukaemia (CML) in 1960 [94] was the first consistent chromosomal change to be seen in human cancer and it was discovered using G-banding [75]. Complementary to this low-resolution method, a targeted and more sensitive method was developed, FISH (fluorescent *in situ* hybridization). FISH allows the precise detection of abnormalities of a few hundred kb based on hybridization of fluorescently labelled probes to specific complementary sequences in the chromosome. When visualising FISH under the microscope, detection of deletions (missing signals), amplifications (extra signals) or translocations (relocated signals) is possible [93]. The approach of identifying novel candidate fusions by 1) detecting a consistent chromosomal aberration using molecular cytogenetic techniques, 2) examining the SV at higher levels of resolution using appropriate molecular techniques until the gene fusion in question is finally identified, and finally 3) the sequence being determined and the case-to-

case variability in terms of precise breakpoint sites being established has been used for several years [95, 96]. Though important, this approach is limited and these cytogenetic techniques do not allow the detection of small SVs. Nevertheless, the recent improvement in genome-wide techniques allowed us to identify the extent of CNVs and also to have a more complete measurement of SVs of all types [97].

Microarray-based techniques – aCGH (array comparative genomic hybridization) and SNP array (single nucleotide polymorphism array) – were the first genome-wide survey of SVs in the human genome. The discovery of the *BRAF* (v-raf murine sarcoma viral oncogene homolog B) tandem duplication at chromosome 7q34 present in 66% pilocytic astrocytomas was first discovered using these microarray based techniques [98, 99]. In addition, due to its affordability, microarrays permitted a comprehensive picture of all CNVs across several types of tumours. However, the information that arrays can reveal about SVs is limited. They can only detect sequences that match the oligonucleotide probes used to make them, which lie in non-repetitive regions, making it difficult to detect SVs with breakpoints in repetitive regions, such as NAHR events or indels of repetitive sequences [88, 97, 100]. Furthermore, microarrays can only detect differences in number of copies, being blind to copy-neutral, or balanced variants, such as inversions or reciprocal translocations [88].

DNA sequencing technology has advanced dramatically and has allowed researchers to simultaneously infer information of both structural and single nucleotide variants. One strategy for genome-wide high-resolution identification of fusion genes and other large scale SVs is paired-end sequencing of genomic DNA or RNA from tumour samples or tumour-derived cell lines [101]. First, the genome is precisely divided into fragments of known size (~500bp) and instead of sequencing the entire piece of DNA, reads are taken only from either end. The resulting paired reads are mapped back to the reference human genome sequence and if they do not match, then a genomic rearrangement is inferred [102]. Algorithms for detection of SVs from next generation sequence data normally rely on discordant paired-end reads or depth of coverage. Read-depth indicates how many copies are present of each gene and if a region is amplified it will show higher read depth, and if deleted, low read depth (less reads). Many algorithms, including Breakdancer and TopHat [103, 104] used to detect SVs rely on the presence of discordant paired reads. In the case of inter-chromosomal translocation, one member of the pair will map to one chromosome and its mate to another. When inversions or intra-chromosomal translocations occur, the two ends map to the same chromosome but in the wrong orientation or the wrong distance apart. These methods can normally detect the breakpoints but with low resolution and normally have low specificity, when one of the paired ends maps to a repetitive region or a high

homology region [105]. In addition, paired-end analysis methods rely on cut-offs to filter out false positives which are normally based on the number of supporting reads. In order to avoid the high-false positive rates inherent to most paired-end approaches, and to better localize the breakpoints, PCR-based validation followed by first generation sequencing (Sanger sequencing) is important.

Current paired-end sequencing can obtain an extremely high probability of breakpoint detection with a very low number of reads. Array CGH can be used by itself to identify the breakpoints involved in deletions and amplifications, using algorithms to analyse CNAs, but when combined with next-generation end sequencing data, breakpoint resolution can improve even more. As the cost of next-generation end sequencing declines and its technology is improved, paired-end sequencing of cancer genomes (both tumour samples and cell lines) will provide reliable and precise detection of fusion genes [101, 106].

1.3.4. General overview of fusion genes in cancer

Translocations are generally classified as reciprocal and non-reciprocal. Reciprocal translocations occur when segments between two chromosomes are exchanged. This can happen between any two chromosomes and at various sites along the length of the chromosome. Non-reciprocal translocations occur when a chromosomal segment is transferred to a non-homologous chromosome (one-way transfer). The juxtaposition of the coding region of a gene near the transcriptionally active promoter/enhancer region of another gene, leading to over-expression of the former gene, is a functionally important outcome of translocations. They can also result in the formation of a unique fusion or chimeric gene, which can be used as a marker for the malignant cells [83].

At present, more than 300 different fusion genes have been identified [75]. Fusion genes can be used as diagnostic markers and most importantly as specific therapeutic targets, leading to the development of successful cancer drugs [107]. Although gene fusions are typically associated with haematological malignancies, rare bone and soft-tissue tumours, they have been also described in solid tumours, namely in prostate, ovarian, brain and lung cancers [108-112]. The development of cutting edge technology for DNA sequencing that has been developed in the past few years has led to an enormous number of publications where novel and recurrent rearrangements are described in numerous different cancers.

Up to 2007 (the latest review documenting fusion genes in cancer) 264 gene fusions, involving 238 different genes, had been identified in haematological disorders and these represented 75% of all fusion genes known in human neoplasia at the time [75]. The first

consistent chromosome change to be seen in human cancer was observed in CML in 1960 [94]. The Philadelphia chromosome is the result of a reciprocal translocation, t(9;22)(q34.1;q11.23), in which a portion of the *BCR* (breakpoint cluster region) gene is joined to the cytoplasmic *ABL1* (c-abl oncogene 1, non-receptor tyrosine kinase) tyrosine kinase, generating a constitutively active enzyme. The resulting chimeric protein contains the catalytic domain of ABL fused to a domain of BCR that mediates constitutive oligomerization of the fusion protein in the absence of physiologic activating signals, thereby promoting aberrant tyrosine kinase activity [113]. This discovery led to the development of the first small molecule inhibitor imatinib, which inhibits the fusion kinase activity [107, 114-116]. Other well-known specific changes seen in practically 100% of a particular leukaemia or lymphoma type are the immunoglobulin heavy locus (*IGH*):*CCND1* that occurs in mantle cell lymphoma, the *MYC* deregulation in Burkitt lymphoma and *PML*:*RARA* (promyelocytic leukaemia / retinoic acid receptor, alpha) in acute promyelocytic leukaemia.

Only more recently have fusion genes in other solid tumours started to be described as novel techniques have made such discoveries possible. Some of the first and most important of these have been identified in prostate cancer. Tomlins and colleagues described a fusion between the 5' untranslated region of a prostate-specific and androgen-inducible gene, *TMPRSS2* (transmembrane protease, serine 2) and the 3' end region of different genes in the ETS (erythroblastosis virus E26 transformation-specific) family of transcription factors (eg. *ERG*, *ETV1*) [108]. The *TMPRSS2:ERG* fusion can occur either through inter-chromosomal insertion or through deletion of the intervening region on chromosome 21 (reviewed in [117]). This alteration has been observed in 50–60% of prostate tumours [118], and leads to increased cell migration [119] and proliferation [120], although the prognostic implications of *TMPRSS2:ERG* have not yet been fully elucidated. Currently, there is evidence that this fusion can be used as potential biomarker for prostate cancer, using non-invasive screening techniques [117].

Bone and soft tissue sarcomas, as in haematological malignancies, have gene fusions that occur more often in some subtypes than in others. There are 41 known gene fusions in 17 different sarcoma types [75]. Close to 100% of Ewing sarcomas, myxoid liposarcomas and synovial sarcomas harbour chimeric genes *EWSR1*, *DDIT3* and *SSX1* (2 and 4), respectively [75].

In 2007, a study to isolate novel transforming genes in non-small-cell lung cancer (NSCLC) reported the presence of a fusion gene that contains the N-terminal portion of the protein encoded by the echinoderm microtubule-associated protein-like 4 (*EML4*) gene and the intracellular signaling portion of the TK encoded by anaplastic lymphoma kinase (*ALK*) gene,

giving rise to a novel kinase that confers a proliferative advantage to cancer cells. Five out of 75 (~7%) of patients with NSCLC were positive for this fusion, and were distinct from those harbouring *EGFR* mutations, another hallmark of this type of cancer [112]. In just 4 years since the discovery of this novel kinase, a new drug – Crizotinib – has been approved by the US Food and Drug Administration (FDA) [121]. Treatment of 82 patients with advanced *ALK* rearrangement-positive NSCLC resulted in tumour shrinkage in 57% of patients and stabilized disease in a further 33% [122].

These results show that fusion genes are potential druggable alterations and they can help to distinguish different subtypes within the same disease.

1.3.5. Gene fusions in brain tumours

As recently as 2008, Jones and colleagues reported the first recurrent rearrangement in brain tumours [98, 99]. This involves a gene fusion involving the *BRAF* gene, a key component of the MAPK pathway, a commonly deregulated pathway in paediatric low grade astrocytomas. A *KIAA1549–BRAF* fusion was found in 80% of sporadic pilocytic astrocytomas [98, 99, 123], with the *KIAA1549* gene replacing the auto-inhibitory domain of the RAF molecule, leading to constitutive activation of the kinase activity [98]. Recently, other fusion partners for *BRAF* have been described, including interstitial deletion of a ~2.5 megabase (Mb) segment that leads to the fusion of *BRAF* gene with an uncharacterized gene located on 7q34, *FAM131B*. This deletion removes the *BRAF* N-terminal inhibitory domains, giving a constitutively active BRAF kinase [124]. *BRAF* and other family members (eg. *RAF1*) fusions have also been described in a small percentage of prostate and gastric cancer, and melanoma [125]. Furthermore, in 2013, Jinghui Zhang and colleagues sequenced 39 paediatric LGGs and reported the presence of intragenic duplications of the portion of the *FGFR1* (Fibroblast growth factor receptor 1) encoding the tyrosine kinase domain, as well as rearrangements of *MYB* (v-myb avian myeloblastosis viral oncogene homolog) which lead to activation of the same pathway as BRAF fusions [126].

In HGG, to date only few fusions have been described in the literature. In 2003, Charest and colleagues described the first fusion gene found in an adult GBM cell line, an intra-chromosomal homozygous deletion on chr6q21 that leads to the formation of *FIG* (Fused in GBM):*ROS* (c-ros oncogene 1, receptor tyrosine kinase) fusion, which displays a constitutive TK activity in the GBM cell line U118MG [127]. *ROS* is a proto-oncogene highly expressed in a variety of tumour cell lines, and belongs to a subfamily of T insulin receptor genes. The second fusion to be described in GBM involves genes *LEO1* (Leo1, Paf1/RNA polymerase II complex component) and *SLC12A1* (solute carrier family 12, member 1), but no further

studies have been done to explore the oncogenic characteristics of this fusion [128]. In addition, a study by Ozawa T. and colleagues described a fusion involving *PDGFRA* and the gene encoding for the vascular endothelial growth factor receptor – 2 (*VEGFR-2*) [129]. The resulting fusion protein has constitutive *PDGFRA* kinase activity and oncogenic transforming potential and occurs in HGG with amplified *PDGFRA*.

Moreover, in 2012, Singh and colleagues identified a tandem duplication on chromosome 4p16.3 which creates a fusion transcript, *FGFR3:TACC3* (transforming, acidic coiled-coil containing protein 3) [130]. This structural variant leads to loss of the 3'-UTR of *FGFR3*, blocking gene regulation of miR99a and enhancing expression of the fusion gene [131]. The chimaera occurs in 3% of adult GBM, displays oncogenic activity and is associated with amplification of both genes [130]. *EGFR* was also found to be fused with *SEPT14* (septin 14) in 7% of adult GBM, and almost always occurred within amplified regions of the fusion partner genes [47]. This fusion gene gives the ability to glioma cells to self-renew and constitutively activate *STAT3* signalling, also conferring sensitivity to EGFR inhibition [47].

The recent sequencing initiatives of pHGG tumours led to the discovery of the first fusion gene in this type of tumour [63]. Gene fusion involving the kinase domain of each of the three *NTRK* (neurotrophin receptor) genes and five different N-terminal fusion partners were identified in 4% of DIPG and 10% of non-brainstem HGGs. Interestingly, 40% of non-brainstem HGGs occur in infants (<3 years old). *NTRK* fusions drive glioma formation *in vivo* and lead to activation of the PI3K/MAPK signalling [63].

All fusion genes occurring in brain tumours until now seem to use the same strategy, the replacement of the regulatory domain of a kinase gene by a constitutively active regulatory domain of another gene. This mechanism leads to a constitutively active kinase, which gives growth advantage to the cells that harbour the fusion. The low frequency of gene fusions in malignant brain tumours can lead to the idea that they play a minor role in the disease pathogenesis, nevertheless fusion genes that have not yet been identified, might characterize the different subtypes of this broad disease and the eventual goal is that novel drugs are developed acting specifically against the alteration.

1.4. References

1. Organization, W.H., *The global burden of disease: 2004 update*. 2008.
2. Jemal, A., et al., *Global cancer statistics*. CA: a cancer journal for clinicians, 2011. **61**(2): p. 69-90.
3. Ferlay, J., et al., *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]*. 2012, Lyon, France: International Agency for Research on Cancer.
4. Bray, F., et al., *Global estimates of cancer prevalence for 27 sites in the adult population in 2008*. Int J Cancer, 2013. **132**(5): p. 1133-45.
5. Institute, N.C., *A Snapshot of Pediatric Cancers - Incidence and Mortality Rate Trends*. U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES, 2011.
6. Group, U.S.C.S.W., *United States Cancer Statistics: 1999–2010 Incidence and Mortality Web-based Report*. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. 2013.
7. UK, C.R. <http://info.cancerresearchuk.org/cancerstats/childhoodcancer/> 12/12/2011.
8. Louis, D.N., et al., *The 2007 WHO classification of tumours of the central nervous system*. Acta Neuropathol, 2007. **114**(2): p. 97-109.
9. Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., *WHO Classification of Tumours of the Central Nervous System, Fourth Edition*. 2007.
10. Kleihues, P., et al., *The WHO classification of tumors of the nervous system*. J Neuropathol Exp Neurol, 2002. **61**(3): p. 215-25; discussion 226-9.
11. Zhu, Y. and L.F. Parada, *The molecular and genetic basis of neurological tumours*. Nat Rev Cancer, 2002. **2**(8): p. 616-26.
12. Ohgaki, H. and P. Kleihues, *Genetic pathways to primary and secondary glioblastoma*. Am J Pathol, 2007. **170**(5): p. 1445-53.
13. Ohgaki, H., et al., *Genetic pathways to glioblastoma: a population-based study*. Cancer Res, 2004. **64**(19): p. 6892-9.
14. Holland, E.C., *Gliomagenesis: genetic alterations and mouse models*. Nat Rev Genet, 2001. **2**(2): p. 120-9.
15. Huse, J.T. and E.C. Holland, *Targeting brain cancer: advances in the molecular pathology of malignant glioma and medulloblastoma*. Nat Rev Cancer, 2010. **10**(5): p. 319-31.
16. Dirks, P.B., *Brain tumor stem cells: the cancer stem cell hypothesis writ large*. Molecular oncology, 2010. **4**(5): p. 420-30.
17. Friedmann-Morvinski, D., et al., *Dedifferentiation of neurons and astrocytes by oncogenes can induce gliomas in mice*. Science, 2012. **338**(6110): p. 1080-4.
18. Arora, R.S., et al., *Age-incidence patterns of primary CNS tumors in children, adolescents, and adults in England*. Neuro Oncol, 2009. **11**(4): p. 403-13.
19. Dolecek, T.A., et al., *CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005-2009*. Neuro Oncol, 2012. **14** Suppl 5: p. v1-49.

20. Qaddoumi, I., I. Sultan, and A. Gajjar, *Outcome and prognostic features in pediatric gliomas: a review of 6212 cases from the Surveillance, Epidemiology, and End Results database*. *Cancer*, 2009. **115**(24): p. 5761-70.
21. Rineer, J., et al., *Characterization and outcomes of infratentorial malignant glioma: a population-based study using the Surveillance Epidemiology and End-Results database*. *Radiother Oncol*, 2010. **95**(3): p. 321-6.
22. Hargrave, D., U. Bartels, and E. Bouffet, *Diffuse brainstem glioma in children: critical review of clinical trials*. *Lancet Oncol*, 2006. **7**(3): p. 241-8.
23. Kramm, C.M., et al., *Thalamic high-grade gliomas in children: a distinct clinical subset?* *Neuro Oncol*, 2011. **13**(6): p. 680-9.
24. Wolff, J.E., et al., *Subpopulations of malignant gliomas in pediatric patients: analysis of the HIT-GBM database*. *J Neurooncol*, 2008. **87**(2): p. 155-64.
25. Wolff, B., et al., *Pediatric high grade glioma of the spinal cord: results of the HIT-GBM database*. *J Neurooncol*, 2012. **107**(1): p. 139-46.
26. Sizoo, E.M., et al., *Symptoms and problems in the end-of-life phase of high-grade glioma patients*. *Neuro Oncol*, 2010. **12**(11): p. 1162-6.
27. *Pediatric Oncology - A Comprehensive Guide*, ed. P. Imbach, T. Kühne, and R.J. Arceci. 2011: Springer Berlin Heidelberg.
28. Broniscer, A. and A. Gajjar, *Supratentorial high-grade astrocytoma and diffuse brainstem glioma: two challenges for the pediatric oncologist*. *Oncologist*, 2004. **9**(2): p. 197-206.
29. Fangusaro, J., *Pediatric high-grade gliomas and diffuse intrinsic pontine gliomas*. *J Child Neurol*, 2009. **24**(11): p. 1409-17.
30. Fangusaro, J., *Pediatric high grade glioma: a review and update on tumor clinical characteristics and biology*. *Front Oncol*, 2012. **2**: p. 105.
31. Fallai, C. and P. Olmi, *Hyperfractionated and accelerated radiation therapy in central nervous system tumors (malignant gliomas, pediatric tumors, and brain metastases)*. *Radiother Oncol*, 1997. **43**(3): p. 235-46.
32. Fine, H.A., et al., *Meta-analysis of radiation therapy with and without adjuvant chemotherapy for malignant gliomas in adults*. *Cancer*, 1993. **71**(8): p. 2585-97.
33. Sposto, R., et al., *The effectiveness of chemotherapy for treatment of high grade astrocytoma in children: results of a randomized trial. A report from the Childrens Cancer Study Group*. *J Neurooncol*, 1989. **7**(2): p. 165-77.
34. Finlay, J.L., et al., *Randomized phase III trial in childhood high-grade astrocytoma comparing vincristine, lomustine, and prednisone with the eight-drugs-in-1-day regimen. Childrens Cancer Group*. *J Clin Oncol*, 1995. **13**(1): p. 112-23.
35. Pollack, I.F., et al., *The influence of central review on outcome associations in childhood malignant gliomas: results from the CCG-945 experience*. *Neuro Oncol*, 2003. **5**(3): p. 197-207.
36. Stupp, R., et al., *Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma*. *N Engl J Med*, 2005. **352**(10): p. 987-96.
37. Cohen, K.J., et al., *Temozolomide in the treatment of high-grade gliomas in children: a report from the Children's Oncology Group*. *Neuro Oncol*, 2011. **13**(3): p. 317-23.
38. Jones, C., L. Perryman, and D. Hargrave, *Paediatric and adult malignant glioma: close relatives or distant cousins?* *Nat Rev Clin Oncol*, 2012. **9**(7): p. 400-13.

39. Tanaka, S., et al., *Diagnostic and therapeutic avenues for glioblastoma: no longer a dead end?* Nat Rev Clin Oncol, 2013. **10**(1): p. 14-26.
40. Li, A., et al., *Unsupervised analysis of transcriptomic profiles reveals six glioma subtypes.* Cancer research, 2009. **69**(5): p. 2091-9.
41. Verhaak, R.G., et al., *Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1.* Cancer Cell, 2010. **17**(1): p. 98-110.
42. *Comprehensive genomic characterization defines human glioblastoma genes and core pathways.* Nature, 2008. **455**(7216): p. 1061-8.
43. Frezza, C., D.A. Tennant, and E. Gottlieb, *IDH1 mutations in gliomas: when an enzyme loses its grip.* Cancer Cell, 2010. **17**(1): p. 7-9.
44. Parsons, D.W., et al., *An integrated genomic analysis of human glioblastoma multiforme.* Science, 2008. **321**(5897): p. 1807-12.
45. Cooper, L.A., et al., *The proneural molecular signature is enriched in oligodendrogliomas and predicts improved survival among diffuse gliomas.* PLoS One, 2010. **5**(9): p. e12548.
46. Noushmehr, H., et al., *Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma.* Cancer Cell, 2010. **17**(5): p. 510-22.
47. Frattini, V., et al., *The integrated landscape of driver genomic alterations in glioblastoma.* Nat Genet, 2013. **45**(10): p. 1141-9.
48. Danussi, C., et al., *RHPN2 drives mesenchymal transformation in malignant glioma by triggering RhoA activation.* Cancer Res, 2013. **73**(16): p. 5140-50.
49. Carro, M.S., et al., *The transcriptional network for mesenchymal transformation of brain tumours.* Nature, 2010. **463**(7279): p. 318-25.
50. Bhat, K.P., et al., *The transcriptional coactivator TAZ regulates mesenchymal differentiation in malignant glioma.* Genes Dev, 2011. **25**(24): p. 2594-609.
51. Brennan, C., et al., *Glioblastoma subclasses can be defined by activity among signal transduction pathways and associated genomic alterations.* PLoS One, 2009. **4**(11): p. e7752.
52. Sumazin, P., et al., *An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma.* Cell, 2011. **147**(2): p. 370-81.
53. Paugh, B.S., et al., *Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease.* J Clin Oncol, 2010. **28**(18): p. 3061-8.
54. Phillips, H.S., et al., *Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis.* Cancer Cell, 2006. **9**(3): p. 157-73.
55. TCGA, *Comprehensive genomic characterization defines human glioblastoma genes and core pathways.* Nature, 2008. **455**(7216): p. 1061-8.
56. Brennan, C.W., et al., *The somatic genomic landscape of glioblastoma.* Cell, 2013. **155**(2): p. 462-77.
57. Sturm, D., et al., *Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma.* Cancer Cell, 2012. **22**(4): p. 425-37.
58. Paugh, B.S., et al., *Integrated Molecular Genetic Profiling of Pediatric High-Grade Gliomas Reveals Key Differences With the Adult Disease.* J Clin Oncol, 2010.

59. Bax, D.A., et al., *A distinct spectrum of copy number aberrations in pediatric high-grade gliomas*. Clin Cancer Res, 2010. **16**(13): p. 3368-77.
60. Sturm, D., et al., *Paediatric and adult glioblastoma: multiform (epi)genomic culprits emerge*. Nat Rev Cancer, 2014. **14**(2): p. 92-107.
61. Furnari, F.B., et al., *Malignant astrocytic glioma: genetics, biology, and paths to treatment*. Genes Dev, 2007. **21**(21): p. 2683-710.
62. Zarghooni, M., et al., *Whole-genome profiling of pediatric diffuse intrinsic pontine gliomas highlights platelet-derived growth factor receptor alpha and poly (ADP-ribose) polymerase as potential therapeutic targets*. J Clin Oncol, 2010. **28**(8): p. 1337-44.
63. Wu, G., et al., *The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma*. Nat Genet, 2014. **46**(5): p. 444-50.
64. Barrow, J., et al., *Homozygous loss of ADAM3A revealed by genome-wide analysis of pediatric high-grade glioma and diffuse intrinsic pontine gliomas*. Neuro Oncol, 2011. **13**(2): p. 212-22.
65. Wu, G., et al., *Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas*. Nat Genet, 2012. **44**(3): p. 251-3.
66. Pollack, I.F., et al., *Age and TP53 mutation frequency in childhood malignant gliomas: results in a multi-institutional cohort*. Cancer Res, 2001. **61**(20): p. 7404-7.
67. Kathryn R Taylor, A.M., Nathalène Truffaux, Yaron Butterfield, Olena Morozova, Cathy Philippe, David Castel, Catherine S Grasso, Maria Vinci, Diana Carvalho, Angel M Carcaboso, Carmen de Torres, Ofelia Cruz, Jaume Mora, Natacha Entz-Werle, Wendy J Ingram, Michelle Monje, Darren Hargrave, Alex N Bullock, Stéphanie Puget, Stephen Yip, Chris Jones, Jacques Grill, *Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma*. Nature Genetics, 2014.
68. Buczkowicz, P., et al., *Genomic analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent activating ACVR1 mutations*. Nat Genet, 2014. **46**(5): p. 451-6.
69. Schwartzenuber, J., et al., *Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma*. Nature, 2012. **482**(7384): p. 226-31.
70. Fontebasso, A.M., et al., *Recurrent somatic mutations in ACVR1 in pediatric midline high-grade astrocytoma*. Nat Genet, 2014. **46**(5): p. 462-6.
71. Lewis, P.W., et al., *Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma*. Science, 2013. **340**(6134): p. 857-61.
72. Bjerke, L., et al., *Histone H3.3 Mutations Drive Pediatric Glioblastoma through Upregulation of MYCN*. Cancer Discov, 2013.
73. Fontebasso, A.M., et al., *Mutations in SETD2 and genes affecting histone H3K36 methylation target hemispheric high-grade gliomas*. Acta Neuropathol, 2013. **125**(5): p. 659-69.
74. Li, F., et al., *The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSalpha*. Cell, 2013. **153**(3): p. 590-600.
75. Mitelman, F., B. Johansson, and F. Mertens, *The impact of translocations and gene fusions on cancer causation*. Nat Rev Cancer, 2007. **7**(4): p. 233-45.
76. Stephens, P.J., et al., *Massive genomic rearrangement acquired in a single catastrophic event during cancer development*. Cell, 2011. **144**(1): p. 27-40.

77. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome*. Nature, 2009. **458**(7239): p. 719-24.
78. Frohling, S. and H. Dohner, *Chromosomal abnormalities in cancer*. The New England journal of medicine, 2008. **359**(7): p. 722-34.
79. Weischenfeldt, J., et al., *Phenotypic impact of genomic structural variation: insights from and for human disease*. Nat Rev Genet, 2013. **14**(2): p. 125-38.
80. Mills, R.E., et al., *Mapping copy number variation by population-scale genome sequencing*. Nature, 2011. **470**(7332): p. 59-65.
81. Mitelman, F. *Cancer cytogenetics update 2005*. *Atlas Genet Cytogenet Oncol Haematol.* 2005; Available from: <http://AtlasGeneticsOncology.org/Deep/CancerCytogenet2005ID20050.html>.
82. Albertson, D.G., et al., *Chromosome aberrations in solid tumors*. Nat Genet, 2003. **34**(4): p. 369-76.
83. Nambiar, M., V. Kari, and S.C. Raghavan, *Chromosomal translocations in cancer*. Biochim Biophys Acta, 2008. **1786**(2): p. 139-52.
84. Vogelstein, B. and K.W. Kinzler, *Cancer genes and the pathways they control*. Nat Med, 2004. **10**(8): p. 789-99.
85. Mani, R.S. and A.M. Chinnaiyan, *Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences*. Nat Rev Genet, 2010. **11**(12): p. 819-29.
86. Hiom, K., *Coping with DNA double strand breaks*. DNA repair, 2010. **9**(12): p. 1256-63.
87. Hastings, P.J., et al., *Mechanisms of change in gene copy number*. Nat Rev Genet, 2009. **10**(8): p. 551-64.
88. Raphael, B.J., *Chapter 6: Structural variation and medical genomics*. PLoS Comput Biol, 2012. **8**(12): p. e1002821.
89. Kloosterman, W.P., et al., *Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline*. Human molecular genetics, 2011. **20**(10): p. 1916-24.
90. Kloosterman, W.P., et al., *Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer*. Genome biology, 2011. **12**(10): p. R103.
91. Malhotra, A., et al., *Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms*. Genome Res, 2013. **23**(5): p. 762-76.
92. Zhang, Y., et al., *The role of mechanistic factors in promoting chromosomal translocations found in lymphoid and other cancers*. Advances in immunology, 2010. **106**: p. 93-133.
93. Vandeweyer, G. and R.F. Kooy, *Detection and interpretation of genomic structural variation in health and disease*. Expert Rev Mol Diagn, 2013. **13**(1): p. 61-82.
94. Hungerford, P.N.a.D., *A minute chromosome in human chronic granulocytic leukemia*. Science, 1960(132): p. 1497.
95. Edwards, P.A., *Fusion genes and chromosome translocations in the common epithelial cancers*. J Pathol, 2010. **220**(2): p. 244-54.
96. Heim, S. and F. Mitelman, *Molecular screening for new fusion genes in cancer*. Nat Genet, 2008. **40**(6): p. 685-6.

97. Baker, M., *Structural variation: the genome's hidden architecture*. Nat Methods, 2012. **9**(2): p. 133-7.
98. Forshew, T., et al., *Activation of the ERK/MAPK pathway: a signature genetic defect in posterior fossa pilocytic astrocytomas*. J Pathol, 2009. **218**(2): p. 172-81.
99. Jones, D.T., et al., *Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas*. Cancer Res, 2008. **68**(21): p. 8673-7.
100. Alkan, C., B.P. Coe, and E.E. Eichler, *Genome structural variation discovery and genotyping*. Nat Rev Genet, 2011. **12**(5): p. 363-76.
101. Bashir, A., et al., *Evaluation of paired-end sequencing strategies for detection of genome rearrangements in cancer*. PLoS Comput Biol, 2008. **4**(4): p. e1000051.
102. Metzker, M.L., *Sequencing technologies - the next generation*. Nat Rev Genet, 2010. **11**(1): p. 31-46.
103. Chen, K., et al., *BreakDancer: an algorithm for high-resolution mapping of genomic structural variation*. Nat Methods, 2009. **6**(9): p. 677-81.
104. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq*. Bioinformatics, 2009. **25**(9): p. 1105-11.
105. Abel, H.J. and E.J. Duncavage, *Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches*. Cancer Genet, 2013. **206**(12): p. 432-40.
106. Mardis, E.R., *The impact of next-generation sequencing technology on genetics*. Trends Genet, 2008. **24**(3): p. 133-41.
107. Druker, B.J., *STI571 (Gleevec) as a paradigm for cancer therapy*. Trends Mol Med, 2002. **8**(4 Suppl): p. S14-8.
108. Tomlins, S.A., et al., *Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer*. Science, 2005. **310**(5748): p. 644-8.
109. Tomlins, S.A., et al., *Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer*. Nature, 2007. **448**(7153): p. 595-9.
110. Robinson, D.R., et al., *Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer*. Nature medicine, 2011. **17**(12): p. 1646-51.
111. Salzman, J., et al., *ESRRA-C11orf20 is a recurrent gene fusion in serous ovarian carcinoma*. PLoS biology, 2011. **9**(9): p. e1001156.
112. Soda, M., et al., *Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer*. Nature, 2007. **448**(7153): p. 561-6.
113. Goldman, J.M. and J.V. Melo, *Chronic myeloid leukemia--advances in biology and new approaches to treatment*. N Engl J Med, 2003. **349**(15): p. 1451-64.
114. Rowley, J.D., *A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining*. Nature, 1973. **243**(5405): p. 290-293.
115. Druker, B.J., et al., *Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome*. N Engl J Med, 2001. **344**(14): p. 1038-42.
116. Druker, B.J., et al., *Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia*. N Engl J Med, 2001. **344**(14): p. 1031-7.

117. Tomlins, S.A., et al., *ETS gene fusions in prostate cancer: from discovery to daily clinical practice*. Eur Urol, 2009. **56**(2): p. 275-86.
118. Markert, E.K., et al., *Molecular classification of prostate cancer using curated expression signatures*. Proceedings of the National Academy of Sciences of the United States of America, 2011.
119. Carver, B.S., et al., *Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate*. Nature genetics, 2009. **41**(5): p. 619-24.
120. Tomlins, S.A., et al., *Role of the TMPRSS2-ERG gene fusion in prostate cancer*. Neoplasia, 2008. **10**(2): p. 177-88.
121. Shaw, A.T., U. Yasothan, and P. Kirkpatrick, *Crizotinib*. Nature reviews. Drug discovery, 2011. **10**(12): p. 897-8.
122. Kwak, E.L., et al., *Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer*. The New England journal of medicine, 2010. **363**(18): p. 1693-703.
123. Sievert, A.J., et al., *Duplication of 7q34 in pediatric low-grade astrocytomas detected by high-density single-nucleotide polymorphism-based genotype arrays results in a novel BRAF fusion gene*. Brain Pathol, 2009. **19**(3): p. 449-58.
124. Cin, H., et al., *Oncogenic FAM131B-BRAF fusion resulting from 7q34 deletion comprises an alternative mechanism of MAPK pathway activation in pilocytic astrocytoma*. Acta neuropathologica, 2011. **121**(6): p. 763-74.
125. Palanisamy, N., et al., *Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma*. Nat Med, 2010. **16**(7): p. 793-8.
126. Zhang, J., et al., *Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas*. Nat Genet, 2013. **45**(6): p. 602-12.
127. Charest, A., et al., *Fusion of FIG to the receptor tyrosine kinase ROS in a glioblastoma with an interstitial del(6)(q21q21)*. Genes Chromosomes Cancer, 2003. **37**(1): p. 58-71.
128. Bralten, L.B., et al., *Integrated genomic profiling identifies candidate genes implicated in glioma-genesis and a novel LEO1-SLC12A1 fusion gene*. Genes Chromosomes Cancer, 2010. **49**(6): p. 509-17.
129. Ozawa, T., et al., *PDGFRA gene rearrangements are frequent genetic events in PDGFRA-amplified glioblastomas*. Genes Dev, 2010. **24**(19): p. 2205-18.
130. Singh, D., et al., *Transforming fusions of FGFR and TACC genes in human glioblastoma*. Science, 2012. **337**(6099): p. 1231-5.
131. Parker, B.C., et al., *The tumorigenic FGFR3-TACC3 gene fusion escapes miR-99a regulation in glioblastoma*. J Clin Invest, 2013. **123**(2): p. 855-65.

2. Aims

2. AIMS

The main goal of this thesis was to better understand the molecular mechanisms underlying the most malignant form of paediatric brain tumours, high grade gliomas (HGG). I decided to study the structural chromosome rearrangements that lead to the occurrence of gene fusions in pHGG. This was an understudied topic in childhood brain tumours and its investigation could lead to the discovery of novel targets for the treatment of this lethal disease.

This objective encompasses the following specific aims:

- Characterize the presence and frequency of *PDGFRA* molecular alterations – mutations and structural rearrangements – in pHGG.
- Investigate novel structural rearrangements associated with copy number alterations in pHGG.
- Utilise whole genome sequencing approaches to identify novel structural rearrangements that lead to fusion genes in paediatric malignant gliomas.

3. *PDGFRA* alterations in paediatric high grade glioma

The results presented throughout this chapter were:

(i) Published as original articles in international peer reviewed journals:

S. Puget, C. Philippe, D. A. Bax, B. Job, P. Varlet, M. P. Junier, F. Andreiuolo, D. Carvalho, R. Reis, L. Guerrini-Rousseau, T. Roujeau, P. Dessen, C. Richon, V. Lazar, G. Le Teuff, C. Sainte-Rose, B. Georger, G. Vassal, C. Jones, and J. Grill, *Mesenchymal Transition and PDGFRA Amplification/Mutation Are Key Distinct Oncogenic Events in Pediatric Diffuse Intrinsic Pontine Gliomas*, PLoS One, 2012, 7(2): p e30313.

B.S. Paugh, X. Zhu, C. Qu, R. Endersby, J. Zhang, D. A. Bax, D. Carvalho, R.M. Reis, A. Onar-Thomas, A. Broniscer, C. Wetmore, C. Jones, J. Zhang, D. W. Ellison, S. J. Baker, *Novel Oncogenic PDGFRA Mutations in Pediatric High-Grade Gliomas*, *Cancer Research*, 2013, 73(20): p. 6219-29

3. *PDGFRA* ALTERATIONS IN PAEDIATRIC HIGH GRADE GLIOMA

3.1. INTRODUCTION

RTKs are important mediators of signalling networks that transmit extracellular signals into cells and activate various cascades of intracellular processes that control cell growth, proliferation, differentiation, survival and metabolism [1]. Stimulation of these RTKs either by autocrine/paracrine secretion and binding of growth factors or by genetic alterations in the genes encoding the receptors themselves (amplifications, indels, fusions, mutations) are commonly found in cancer. These alterations result in RTKs with a constitutive or strongly enhanced signalling capacity, which provides rationale for the development of small molecules that selectively inhibit RTK activity and downstream signalling [1, 2].

The PDGFR/PDGF system includes two receptors (*PDGFRA* and *PDGFRB*) and four ligands (*PDGFA*, *PDGFB*, *PDGFC*, and *PDGFD*) which belong to the type III family of RTKs [3]. Receptor dimerization is triggered by ligand binding, enabling autophosphorylation of multiple tyrosine residues and subsequent recruitment of a variety of signalling transduction molecules. Activated PDGFRs transduce signals via numerous downstream pathways including PI3K/Akt, RAS/MAP kinase, Src kinase family and PLC/PKC pathways, which have all been implicated in tumourigenesis [3-5]. *PDGFR* regulates normal cellular growth and differentiation [6], and expression of activated *PDGFR* promotes oncogenic transformation.

PDGFRA is the predominant target of focal amplification in pHGG including DIPG (12% and 30%, respectively), in contrast to the disease in adults, where *EGFR* is the most common target (43% - Classical Subgroup) [7-10]. However, the Proneuronal adult GBM subgroup is also driven by PDGF signalling, with approximately one third of the Proneuronal/PDGF tumours harbouring amplification of the *PDGFRA* locus, while the remainder shows overexpression of the PDGF ligand [10-12]. Furthermore, the gene expression signature associated with *PDGFRA* is significantly overexpressed in pHGG, even in tumours that do not harbour the RTK amplification. Besides *PDGFRA* itself, *PDGFB* ligand is also commonly amplified and/or overexpressed in paediatric GBM, indicating a potential for autocrine-paracrine loops boosting oncogenic signalling pathway through the PDGF network [8, 13].

Abnormally activated *PDGFRA* signalling driven by viral expression of *PDGFB* ligand is sufficient to induce glioma formation *in vivo*, indicating that activation of *PDGFR* pathways is potentially an early event in tumourigenesis [14-16]. Furthermore, simultaneous overexpression of *PDGFB* and loss of TP53 induced murine HGG with increased incidence and shorter latency indicating cooperativity between these pathways [17, 18]. However,

these studies focused on autocrine and paracrine activation of PDGFR signalling pathways by PDGFB ligand overexpression.

Activating mutations in *PDGFRA* have been previously reported in gastrointestinal stromal tumours (GIST) [2, 19, 20]. GIST mutations occur in 6-7% of cases and are mainly located in the TK domain of *PDGFRA* (exons 12 and 18), leading to a constitutive phosphorylation of *PDGFRA* and ligand-independent kinase activity when expressed *in vitro* [21, 22]. On the other hand, TK domain *PDGFRA* mutations reported in GBM are only sporadic. Rand and colleagues reported a 2-bp deletion in exon 23 (the final exon) within the C-terminus of *PDGFRA* which results in an alternate, shorter C-terminus [19]; four GBM missense mutations (I1050T, W349C, C235Y, V536E) have been reported in the extracellular and TK domain of *PDGFRA* [9, 23]. The functional consequences of these mutations were not analysed. Interestingly, mutations in the extracellular domain of *EGFR* have been detected in 13.6% of cases in a series of 132 adult GBM [24]. These mutations are associated with *EGFR* copy number alterations and with an increased sensitivity to targeted inhibitors of *EGFR* kinase domain and they are oncogenic in a fibroblast cell line [24].

Larger structural rearrangements involving *PDGFRA* are a more common event in adult GBMs. Ozawa and colleagues reported an in-frame deletion of 81 amino acids (aa) in the immunoglobulin-like (Ig-like) domains of *PDGFRA* (termed *PDGFRA* Δ 8,9) to be present in 40% of adult GBM with *PDGFRA* amplification (Figure 3.1) [25-27].

The mutant *PDGFRA* has the deletion in the extracellular domain with retention of normal transmembrane and intracellular domains. This produces a novel transcript fusing exons 7–10 resulting in loss of coding regions for part of the fourth and fifth Ig-like domains due to a complete loss of *PDGFRA* gene exons 8 and 9. Interestingly, this type of deletion is reminiscent of the oncogenic deletion frequently observed in human GBMs, which generates a constitutively active *EGFRvIII* protein [28-34]. Moreover, a gene fusion between the kinase insert domain receptor (KDR) TK domain (VEGFR2) and *PDGFRA* extracellular domain was identified the same study. The *KDR:PDGFRA* (*KP*) fusion protein consists of an extracellular domain containing the first to sixth Ig-like domain of KDR and the disrupted fifth Ig-like domain of *PDGFRA*, followed by the intact transmembrane domain and TK domain of *PDGFRA* (Figure 3.1). Both *PDGFRA* rearrangements showed constitutively elevated TK activity and transforming potential that is reversed by PDGFR blockade [27].

To explore the possibility that *PDGFRA* might be a preferential target of oncogenic mutations in the paediatric setting, we screened a large series of cases for single base changes and small indels, as well as the Δ 8,9 and *KP* fusion, informing two studies in collaboration with

colleagues at the Institut Gustav Roussy, Paris, and St Jude Childrens Research Hospital, Memphis.

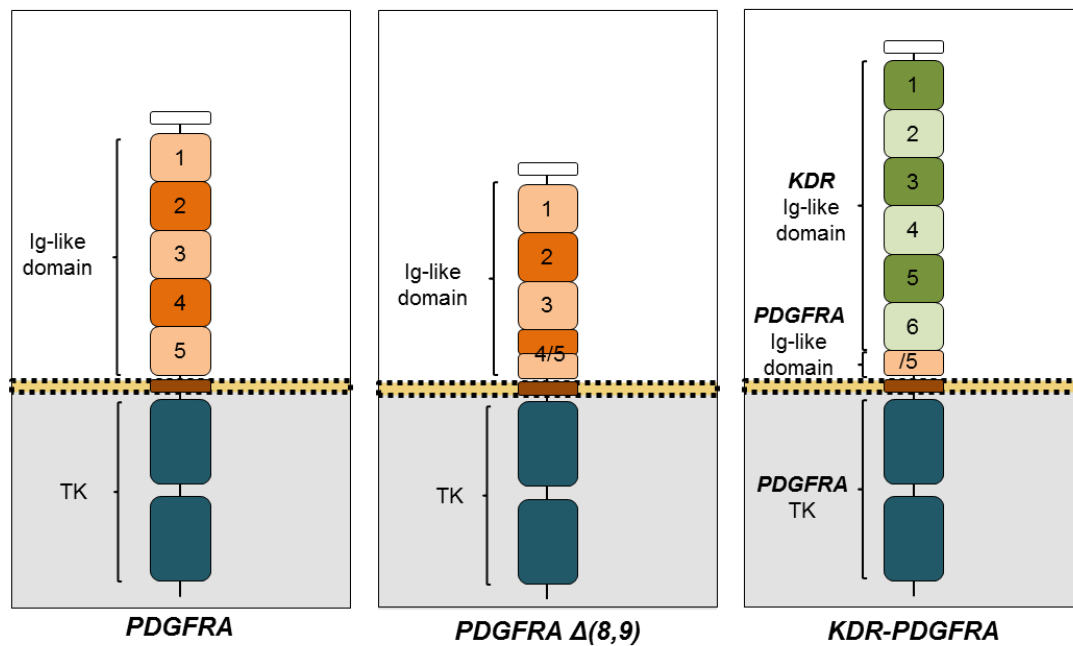


Figure 3.1. Schematic of PDGFRA, PDGFRA Δ 8,9 and KDR:PDGFRA. PDGFRA is constituted by five immunoglobulin-like (Ig-like) domains and two tyrosine kinase (TK) domains. PDGFRA Δ 8,9 has an in-frame deletion of exons 8 and 9 resulting in the deletion of 81 amino acids. This deletion results in loss of approximately half of Ig-like domain 4 and half of Ig-like domain 5. KDR:PDGFRA results from the fusion of the extracellular domain of KDR (Ig-like domains 1-6) with half the Ig-like domain 5 and the TK domain of PDGFRA.

3.2. MATERIALS AND METHODS

3.2.1. Patient samples

A total of 306 tumour samples were included in this study. Both formalin-fixed, paraffin-embedded (FFPE) and snap-frozen HGG specimens from children, young adult (<23 years) and older adults were obtained from the following institutions: King's College Hospital (UK), St George's Hospital (UK), Newcastle Royal Infirmary (UK), Nottingham University Hospital (UK), St Jude Children's Research Hospital (US) and Institut Gustave Roussy (France). (Appendix I).

3.2.2. DNA and RNA extraction

DNA from both frozen and FFPE specimens was extracted using the DNeasy Tissue Kit (Qiagen, Manchester, UK). RNA was extracted from the FFPE samples using the RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE (Life Technologies, Paisley, UK) and cDNA was synthesized using the SuperScript system (Life Technologies) according to the manufacturer's protocol. Nucleic acids were quantified using a ND-1000 Nanodrop (Thermo Scientific, Loughborough, UK).

3.2.3. Screening of novel PDGFRA mutations

All coding exons of *PDGFRA* were PCR amplified from genomic DNA in the tumours listed in Appendix I. Each reaction contained 20ng of genomic DNA and was carried out using Taq DNA Polymerase (Life Technologies) in a final volume of 25 µl. The following program was used: Initial denaturation at 94°C for 3min, followed by a 2-Step-Touchdown: 1. (94°C 30 sec, 68°C 45 sec, 72°C 1 min) 18 cycles, 2. (94°C 30 sec, 50°C 45 sec, 72°C 1 min) 30 cycles; followed by an additional cycle of 72°C 10 min. The primers used are listed in Table 3.1. DNA sequencing of the PCR products was performed by DNA Sequencing & Services (MRCPPU, College of Life Sciences, University of Dundee, Scotland, www.dnaseq.co.uk) using Applied Biosystems Big-Dye Ver 3.1 chemistry on an Applied Biosystems model 3730 automated capillary DNA sequencer. Sequences were analysed using Mutation Surveyor (SoftGenetics, US) and Chromas Lite 2.1 (Technelysium, Australia).

3.2.4. Screening of PDGFRA Δ 8,9 and KDR:PDGFRA fusion gene

83 paediatric plus 75 adult cases were screened by RT-PCR for the 243 bp deletion in exons 8 and 9, and 83 samples were screened for the *KP* fusion (Appendix I). The KDR-Exon9-F (K9) / PDGFRA-Exon12-R (P12) primer pair was used for the detection of the *KP* fusion. The screening of the PDGFRA Δ 8,9 was performed with PDGFRA Δ 8,9 Ex7_ Fwd1/ Ex10_ Rev1 and PDGFRA Δ 8,9 Ex9_ Fwd1/ Ex10_ Rev2 primer pairs for detection of both wild type and mutant allele (Table 3.1). PCR was performed as described in the previous section using 20ng of cDNA.

3.2.5. Statistical analysis

Association between histological tumour type and the presence of PDGFRA mutation in pHGG was assessed by exact chi-square test for independence

Table 3.1. Primers sequences used in *PDGFRA* mutations screen.

Primer	Sequence
PDGFRA Ex2 Fwdb	CTAGGCTCCAGGGTTGTTTCT
PDGFRA Ex2 Revb	AGGAACTCAGAGAGGACTGGG
PDGFRA Ex3a Fwdb	AGACTGTCCTTTCTGACTGCATC
PDGFRA Ex3a Revb	CATTTCTGATTTCCACATCGG
PDGFRA Ex3b Fwdb	CCGATGTGGAAATCAGAAATG
PDGFRA Ex3b Revb	CAGGAAGAGGAGAAGAAGCTTGGT
PDGFRA Ex4a Fwdb	CTGGATTTATGTGTAAAGGTGAAAT
PDGFRA Ex4a Revb	TTGTGTAAGGTTACAGGAGTCTCG
PDGFRA Ex4b Fwdb	ACTGATCCCGAGACTCCTGTAA
PDGFRA Ex4b Revb	GCACCTTATGATTTTGCCTGTT
PDGFRA Ex5 Fwdb	CCTGTGGATTTTTAGGCCCTT
PDGFRA Ex5 Revb	AAGCCATTGCACGTTTTGA
PDGFRA Ex6 Fwdb	TTCACTCCTAGGAGCGAGCT
PDGFRA Ex6 Revb	GAGCAGCATGGACAACTGAC
PDGFRA Ex7 Fwdb	CGGGATCCATATGTGGTAATC
PDGFRA Ex7 Revb	TGGGCAGAGAGTTTCTTTACCT
PDGFRA Ex8 Fwdb	GGAACCTACTTAGCTACTGCTTGTT
PDGFRA Ex8 Revb	CCTTGGAAGACACTCATCTACAGA
PDGFRA Ex9 Fwdb	CACGAGCTATTCCATTCTGACTT
PDGFRA Ex9 Revb	ACCACGAAAGAAGAAGACACATC
PDGFRA Ex10 Fwdb	GAATTGGCCCTATACTTAGGCC
PDGFRA Ex10 Revb	GTTGTCCTGACTGTTGAGGAACT
PDGFRA Ex11 Fwdb	TTTCATTGTGCCTCTCTCTTTG
PDGFRA Ex11 Revb	GCTATGCTTGTTCATTGGC
PDGFRA Ex12 Fwdb	CTGGGACTTTGGTAATTCACCA
PDGFRA Ex12 Revb	AGTCTTGGGAGGTTACCCCAT
PDGFRA Ex18 Fwdb	GATGGCTTGATCCTGAGTCA
PDGFRA Ex18 Revb	GCCTGACCAGTGAGGGAA
PDGFRA Ex23 Fwdb	GCAGGAGTTGTAATATTTGCTCTTC
PDGFRA Ex23 Revb	CAGTTACAGGAAGCTGTCTTCCA
PDGFRA _{del8,9} Ex7 Fwd1	TGAGATCACCCTGATGTGGA
PDGFRA _{del8,9} Ex9 Fwd1	CATCCATTCTGGACTTGGT
PDGFRA _{del8,9} Ex10 Rev1	TGGCCAAAATAGTCCAGGAA
PDGFRA _{del8,9} _Ex10_Rev2	CTGTCTCGGGAGTGGATCTC
KDR-Exon9-F (K9)	CCCTTGAGTCCAATCACACA
PDGFRA-E12-R (P12)	GGTATGAAATTCGCTGGAGG
KDR-Exon13-F (K13)	AATCCATGTGGGAGAGTTGC

3.3. RESULTS

Encouraged by the fact that *PDGFRA* plays an important role in pHGG and that oncogenic alterations in *PDGFRA* have been previously reported in other cancers including adult GBM, we decided to sequence all coding exons of *PDGFRA* in pHGG samples and screen for previously reported structural rearrangements (Figure 3.1). Two different cohorts of samples were used in this study. The first cohort (cohort 1) consisted of a series of 34 DIPGs biopsies. The biopsy material was obtained at the Neurosurgery Department of Necker Sick Children's Hospital in Paris and after DNA and RNA extraction, was also used for aCGH and gene expression profile at the Institut Gustave Roussy, France. The second cohort (cohort 2) contained 133 (90 pHGGs outside the brainstem and 43 DIPGs) samples that were obtained from St Jude Children's Research Hospital, USA and the Royal Marsden Hospital, UK. Both cohorts were sequenced for novel mutations in *PDGFRA* and a broad spectrum of alterations was discovered in both: novel missense mutations, small structural rearrangements (in-frame indels) and two previously reported large structural rearrangements – *KP* and *PDGFRAΔ8,9*.

Overall, 18 cases as well as further two pHGG established as primary xenografts with altered *PDGFRA* were identified. All alterations were confirmed by independent PCR reactions and were validated as somatic, found only in the tumour and not matching germline DNA where a matched normal sample was available. In cohort 1, novel missense mutations were only observed in the *PDGFRA* extracellular domain in 3/34 (8.8%) cases, and in 2 additional primary xenografts. 66% (2/3) of the mutations were heterozygous and were also amplified. All mutations were found in DIPG cases. In cohort 2, 15/123 (12.2%) cases carried *PDGFRA* alterations, of which 60% (9/15) were present as heterozygous alleles, and 40% (6/15) occurred in cases with *PDGFRA* amplification. SNP array analyses previously showed copy number imbalances for the majority of the samples, demonstrating sufficient tumour purity to detect clonal alterations [8, 35]. Importantly, all mutations reported were readily detected by Sanger sequencing, indicating clonal expansion of the population containing the mutation, particularly in the cases in which *PDGFRA* was not amplified. In cohort 2, there was no association between the presence of *PDGFRA* mutation and histopathological features ($p=0.26$). Three mutations were found in anaplastic astrocytomas, one in an anaplastic oligodendroglioma, and the remaining mutations in grade IV GBMs, including two DIPGs.

The coding mutations (G79D, E229K, C235R, C235Y, T276P, Y288C, C290R, T345H, W449C, R479Q, N659K, D842V) were located in the extracellular, transmembrane and TK domains of *PDGFRA* (Figure 3.2). The mutations found in the kinase domain, N659K and

D842V, were previously reported in GIST but not in gliomas [20, 36]. One of the mutations in the IGR-G82 paediatric glioma xenograft has been previously described in adult GBM (C235Y) [9]. In addition, 2 of the DIPG biopsy samples mutated in cohort 2 were included in a gene expression profile study, as referred before. Both mutant-positive cases were part of the group 1 DIPG which is characterized by a Proneural phenotype, oligodendroglial features (Appendix II Figure 4A) and is largely driven by *PDGFRA*, either by amplification (Appendix II Figure 5D) and/or these novel missense mutations (Figure 3.2, Appendix II Figure 5G).

Moreover, 6 small structural rearrangements were identified in the extracellular domain of *PDGFRA* across 8 different patients. Three in-frame insertions (C450ins, A491ins and V544ins) and 3 in-frame deletions (E7del, E10del and E10del2) (Figure 3.3) not found in previous studies of a large collection of pHGGs and aHGGs [37, 38]. Three of the identified rearrangements (E7del, E10del2 and N659K) were recurrent, each of them found in 2 different HGG cases. Analysis of cDNA showed that one of the somatic point mutations, N468S, generated a new splice site that removed 13 amino acids (E10del2) (Figure 3.3).

We also evaluated whether activating *PDGFRA* alterations previously shown in aHGG occur frequently in our series of pHGG. *PDGFRA* Δ 8,9, an in-frame deletion of 243 base pairs is a known transforming oncogene due to ligand-independent receptor activation [26, 27].

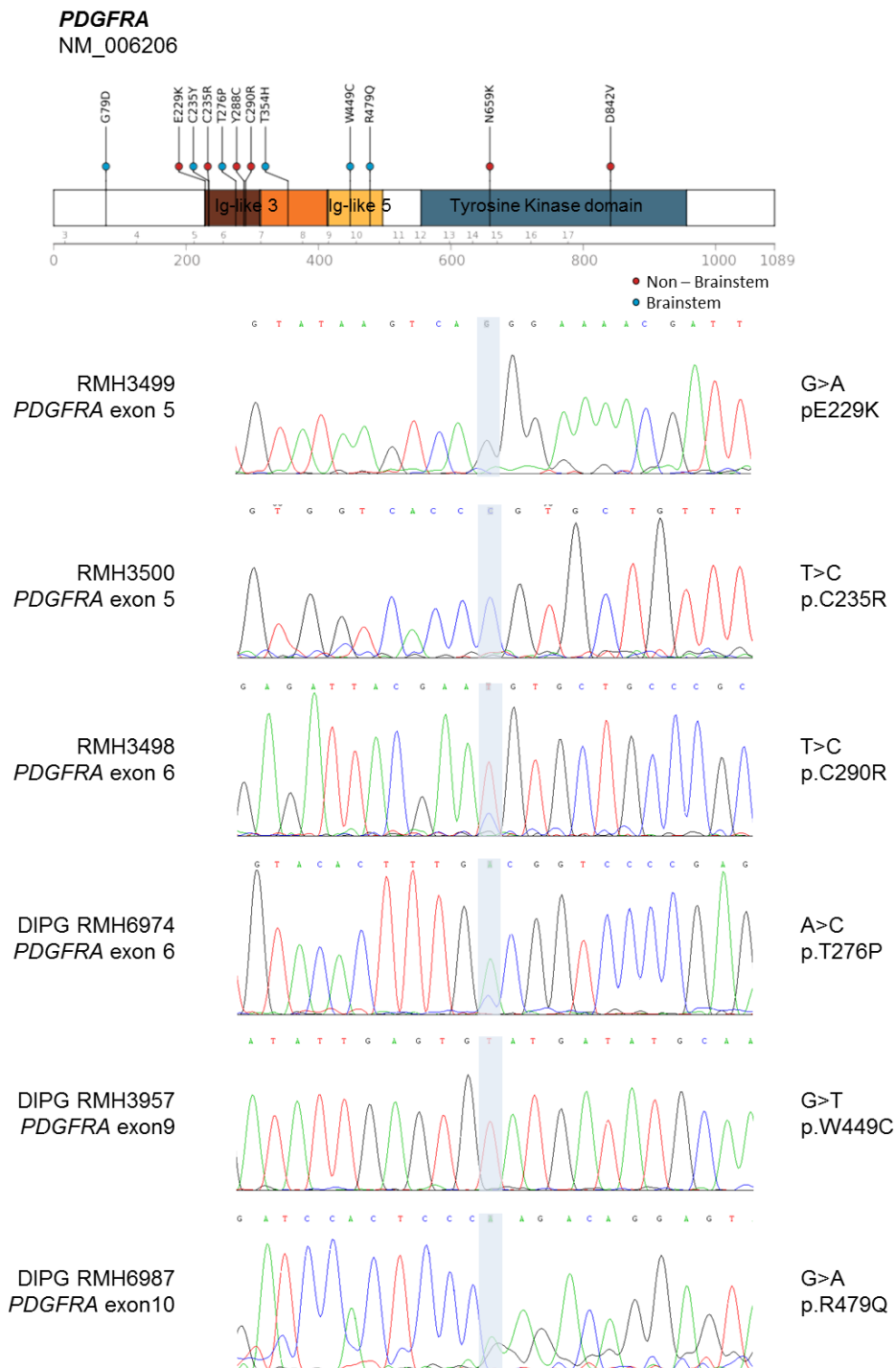


Figure 3.2. Missense mutations in *PDGFRA*. Cartoon showing missense mutations in *PDGFRA* overlaid with functional protein domains and exon boundaries. Chromatograms of three non-brainstem high grade glioma (HGG) cases - RMH3499 (E229K), RMH3500 (C235R) and RMH3496 (C290R) and three brainstem HGG cases – RMH 6974 (T276P), RMH3957 (W449C) and RMH6987 (R479Q). Ig-like: Immunoglobulin-like.

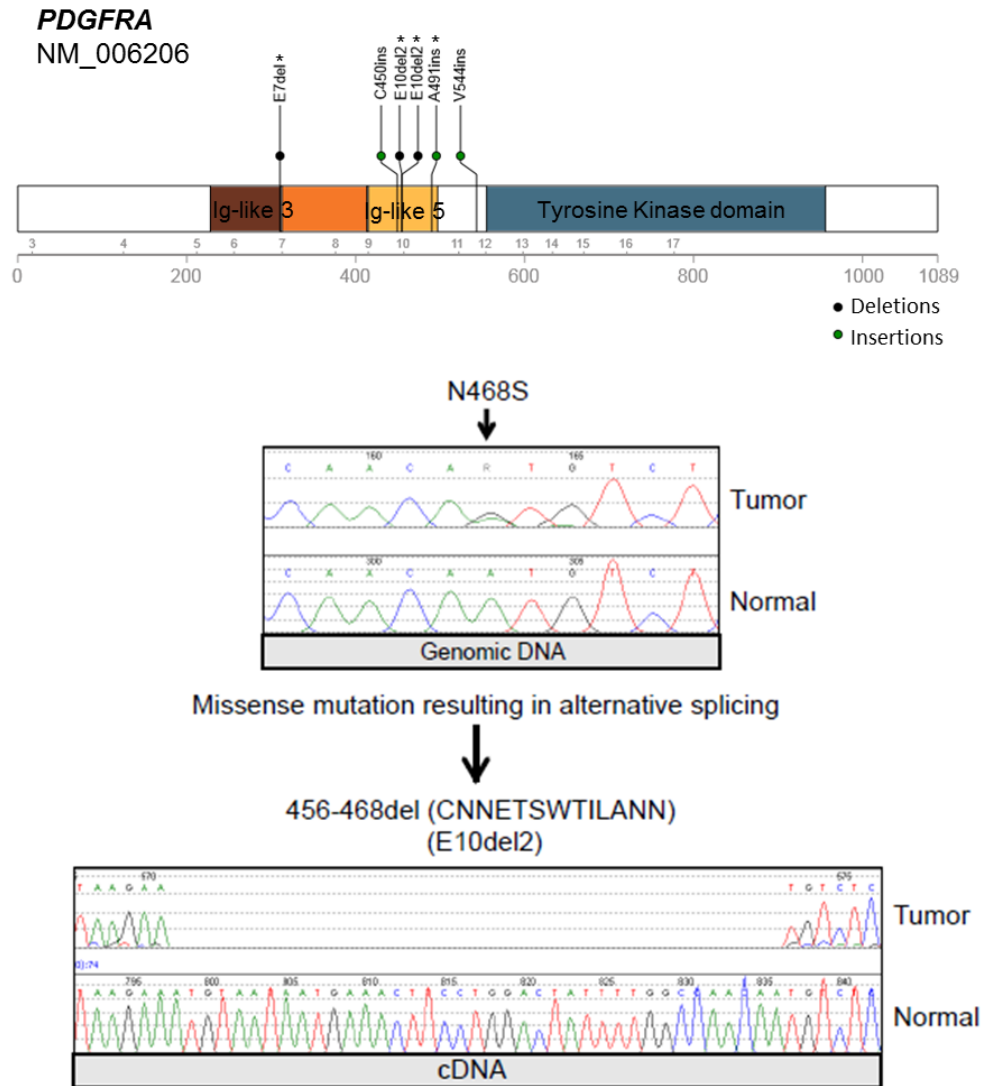


Figure 3.3. Structural rearrangements in *PDGFRA*. Cartoon showing small insertions and deletions present in *PDGFRA* overlaid with functional protein domains and exon boundaries. Sequencing chromatograms are shown for HGG028 for genomic DNA and cDNA. Sanger sequencing of *PDGFRA* from tumour and normal reference sample identified a heterozygous somatic missense mutation in genomic DNA from HGG028 and 2HGG171T (N468S). Subsequent validation of expression of the mutated allele by sequencing cDNA amplified by RT-PCR revealed deletion of 13 amino acids, indicating that this point mutation lead to generation of an alternative splicing site and subsequent deletion of amino acids 456-468 within exon 10 (E10del2) rather than encoding an amino acid substitution.

We used RT-PCR to screen for *PDGFRA* Δ 8,9 mutant in pHGG and in an extended series of aHGGs (Appendix I). All adult cases harboured *PDGFRA* amplification. Two cases of *PDGFRA* Δ 8,9 transcripts were identified in the adult group but none were found in the paediatric samples (Figure 3.4). Both mutant tumours expressed wild-type and deletion mutant transcripts. A gene fusion involving *PDGFRA* and *KDR* (Figure 3.1), which rendered a fusion transcript receptor constitutively active and tumorigenic *in vivo*, that has been reported in 1 out of 215 cases of adult GBMs [27]. Since *PDGFRA* is a common target in pHGG, we sought to use RT-PCR to screen for the *KP* fusion in pHGG. We found 1 case of *KP* in 83 non-brainstem pHGG analysed (Figure 3.5). In contrast to the case found in the *T. Ozawa* study, *PDGFRA* was not amplified (Figure 3.5), thus the mechanism driving the rearrangement was not associated with copy number alteration. However, the exact same gene regions were involved in the fusion, creating the same *KP* transcript as reported before (Figure 3.5).

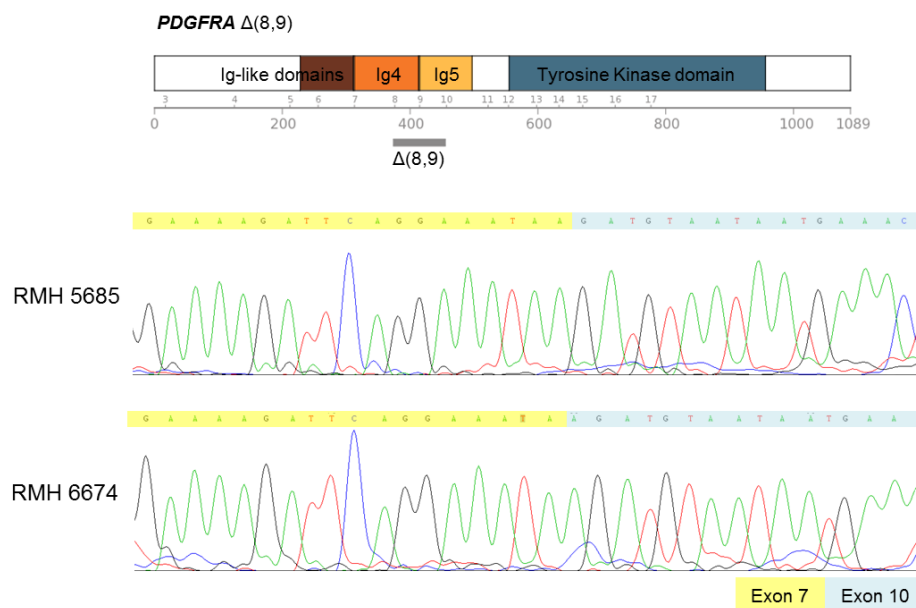


Figure 3.4. Identification of the *PDGFRA* Δ 8,9 fusion in two adult high grade glioma tumours. Cartoon showing *PDGFRA* Δ 8,9 gene structure. Sequencing chromatograms are shown for RMH5685 and RMH6674 illustrating the sequence fusion of exon 7 and exon 10 in *PDGFRA*.

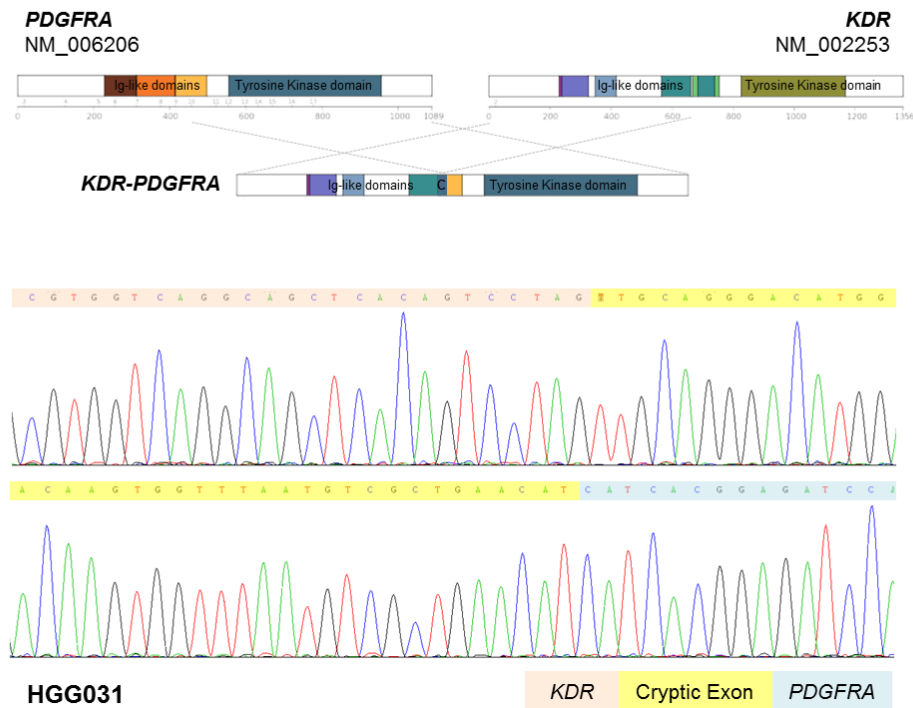


Figure 3.5. Identification of the *KDR:PDGFRA* (*KP*) fusion in one paediatric high grade glioma tumour. Schematic diagram showing the formation of the *KP* fusion transcript, as well as the gene structure of *PDGFRA* and *KDR*. Partial sequence of *KP* fusion transcript found in case HGG031.

3.4. DISCUSSION

The study described in this chapter contributed to two publications, Puget S et al PlosOne 2012 (Appendix II) and Paugh S. et al Cancer Research 2013 (Appendix III). The former was focussed on DIPG, and provided evidence that *PDGFRA* amplification/mutation was associated with a specific subgroup of tumours, whilst the latter demonstrated that the novel mutations identified in pHGG were constitutively activating and tumourigenic.

Aberrations of *PDGFRA* signalling via amplification and/or mutation and/or structural rearrangement are a frequent event in pHGG. In adult GBMs, the frequency of *PDGFRA* amplification and mutations is lower than in paediatric tumours and *PDGFRA* mutations comprise a distinct set of alterations compared to childhood disease (*PDGFRA* Δ 8,9). In this study we identified novel somatic mutations and small indels in pHGG, as well as a case with a previously reported fusion gene, *KP*. A recent study reported *PDGFRA* mutations in pHGG, but none of the described mutations overlapped with our findings [37]. Interestingly, the novel mutations were not represented by a single hotspot as observed in GIST, but mainly spread across the extracellular domain of *PDGFRA*.

The mutations found in Cohort 1 were reported in *Puget S et al* study and contributed to the characterization of two molecularly distinct subgroups of DIPG (Appendix II Figure 2B). Group 1 DIPG was characterized by a Proneural phenotype, an oligodendroglial differentiation, and *PDGFRA* amplification/mutation, whereas group 2 DIPG exhibited a mesenchymal and pro-angiogenic phenotype. The gene expression profile of group 1 DIPG was significantly enriched with genes describing the signature of *PDGFRA* amplified gliomas [39, 40] supporting the hypothesis that *PDGFRA* amplification/mutation is associated with a robust gene expression profile across tumour location and patient's age (Appendix II Figure 5A). It is now known that the different gene expression profile is driven by the recently discovered mutations in the histone genes, H3.3 (*H3F3A*) and H3.1 (*HIST1H3B*), which are unique to DIPG tumours [41, 42]. This *PDGFRA* gene expression signature has also been previously described in adult tumours [11, 43-45], and includes the expression of genes involved in neurogenesis and oligodendrocyte development, such as Olig transcription factors, *PDGFRA* and *SOX10* [46]. DIPG with oligodendroglial phenotype and *Olig2* overexpression exhibited an even worse evolution and resistance to radiation than the other DIPG in this series (Appendix II, Figure 4F). This could be explained by the fact that the CNS-restricted transcription factor *Olig2* opposes p53 response to genotoxic damage in neural progenitors and malignant glioma [47]. In addition, Group I is linked with the K27M mutation in the H3.3 variant which is also associated with a worst outcome [48]. Interestingly, this is in contrast to the adult gliomas where oligodendroglial differentiation and proneural phenotype are linked with a better prognosis [10]. Moreover, we did not observe *IDH1/2* mutations in 10 DIPG, while in adult Proneural gliomas *IDH1* mutations are frequent. Furthermore, a recent study by K. Taylor et al has also confirmed that *IDH1* mutations are absent in a large series of DIPG biopsies [49]. Integrative genomics showed that the gene expression profiles of this group of DIPG were driven by copy number changes, in contrast to the other group, suggesting that chromosomal instability plays an important role in the phenotype of these tumours. We found 28% (9/32) of *PDGFRA* gains or amplifications, all but one being included in the group 1 defined by unsupervised gene expression clustering. In addition, the *PDGFRA* amplified cases are also associated with H3.3 K27M mutation, but not H3.1 K27M (J. Grill unpublished observations). The PDGF autocrine/paracrine loop has been frequently implicated in oligodendrogliomas [50] and has been used to create preclinical models of glioma [16, 51], including brainstem tumours [52, 53]. *PDGFRA* amplification has been shown to be more frequent in pHGG than in adult , and a recent report found *PDGFRA* gain or amplification in four out of eleven post-mortem samples of DIPG [39, 54].

Cohort 2 *PDGFRA* missense mutations and structural rearrangements were reported in the Paugh S et al. study (Appendix III). Contrary to the previous study, several in-frame indels were detected in this series and these were distributed among different regions of the *PDGFRA* protein, including the kinase domain. On the other hand, only 1 case of pHGG carried the *KP* fusion and none harboured the *PDGFRA* Δ 8,9, a recurrent alteration in adult GBM [25, 27]. Thus alternate mechanisms are used to generate oncogenic mutations of *PDGFRA* in childhood and adult HGGs. The role of six of the novel *PDGFRA* variants described in this chapter was explored. The *PDGFRA* mutants analysed represented alterations in different functional domains, and all were constitutively active and tumourigenic. In a p53-null background the D842V, V544ins, C450ins, E10del, E10del2 and E7del *PDGFRA* mutants showed ligand-independent phosphorylation while wild-type receptor activation required ligand stimulation (Appendix III Figure 3A), suggesting lack of sufficient ligand, or different selective advantage conferred by wild-type *PDGFRA* alteration *versus* mutation with or without amplification. Both wild-type and mutant PDGFR α conferred a proliferative advantage to p53-null primary mouse astrocytes *in vitro*, with all 6 mutants inducing HGG with an activated *PDGFRA* signalling (Appendix III Figure 4D). In contrast to the previous study, the tumours arising from *PDGFRA* mutants were not associated with a specific gene expression signature, instead they displayed a range of gene expression signatures similar to the spectrum seen in human HGG (Proneural, Proliferative and Mesenchymal) (Appendix III Figure 5). When mutant cells were treated with small molecule inhibitors including dasatinib a potent multi-TK inhibitor, and crenolanib, a more specific *PDGFRA* inhibitor, *PDGFRA* signalling was abrogated (Appendix III Figure 3A) and proliferation was inhibited (Appendix III Figure 3B). In contrast to *PDGFRA* D842V, a recurrent hotspot mutation in GIST that confers resistance to imatinib [20], the mutations found in this study do not confer resistance to available small molecule PDGFR inhibitors, although these compounds induced a cytostatic, but not cytotoxic response. The resistance/sensitivity to certain inhibitors is mainly dependent on the specific mutation rather than the therapeutic target, as illustrated by the different spectrum of *EGFR* mutations in adult GBM that are not sensitive to the small molecule inhibitors that work well in lung cancer [55]. This data suggests that clinical use of PDGFR inhibitors as a single agent may not be sufficient to cause regression of pHGG, but could be a useful addition to other therapeutic approaches.

In summary, this chapter demonstrated that *PDGFRA* alterations are a common event in both non-brainstem and brainstem pHGG. These alterations range from missense mutations to small indels and are important for tumour development. Moreover, they have helped characterize two biologically and clinically different subgroups of DIPG.

3.5. REFERENCES

1. Zwick, E., J. Bange, and A. Ullrich, *Receptor tyrosine kinases as targets for anticancer drugs*. Trends Mol Med, 2002. **8**(1): p. 17-23.
2. Burger, H., et al., *Activating mutations in c-KIT and PDGFRalpha are exclusively found in gastrointestinal stromal tumors and not in other tumors overexpressing these imatinib mesylate target genes*. Cancer Biol Ther, 2005. **4**(11): p. 1270-4.
3. Andrae, J., R. Gallini, and C. Betsholtz, *Role of platelet-derived growth factors in physiology and medicine*. Genes Dev, 2008. **22**(10): p. 1276-312.
4. Tallquist, M. and A. Kazlauskas, *PDGF signaling in cells and mice*. Cytokine Growth Factor Rev, 2004. **15**(4): p. 205-13.
5. Heldin, C.H., A. Ostman, and L. Ronnstrand, *Signal transduction via platelet-derived growth factor receptors*. Biochim Biophys Acta, 1998. **1378**(1): p. F79-113.
6. Wang, C. and B. Song, *Cell-type-specific expression of the platelet-derived growth factor alpha receptor: a role for GATA-binding protein*. Mol Cell Biol, 1996. **16**(2): p. 712-23.
7. Bax, D.A., et al., *A distinct spectrum of copy number aberrations in pediatric high-grade gliomas*. Clin Cancer Res, 2010. **16**(13): p. 3368-77.
8. Paugh, B.S., et al., *Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease*. J Clin Oncol, 2010. **28**(18): p. 3061-8.
9. TCGA, *Comprehensive genomic characterization defines human glioblastoma genes and core pathways*. Nature, 2008. **455**(7216): p. 1061-8.
10. Verhaak, R.G., et al., *Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1*. Cancer Cell, 2010. **17**(1): p. 98-110.
11. Phillips, H.S., et al., *Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis*. Cancer Cell, 2006. **9**(3): p. 157-73.
12. Brennan, C., et al., *Glioblastoma subclasses can be defined by activity among signal transduction pathways and associated genomic alterations*. PLoS One, 2009. **4**(11): p. e7752.
13. Martinho, O., et al., *Expression, mutation and copy number analysis of platelet-derived growth factor receptor A (PDGFRA) and its ligand PDGFA in gliomas*. Br J Cancer, 2009. **101**(6): p. 973-82.
14. Uhrbom, L., et al., *Induction of brain tumors in mice using a recombinant platelet-derived growth factor B-chain retrovirus*. Cancer Res, 1998. **58**(23): p. 5275-9.
15. Dai, C., et al., *PDGF autocrine stimulation dedifferentiates cultured astrocytes and induces oligodendrogliomas and oligoastrocytomas from neural progenitors and astrocytes in vivo*. Genes Dev, 2001. **15**(15): p. 1913-25.
16. Lindberg, N., et al., *Oligodendrocyte progenitor cells can act as cell of origin for experimental glioma*. Oncogene, 2009. **28**(23): p. 2266-75.
17. Hesselager, G., et al., *Complementary effects of platelet-derived growth factor autocrine stimulation and p53 or Ink4a-Arf deletion in a mouse glioma model*. Cancer Res, 2003. **63**(15): p. 4305-9.

18. Hambardzumyan, D., et al., *Genetic modeling of gliomas in mice: new tools to tackle old problems*. *Glia*, 2011. **59**(8): p. 1155-68.
19. Rand, V., et al., *Sequence survey of receptor tyrosine kinases reveals mutations in glioblastomas*. *Proc Natl Acad Sci U S A*, 2005. **102**(40): p. 14344-9.
20. Corless, C.L., et al., *PDGFRA mutations in gastrointestinal stromal tumors: frequency, spectrum and in vitro sensitivity to imatinib*. *J Clin Oncol*, 2005. **23**(23): p. 5357-64.
21. Hirota, S., et al., *Gain-of-function mutations of platelet-derived growth factor receptor alpha gene in gastrointestinal stromal tumors*. *Gastroenterology*, 2003. **125**(3): p. 660-7.
22. Lasota, J., et al., *A great majority of GISTs with PDGFRA mutations represent gastric tumors of low or no malignant potential*. *Lab Invest*, 2004. **84**(7): p. 874-83.
23. Parsons, D.W., et al., *An integrated genomic analysis of human glioblastoma multiforme*. *Science*, 2008. **321**(5897): p. 1807-12.
24. Lee, J.C., et al., *Epidermal growth factor receptor activation in glioblastoma through novel missense mutations in the extracellular domain*. *PLoS Med*, 2006. **3**(12): p. e485.
25. Kumabe, T., et al., *Amplification of alpha-platelet-derived growth factor receptor gene lacking an exon coding for a portion of the extracellular region in a primary brain tumor of glial origin*. *Oncogene*, 1992. **7**(4): p. 627-33.
26. Clarke, I.D. and P.B. Dirks, *A human brain tumor-derived PDGFR-alpha deletion mutant is transforming*. *Oncogene*, 2003. **22**(5): p. 722-33.
27. Ozawa, T., et al., *PDGFRA gene rearrangements are frequent genetic events in PDGFRA-amplified glioblastomas*. *Genes Dev*, 2010. **24**(19): p. 2205-18.
28. Ekstrand, A.J., et al., *Amplified and rearranged epidermal growth factor receptor genes in human glioblastomas reveal deletions of sequences encoding portions of the N- and/or C-terminal tails*. *Proc Natl Acad Sci U S A*, 1992. **89**(10): p. 4309-13.
29. Sugawa, N., et al., *Identical splicing of aberrant epidermal growth factor receptor transcripts from amplified rearranged genes in human glioblastomas*. *Proc Natl Acad Sci U S A*, 1990. **87**(21): p. 8602-6.
30. Nishikawa, R., et al., *A mutant epidermal growth factor receptor common in human glioma confers enhanced tumorigenicity*. *Proc Natl Acad Sci U S A*, 1994. **91**(16): p. 7727-31.
31. Batra, S.K., et al., *Epidermal growth factor ligand-independent, unregulated, cell-transforming potential of a naturally occurring human mutant EGFRvIII gene*. *Cell Growth Differ*, 1995. **6**(10): p. 1251-9.
32. Huang, H.S., et al., *The enhanced tumorigenic activity of a mutant epidermal growth factor receptor common in human cancers is mediated by threshold levels of constitutive tyrosine phosphorylation and unattenuated signaling*. *J Biol Chem*, 1997. **272**(5): p. 2927-35.
33. Holland, E.C., et al., *A constitutively active epidermal growth factor receptor cooperates with disruption of G1 cell-cycle arrest pathways to induce glioma-like lesions in mice*. *Genes Dev*, 1998. **12**(23): p. 3675-85.
34. Feldkamp, M.M., et al., *Normoxic and hypoxic regulation of vascular endothelial growth factor (VEGF) by astrocytoma cells is mediated by Ras*. *Int J Cancer*, 1999. **81**(1): p. 118-24.

35. Paugh, B.S., et al., *Genome-wide analyses identify recurrent amplifications of receptor tyrosine kinases and cell-cycle regulatory genes in diffuse intrinsic pontine glioma*. J Clin Oncol, 2011. **29**(30): p. 3999-4006.
36. Medeiros, F., et al., *KIT-negative gastrointestinal stromal tumors: proof of concept and therapeutic implications*. Am J Surg Pathol, 2004. **28**(7): p. 889-94.
37. Sturm, D., et al., *Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma*. Cancer Cell, 2012. **22**(4): p. 425-37.
38. Persons, D.A., et al., *Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis*. Blood, 1999. **93**(2): p. 488-99.
39. Paugh, B.S., et al., *Integrated Molecular Genetic Profiling of Pediatric High-Grade Gliomas Reveals Key Differences With the Adult Disease*. J Clin Oncol, 2010.
40. Noushmehr, H., et al., *Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma*. Cancer Cell, 2010. **17**(5): p. 510-22.
41. Schwartzenuber, J., et al., *Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma*. Nature, 2012. **482**(7384): p. 226-31.
42. Wu, G., et al., *Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas*. Nat Genet, 2012. **44**(3): p. 251-3.
43. Cooper, L.A., et al., *The proneural molecular signature is enriched in oligodendrogliomas and predicts improved survival among diffuse gliomas*. PLoS One, 2010. **5**(9): p. e12548.
44. Ducray, F., et al., *Anaplastic oligodendrogliomas with 1p19q codeletion have a proneural gene expression profile*. Mol Cancer, 2008. **7**: p. 41.
45. Freije, W.A., et al., *Gene expression profiling of gliomas strongly predicts survival*. Cancer Res, 2004. **64**(18): p. 6503-10.
46. Rowitch, D.H. and A.R. Kriegstein, *Developmental genetics of vertebrate glial-cell specification*. Nature, 2010. **468**(7321): p. 214-22.
47. Mehta, S., et al., *The central nervous system-restricted transcription factor Olig2 opposes p53 responses to genotoxic damage in neural progenitors and malignant glioma*. Cancer Cell, 2011. **19**(3): p. 359-71.
48. Kathryn R Taylor, A.M., Nathalène Truffaux, Yaron Butterfield, Olena Morozova, Cathy Philippe, David Castel, Catherine S Grasso, Maria Vinci, Diana Carvalho, Angel M Carcaboso, Carmen de Torres, Ofelia Cruz, Jaume Mora, Natacha Entz-Werle, Wendy J Ingram, Michelle Monje, Darren Hargrave, Alex N Bullock, Stéphanie Puget, Stephen Yip, Chris Jones, Jacques Grill, *Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma*. Nature Genetics, 2014.
49. Taylor, K.R., et al., *Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma*. Nat Genet, 2014.
50. Ferletta, M., et al., *Sox10 has a broad expression pattern in gliomas and enhances platelet-derived growth factor-B--induced gliomagenesis*. Mol Cancer Res, 2007. **5**(9): p. 891-7.
51. Appolloni, I., et al., *PDGF-B induces a homogeneous class of oligodendrogliomas from embryonic neural progenitors*. Int J Cancer, 2009. **124**(10): p. 2251-9.
52. Masui, K., et al., *Glial progenitors in the brainstem give rise to malignant gliomas by platelet-derived growth factor stimulation*. Glia, 2010. **58**(9): p. 1050-65.

53. Becher, O.J., et al., *Preclinical evaluation of radiation and perifosine in a genetically and histologically accurate model of brainstem glioma*. *Cancer Res*, 2010. **70**(6): p. 2548-57.
54. Zarghooni, M., et al., *Whole-genome profiling of pediatric diffuse intrinsic pontine gliomas highlights platelet-derived growth factor receptor alpha and poly (ADP-ribose) polymerase as potential therapeutic targets*. *J Clin Oncol*, 2010. **28**(8): p. 1337-44.
55. Vivanco, I., et al., *Differential sensitivity of glioma- versus lung cancer-specific EGFR mutations to EGFR kinase inhibitors*. *Cancer Discov*, 2012. **2**(5): p. 458-71.

4. Intragenic copy number breakpoint analysis identifies a novel complex fusion *DHX57:TMEM178:MAP4K3* in paediatric high grade glioma

The results presented throughout this chapter were:

- (i) Published as an original article in an international peer reviewed journal:

D. Carvalho, A. Mackay, L. Bjerke, R.G. Grundy, C. Lopes, R. M. Reis and C. Jones, The prognostic role of intragenic copy number breakpoints and identification of novel fusion genes in pediatric high grade glioma. Acta Neuropathologica Communications, 2014, 2(1): p23

4. INTRAGENIC COPY NUMBER BREAKPOINT ANALYSIS IDENTIFIES A NOVEL COMPLEX FUSION *DHX57:TMEM178:MAP4K3* IN PAEDIATRIC HIGH GRADE GLIOMA

4.1. INTRODUCTION

DNA copy number and gene expression studies have highlighted key distinctions between HGG arising in childhood and far more commonly, much later in adult life [1-4]. Indeed, recent exome-level sequencing initiatives have conclusively shown the existence of subgroups of HGG marked by distinct driver mutations [5], which are significantly enriched in young children (*H3F3A* K27M), teenagers and young adults (*H3F3A* G34R/V), and middle-aged adults (*IDH1/2*) [6]. Specific driving events for infants and elderly patients remain to be elucidated, however they too represent biological sub-entities, with infants having few genomic alterations [2], and elderly patients harbouring frequent amplification of *EGFR* and other genomic events [3, 4].

The identification of driving genetic alterations at the DNA copy level are necessarily most commonly focussed on assessing the amplification/deletion of genes in their entirety, and approaches to ascribe significance to genomic events make use of overlapping regions across multiple samples to find genes consistently within regions of gain/loss [7]. This approach has the result of ignoring genes for whom the breakpoint, *i.e.* the specific location of copy number change, is found within the coding regions. Such events may be more than mere bystanders of the “driving” aberration, and may themselves play significant roles in tumour initiation and maintenance.

One key implication of copy number breakpoints occurring within genes is the possibility of generating novel fusions. Gene fusions can occur through both intra- and inter-chromosomal translocations, bringing together coding regions from two or more genes within a single reading frame allowing expression of a novel protein. Such gene fusions are common in cancer, but have historically been thought to be largely restricted to haematological malignancies and selected solid tumours such as sarcomas. Recent evidence has overturned this, with numerous novel gene fusions being discovered in a wide range of cancer types, exemplified by the identification of common *TMPRSS2:ERG* fusions in prostate cancer [8] and the *EML4:ALK* fusion in NSCLC [9].

The first fusion gene found in GBM was the rearrangement located at an amplified region at chromosome 4q12, resulting in the fusing of the kinase domain of PDGFRA with the regulatory domains of KDR (*VEGFR2*) [10]. This *KDR:PDGFRA* was found to be activating

and tumourigenic, however to date only a single additional case has been found, in a pHGG [11], and thus these fusions do not represent a common event. Another low frequency fusion has more recently been identified in approximately 3% of adult HGG, involving *FGFR1* or *FGFR3* partnering with *TACC1* or *TACC3* [12]. These *FGFR:TACC* fusions have been shown to localize to mitotic spindle poles, have constitutive kinase activity and induce mitotic and chromosomal segregation defects and aneuploidy [12]. The types of integrated analysis that identified these mutations have also begun to identify more common rearrangements, such as numerous fusions involving *EGFR*, the most frequently seen partner producing the *EGFR:SEPT14* fusion demonstrated to activate STAT3 signalling and confer mitogen independence and sensitivity to EGFR inhibition [13].

Such analyses are clearly proving extremely valuable in furthering our understanding of HGG biology and generating novel targets for therapeutic intervention. As similar approaches are yet to be undertaken in the paediatric setting, we have applied an algorithm designed to identify intragenic copy number breakpoints in our previously published study of DNA copy number [2]. We identify numerous potentially functional gene disruptions and a novel validated complex fusion, *DHX57:TMEM178:MAP4K3*.

4.2. MATERIALS and METHODS

4.2.1. Published DNA copy number data

Our lab previously carried out a DNA copy number profiling study of 100 pHGG cases on Affymetrix 500K SNP arrays [2]. The data have been deposited at the Gene Expression Omnibus (GEO, www.ncbi.nlm.nih.gov/geo/) with accession number GSE19578. Copy number assignment was carried out as per the original publication, using Affymetrix Genotyping Analysis Software (GTYPE version 4.s) improved using Bayesian Linear Model with Mahalanobis distance classifier algorithm (BRLMM) and standard dChipSNP normalization and model-based expression algorithms [2]. Log₂-transformed data was used for all subsequent analysis in the present study.

4.2.2. iCNA algorithm

We implemented the iCNA algorithm developed as part of the GTS package under R2.11.0 (cbio.mskcc.org/~brennan) [14]. Breakpoints are calculated according to user-defined 'delta' values representing shifts in log₂ ratios between two contiguous genomic regions after segmenting the copy number data using circular binary segmentation (cbs) [15]. Using a delta of 0.4, breakpoint boundaries are identified and errors estimated by permutation-based calculations of neighbouring probe data. Confidence intervals are assigned and those falling

within the 95% window considered 'high confidence'. An estimate is calculated for the expected rate of breaks for each gene based upon gene size and rate of breaks per sample, with a p value obtained based upon (observed-expected)/standard error. A corrected p value of < 0.05 is considered significant. Manual inspection of copy number plots was undertaken to ensure sufficient probe coverage was present at identified loci in order to prioritize the most convincing breakpoints. Those with substantial gaps at either side of the break were excluded.

4.2.3. Custom oligonucleotide aCGH

We designed two fine-tiling oligonucleotide microarrays to cover the specific amplicons observed at chromosomes 2p22.1 and 10q11.21 This was undertaken using the Agilent custom array design tool e-Array (Agilent, Santa Clara, CA, USA; <https://earray.chem.agilent.com/>), and comprised 700 probes covering 43.56–43.70Mb on chromosome 10 and 5000 probes covering 39–40Mb on chromosome 10 with a median probe interval of 200bp on 2x105K microarray. Due to limited amount of material, DNA was whole genome amplified (WGA) using the GenomePlex® Complete Whole Genome Amplification Kit (Sigma, Gillingham, UK) starting with 10ng of sample and control DNA, and following the manufacturer's protocol. WGA DNA was labelled using the Agilent Genomic DNA ULS labelling kit, hybridised as per manufacturer's instructions, and scanned on the Agilent 2505B Microarray Scanner System. Data has been submitted to ArrayExpress with accession number E-MTAB-2340.

4.2.4. siRNA knockdown

siRNA was carried out using a Dharmacon SMARTpool™ (Dharmacon, Lafayette, CO, USA) against *MAP4K3* [#003588] with paediatric glioma cells SF188, KNS42, UW479, Res259 and Res186 [16] and a panel of breast carcinoma lines. Cells were plated and transfected 24 hours later with siRNA using Lipofectamine RNAiMax™ (Invitrogen, Paisley, UK) as per manufacturer's instructions, alongside transfections of siControl. Twenty four hours following transfection, cells were trypsinised and media replenished after 48 hours and 96 hours, with cell viability assessed after seven days using CellTiter-Glo™ Luminescent Cell Viability Assay (Promega, Madison, WI, USA) as per manufacturer's instructions.

4.3. RESULTS

We have previously carried out DNA copy number profiling of a large series of pHGG samples using Affymetrix 500K SNP arrays, and reported numerous genes encompassed within areas of focal amplification and deletion [2]. We now applied an algorithm (iCNA [14]) designed to identify copy number breakpoints contained within the sequence of known genes. A full schema of the analytical process is given in Figure 4.1.

This algorithm was applied to 100 pHGG and 26 matched normal DNA samples, resulting in the identification of 1099 unique DNA copy number breaks contained within gene sequences across all tumour samples (Appendix IV - Supplementary Table S1). Of these, 479 were found to map to known regions of copy number variations found commonly in the germlines of the general population by cross-referencing the breakpoints with The Centre for Applied Genomics Database of Genomic Variants [17], leaving a total of 620 events.

These were filtered to 500 after excluding those with p values > 0.05, and further reduced to 388 with at least one sample harbouring a given aberration at 'high confidence'. A further 100 of these were excluded as they were also found in at least one of the normal samples profiled, representing either technical artefacts associated with the array platform used, or low frequency normal copy number polymorphisms. A final list of 288 iCNA is provided in Appendix IV - Supplementary Table S2.

Most pHGG samples harboured at least one iCNA (median=3), although seven cases were found to contain none. Several cases were found to contain many more aberrations (maximum=19), though these were in the minority (Figure 4.2a). The number of iCNA events per sample was found to be prognostic in this multi-institutional series of cases, with pHGG containing more than 10 iCNA (n=9) found to have a significantly poorer survival (median=7.8 months), and those with no iCNA a better survival (median=24 months) than the rest of the tumours (median=13.2 months) (p=0.026, log-rank test) (Figure 4.2b).

There were no differences in the number of iCNA between grade III (n=20, median=3, range 0-7) and grade IV tumours (n=58, median=3, range 0-19) (p=0.456, t-test), though the cases with the highest number of iCNA were all grade IV GBM (Figure 4.2c). Similarly, there were no differences between primary tumours (n=68, median=3, range 0-19) and those which arose as secondary malignancies after cranio-spinal irradiation (n=10, median=2.5, range 0-11) (p=0.698, t-test) (Figure 4.2d). Infants (less than 3 years at diagnosis) had significantly fewer iCNA (n=10, median=2.5, range 0-6) than older children (n=68, median=4, range 0-19) (p=0.050, t-test) (Figure 4.2e). Tumours with the K27M mutation in the gene encoding the

histone variant H3.3, *H3F3A*, harboured significantly more iCNA (n=5, median=6, range 1-16) than either G34R/V mutant tumours (n=4, median=4.5, range 3-7) or wild-type (n=14, median=3, range 0-11) (p=0.043, ANOVA) (Figure 4.2f). This was independent of location of tumour, with no differences in number of iCNA between supratentorial GBM (n=51) and DIPG (n=7, p=0.684, t-test).

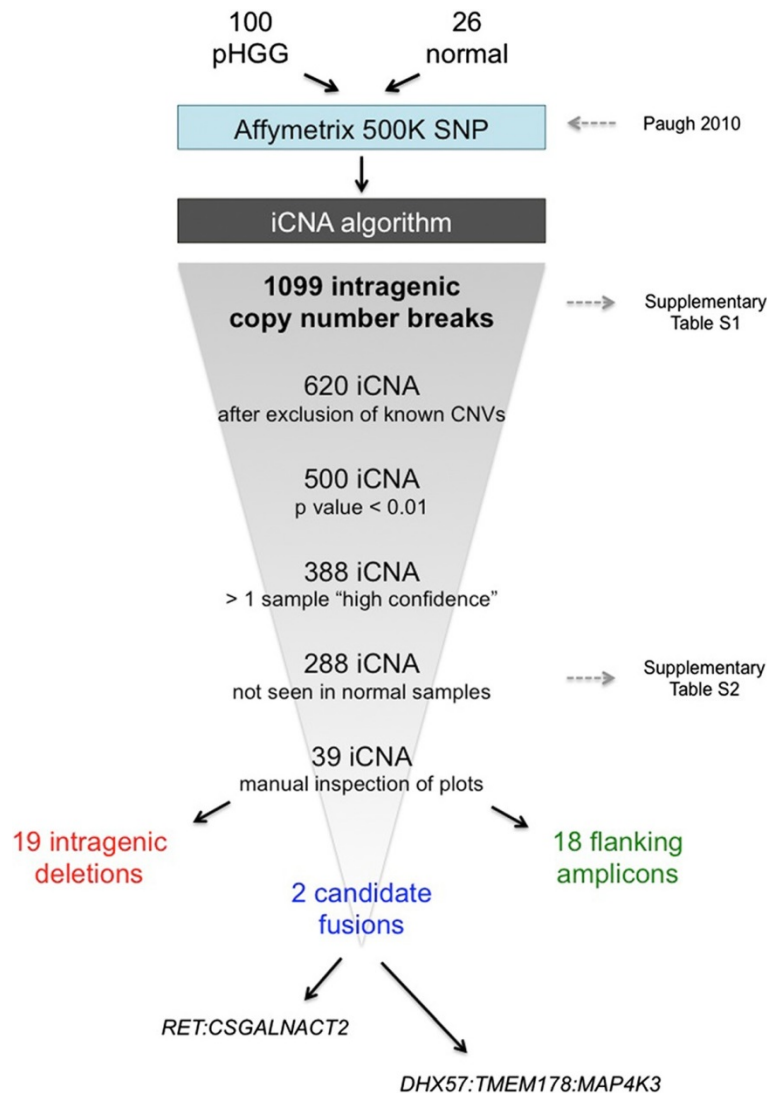


Figure 4.1. Schema of iCNA algorithm applied to paediatric high grade glioma. 1099 intragenic breakpoints were initially identified in a published series of 100 pHGG. From these, known copy number polymorphisms were excluded, as were these seen in a series of 26 matched normal DNA samples. After filtering for statistical significance and manual inspection of copy number plots, a series of 19 intragenic deletions, 18 amplicons, and 2 candidate fusions were identified.

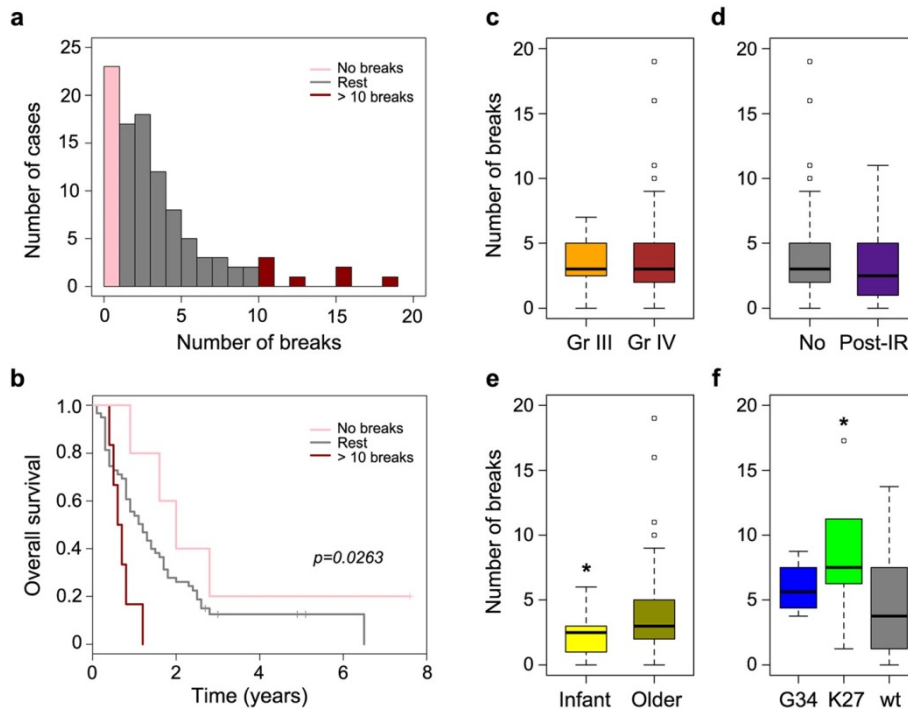


Figure 4.2. Clinicopathological correlates of intragenic copy number breaks in paediatric high grade glioma. (a) Number of intragenic breaks per sample in a series of 100 pHGG. (b) Kaplan-Meier plot of overall survival stratified by number of intragenic breaks per sample. Pink: no breaks; Red: more than 10 breaks per sample; Grey: rest. (c-f) Boxplot of number of intragenic breaks per sample separated by – (c) WHO grade. Orange: grade III; Brown: grade IV. (d) Previous radiation treatment for an earlier malignancy. Grey: primary pHGG; Purple: post-irradiation; (e) Age at diagnosis. Yellow: infant (<3 years old); Dark yellow: older children (>3 years old). (f) *H3F3A* status. Blue: G34R/V; Green: K27M; Grey: wild-type

The 288 iCNA were further subjected to individual manual inspection of the data plots in order to identify the most robust copy number shifts associated with intragenic breaks. This resulted in a list of 39 unique events in 51 samples (Table 4.1). The recurrent changes included copy number loss, resulting primarily in either the absence of the 3' end of a gene or small deletions wholly within the coding sequence. These intragenic deletions included those targeting known tumour suppressors in GBM such as *NF1* (17q11.2, n=2) (Figure 4.3a) and *RB1* (13q14.2, n=1) (Appendix IV - Supplementary Figure S3), as well as putative novel GBM-associated genes including *FAF1* (1p33, n=2) and *MTAP* (9p21.3, n=2) [Appendix IV - Supplementary Figure S3]. In addition, there were novel deletions in the protein phosphatase *PTPRE* (10q26.2, n=2) (Figure 4.3b) and recurrent internal microdeletions in the gene *CSMD3* (CUB and Sushi multiple domains 3) (8q23.3, n=3), all of which overlapped to result in the loss of exon 4 (Appendix IV - Supplementary Figure S3).

Table 4.1. Nominated intragenic copy number aberration candidates. 39 unique intragenic breakpoints found within 51 cases of paediatric high grade glioma. Direction of copy number shift (gain/loss) is reported, as well as candidate fusion events found within 51 cases of pHGG. Direction of copy number shift (gain/loss) is reported, as well as candidate fusion events.

Cand.ID	Gene	Chromosome	Sample	Copy number change	Comments
1	FAF1	1	HGG091	Deletion	
	FAF1	1	HGG140	Deletion	
2	CD84	1	HGG088	Amplification	
3	LGAL	1	HGG070	Deletion	
4	KIDIN	2	HGG077	Amplification	
5	DHX5	2	HGG063	Amplification	Fusion
6	TME	2	HGG063	Amplification	Fusion
7	WDR	3	HGG010	Amplification	
8	PDC	3	HGG157	Deletion	
9	CHIC	4	HGG077	Amplification	
10	ITGA	5	HGG029	Deletion	
11	EPHA	6	HGG139	Deletion	
12	LANC	7	HGG060	Amplification	
13	ECO	7	HGG060	Amplification	
14	KCN	7	HGG152	Amplification	
	KCN	7	HGG162	Amplification	
14	SND1	7	HGG090	Deletion	
15	CSM	8	HGG054	Deletion	
	CSM	8	HGG140	Deletion	
	CSM	8	HGG153	Deletion	
16	SLC2	9	HGG151	Deletion	
	SLC2	9	HGG011	Deletion	
17	MTA	9	HGG022	Deletion	
	MTA	9	HGG007	Deletion	
18	ANK	10	HGG068	Deletion	
19	RET	10	HGG139	Amplification	Fusion
20	CSG	10	HGG139	Amplification	Fusion
21	PTPR	10	HGG086	Deletion	
	PTPR	10	HGG145	Deletion	
22	RAB6	11	HGG092	Amplification	
23	PSM	11	HGG092	Amplification	
24	TMT	12	HGG010	Amplification	
	TMT	12	HGG068	Amplification	
25	LRRK	12	HGG068	Amplification	

26	MYO	12	HGG029	Amplification
27	XRC	12	HGG029	Amplification
28	OSB	12	HGG065	Amplification
	OSB	12	HGG162	Amplification
29	RB1	13	HGG154	Deletion
30	PCD	13	HGG059	Deletion
31	CD27	15	HGG006	Deletion
32	MEF2	15	HGG011	Deletion
33	DNA	17	HGG143	Amplification
34	NF1	17	HGG154	Deletion
	NF1	17	HGG140	Deletion
35	BRIP	17	HGG077	Deletion
36	KCN	20	HGG139	Amplification
37	SYN3	22	HGG017	Deletion
	SYN3	22	HGG146	Deletion
38	TIMP	22	HGG146	Deletion
39	PHF2	22	HGG072	Deletion

39 unique intragenic breakpoints found within 51 cases of paediatric high grade glioma. Direction of copy number shift (gain/loss) is reported, as well as candidate fusion events.

Copy number gains within gene coding regions tended to be associated with regions flanking known oncogenic amplicons. These included amplification of the MYCN locus at chromosome 2p24.3, which in case HGG077 breaks within the coding region of the kinase D-interacting substrate *KIDINS220* (Figure 4.4a); amplification of *PDGFRA* at 4q12, harbouring an iCNA in *CHIC2* in the same case (though only covered by two probes); and recurrent breakpoints in the gene encoding the potassium voltage-gated channel *KCND2* at 7q31.31 in association with amplification of *MET*, though curiously this targeted either 5' or 3' ends in two different cases (Figure 4.4b). Similarly, common amplification events encompassing *EGFR* (7p12) and *CDK4* (12q14) had intragenic breakpoints at both ends in cases HGG060 (*LANCL2* and *ECOP*) and HGG029 (*MYO1A* and *XRCC6BP1*), respectively (Appendix IV - Supplementary Figure S4).

For the most part, iCNA events resulted in an imbalance of certain regions of coding genes in isolation, with the predicted consequence a disruption of full-length gene expression. For certain events however, a 5' end of one gene was found amplified at a similar copy number to a 3' end of a second gene within the same case. We reasoned that such instances may represent candidate fusion genes, and we identified two such examples in our cohort.

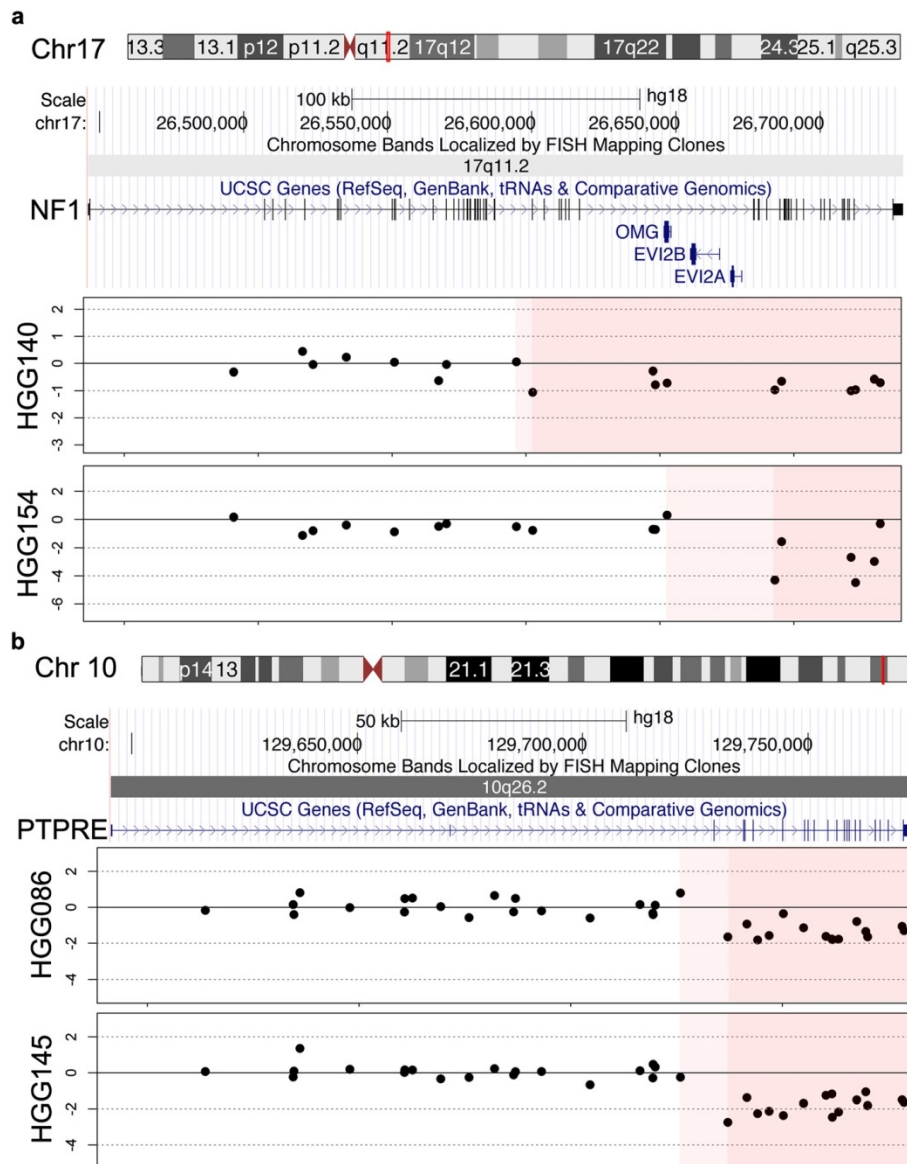


Figure 4.3. Intragenic deletions in paediatric high grade glioma. (a) Recurrent copy number breakpoint within *NF1* on chromosome 17q11.2 in two cases of pHGG. (b) Recurrent copy number breakpoint within *PTPRE* on chromosome 10q26.2 in two cases of pHGG. Dark pink: confirmed region of loss; Light pink: region within which breakpoint lies, as defined by the resolution of probes on the array.

The first was at chromosome 10q11.21 and reflected a single amplicon, breaking within the genes *RET* and *CSGALNACT2* such that we propose a hypothetical fusion gene encompassing the 5' regulatory regions of *CSGALNACT2* and the 3' kinase domain of *RET*. In order to determine the precise breakpoints to allow validation of this novel fusion, we designed custom oligonucleotide arrays spanning the amplicon in order to carry out high-resolution aCGH on the reference case HGG139, a relapse sample of GBM in which this genomic event was not present in the primary tumour. Although the breakpoint for *CSGALNACT2* was identified within intron 2, leaving the catalytic domains intact, the breakpoint within *RET* could not be accurately determined to closer than 10kb between introns 1 and 2 (Appendix IV - Supplementary Figure S5)..

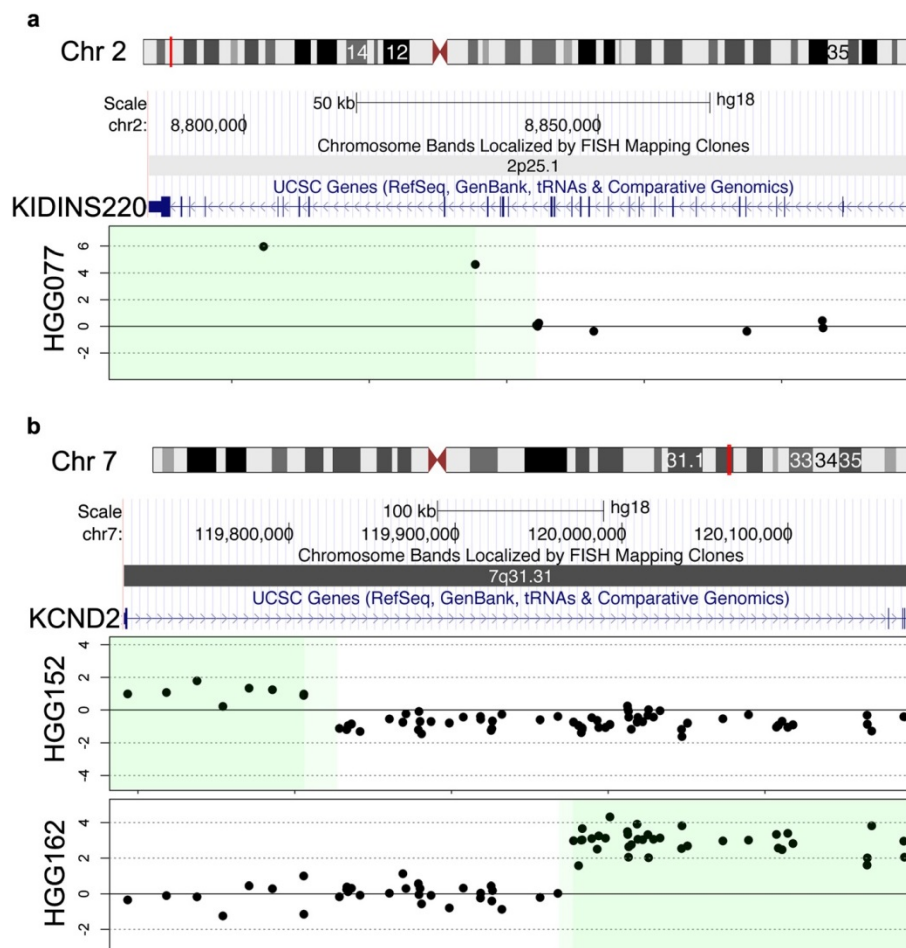


Figure 4.4. Intragenic amplifications in paediatric high grade glioma. (a) Copy number breakpoint within *KIDINS200* on chromosome 2p25.1, flanking the *MYCN* amplicon in a case of pHGG. [b] Recurrent copy number breakpoint within *KCND2* on chromosome 7q31.31 in two cases of pHGG, in both cases part of the *MET* amplicon, though targeting either the 5' or 3' end of the gene. Dark green: confirmed region of gain; Light green: region within which breakpoint lies, as defined by the resolution of probes on the array.

As material was limited for this case, we were unable to confirm the precise nature of the putative *CSGALNACT2:RET* fusion by PCR-based techniques

The second fusion candidate was located at an amplified region of chromosome 2p22.1 in case HGG063, an anaplastic astrocytoma. At Affymetrix 500K SNP resolution, this appeared to be a single amplicon with breaks within the coding regions of the RNA helicase *DHX57* and the transmembrane protein *TMEM178* (Figure 4.5a). Applying the same approach as above, using custom-designed oligonucleotide arrays for high-resolution aCGH revealed two amplicons within this structure, with further intragenic breakpoints within the mitogen-activated protein kinase *MAP4K3* (Figure 4.5b). Designing PCR primers to amplify across the highly specific breakpoints confirmed the presence of the fusion, which was further validated by direct sequencing (Figure 4.6).

The resultant fusion gene, *DHX57:TMEM178:MAP4K3*, is a complex three gene fusion formed from a series of intragenic breaks, amplifications and inversions to produce a sequence comprising exons 1-12 of *DHX57*, exons 2-4 of *TMEM178* and exons 13-34 of *MAP4K3*, associated with regions of microhomology (Figure 4.6). This would produce a protein with the zinger finger and DEAD-like helicase domains of *DHX57*, the claudin family transmembrane domains of *TMEM178* and the citron domain of *MAP4K3*. Selective knockdown of *MAP4K3* by siRNA leads to a significant reduction in cell viability in five paediatric glioma cell lines as assayed by CellTiter Glo, an effect not seen in 18/20 breast cancer cells ($p=0.0017$, pHGG vs breast cancer, t-test) (Appendix IV - Supplementary Figure S6).

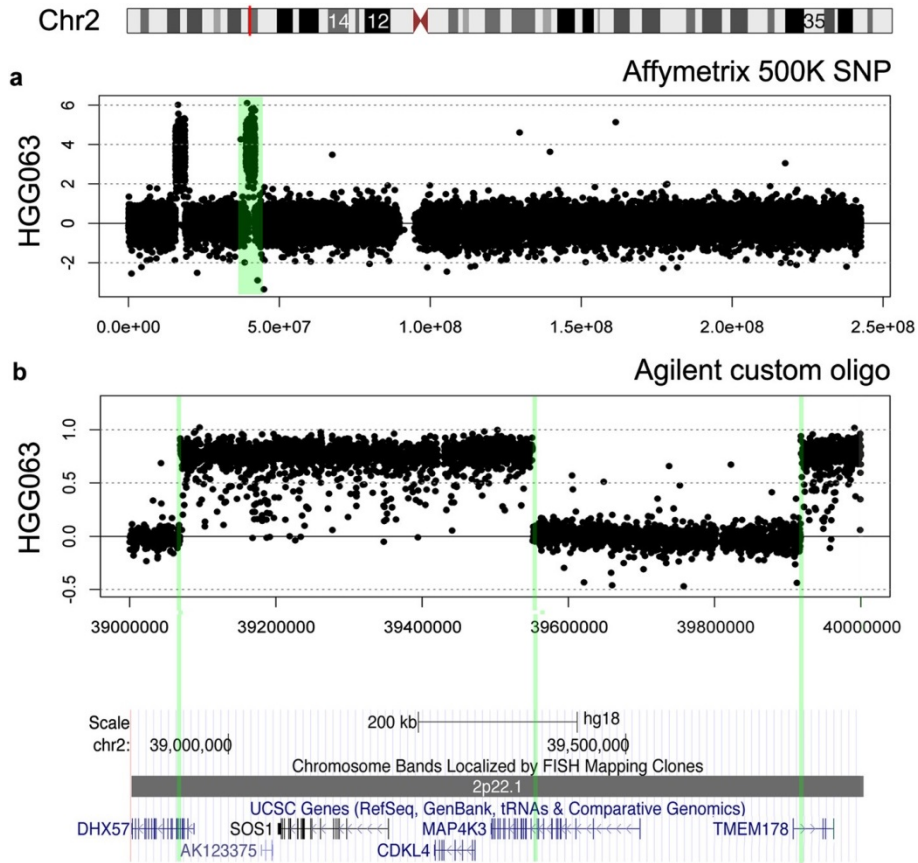


Figure 4.5. Identification of a novel complex fusion *DHX57:TMEM178:MAP4K3*. (a) Affymetrix 500K SNP array of chromosome 2, highlighting two amplicons, the most telomeric encompassing *MYCN*, the more centromeric as 2p22.1 involving intragenic breakpoints in *DHX57* and *TMEM178* (green). (b) Custom oligonucleotide array of the 2p22.1 amplicon, revealing two amplified structures and three intragenic breakpoints, in *DHX57*, *MAP4K3* and *TMEM178* (green).

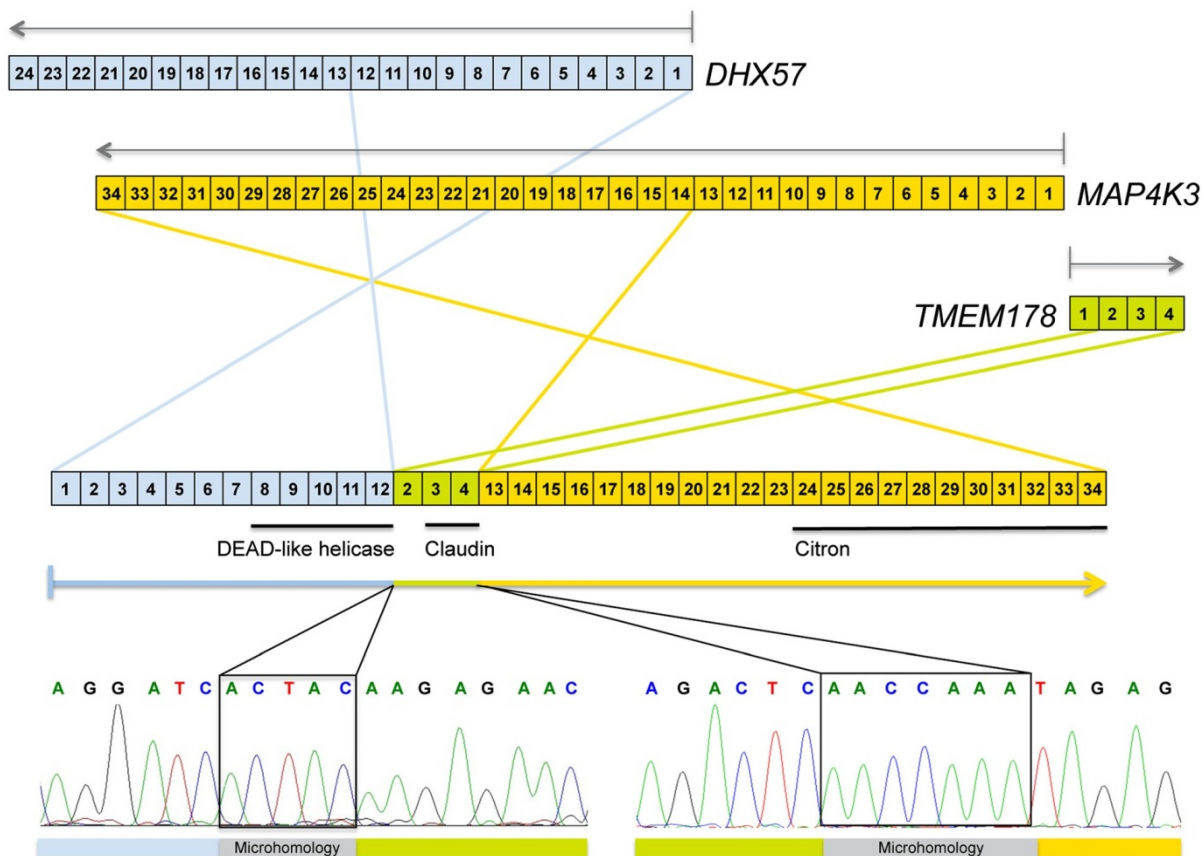


Figure 4.6. Cartoon depicting the structure of the *DHX57:TMEM178:MAP4K3* fusion. A complex microhomology-mediated rearrangement of exons 1-12 of *DHX57* (blue), exons 2-4 of *TMEM178* (green) and exons 13-34 of *MAP4K3* (orange) was confirmed by direct sequencing. Regions of microhomology are highlighted in grey.

4.4. DISCUSSION

Comprehensive copy number profiling of adult and pHGGs was among the first data to demonstrate the biological differences between these similar-looking histological malignancies [18]. In this context, the focus has been on large-scale genomic copy number changes. A more refined analysis of copy number and exon-level expression data has identified new insights into genomic architecture and novel fusion proteins in adult GBM [12, 13]. Here we leverage a large dataset we have previously generated [2] in the paediatric disease to carry out a scan of intragenic breakpoints, leading to the identification of novel gene disruptions and candidate gene fusions.

The presence of intragenic copy number aberrations was confirmed in the vast majority of pHGG cases, and was itself prognostic, with an absence of iCNAs conferring a longer overall survival in paediatric patients. This was associated with the infant age group, known

to have a better clinical outcome than older children [19], and further highlights the biological distinctiveness of this age group. By contrast, the presence of large numbers of intragenic breaks conferred a shorter survival time, but was not a result of the grade of the tumour, nor associated with a second malignancy due to radiation treatment for an earlier cancer. We had previously reported an association of post-irradiated HGG with *PDGFRA* amplification and chromosome 1q gain [2], so it appears these are relatively selective radiation-induced changes, rather than reflecting a generalised genomic instability in secondary tumours from these patients. Importantly, we identified an increased number of iCNA in tumours harbouring an *H3F3A* K27M mutation, regardless of anatomical location. This is a group of thalamic and pontine HGG associated with a particularly dismal prognosis [18], for whom understanding the mechanisms of genomic instability and the identification of novel gene disruptions is of considerable interest.

The majority of intragenic breakpoints we identified were associated with gene disruption. This includes deletions of known tumour suppressors such as *RB1* and *NF1*, but also more novel GBM associated genes. *FAF1* and *MTAP* were both recurrently targeted by intragenic deletion events in pHGG. These genes are localised close to known cyclin-dependent kinase inhibitors and tumour suppressors *CDKN1C* and *CDKN2A/B*, respectively, but both *FAF1* and *MTAP* have recently been proposed to harbour tumour suppressor activity in their own right. *FAF1* is associated with a FAS-mediated apoptosis response and restoration of the FAF1 protein in adult glioma cell lines significantly increases cell death [20], whilst in *MTAP*-deficient cells, methylthioadenosine, generated during polyamine biosynthesis, is not cleaved and the salvage pathway for adenine and methionine is absent. It seems that such mechanisms are also likely in a subset of paediatric tumours.

Of note we identified novel deletions in the protein phosphatase epsilon, *PTPRE*. This has not been reported previously, although there are several reports of the tumour suppressive capacity of the related *PTPRD* [21, 22]. This gene also appears targeted by intragenic deletions, and human astrocytes lacking *PTPRD* exhibited increased growth, as it is thought the protein usually functions to dephosphorylate the oncoprotein STAT3 [21]. The wholly intragenic microdeletions observed in *CSMD3* in four cases may represent another novel mechanism of gene disruption. *CSMD3* encodes a gene with multiple CUB and Sushi domains whose function is poorly understood. Recently, *CSMD3* was identified as the second most frequently mutated gene (next to TP53) in lung cancer, where it was demonstrated that loss of *CSMD3* results in increased proliferation of airway epithelial cells [23].

Gene disruption may also play a significant functional role when known gain-of-function oncogenes are amplified. We report numerous intragenic breakpoints which may have been overlooked in the context of identifying the ‘driver’ event within a common amplicon, but which may themselves be tumourigenic. These include disruptions of *KIDINS220*, a functional mediator of multiple receptor signalling pathways and essential for cortical development [24, 25]; *CHIC2*, frequently deleted/rearranged in myeloid malignancies [26]; and *KCND2*, encoding a potassium voltage-gated channel, which is expressed in both neuronal and glial cells and has been shown to regulate ERK signaling in ganglioglioma [27]. All of these gene disruptions represent novel avenues for understanding the underlying biology of pHGG.

Of most interest was the use of the iCNA algorithm to identify potential novel fusion genes, as was demonstrated in adult GBM with the identification of the *KDR:PDGFRA* fusion [10], which we also found in a case of pHGG [11]. Our analysis nominated two potential candidates – the first we were unable to conclusively validate, *CSGALNACT2:RET*. Such a putative fusion would retain the kinase domain of the RET oncoprotein, but would lose the autoregulatory portion of the protein, instead fusing it to the N terminal of chondroitin sulfate N-acetyl-galactosaminyltransferase 2. Although a precise cancer-related function has not been ascribed to the latter enzyme, it is thought to play an important role in morphogenesis in zebrafish models [28, 29]. Whilst not validated, oncogenic *RET* rearrangements and fusions are common in thyroid and lung cancer [30, 31], and the presence of infrequent activating fusions in HGG do not seem unlikely.

We were able to validate a novel complex fusion involving three genes with intragenic breakpoints and amplification/rearrangement on chromosome 2p22.1. The resulting fusion gene, *DHX57:TMEM178:MAP4K3* encompasses key regulatory domains from all three proteins, though a specific function is hard to predict. The helicase properties of the DHX57 component may be a candidate for oncogenicity, with numerous other DEAD-box helicases appearing to play a role in regulation of DNA repair, apoptosis and drug sensitivity [32]. *MAP4K3* has been associated with several malignancies in both an oncogenic and tumour suppressor capacity [33, 34]. In particular, one function that has been ascribed includes activation of mTOR signalling via the TORC1 complex [35], a pathway commonly activated by diverse mechanisms in pHGG [18].

In the context of pHGG, although the kinase domain is not retained in the fusion, *MAP4K3* plays some functional role as selective knockdown by siRNA leads to a significant and selective reduction in cell viability in paediatric glioma cell lines. Thus we hypothesise that

the *DHX57:TMEM178:MAP4K3* is activating as disruption of the protein would otherwise seem incompatible with tumour cell growth and proliferation.

In summary these data represent a key addition to our understanding of the genomic alterations driving pHGG and provide novel avenues for developing sorely-needed novel therapeutic strategies for children with these otherwise incurable tumours.

4.5. REFERENCES

1. Bax, D.A., et al., *A distinct spectrum of copy number aberrations in paediatric high grade gliomas*. Clin Cancer Res, 2010. **16**[13]: p. 3368-3377.
2. Paugh, B.S., et al., *Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease*. J Clin Oncol, 2010. **28**[18]: p. 3061-8.
3. McLendon, R., et al., *Comprehensive genomic characterization defines human glioblastoma genes and core pathways*. Nature, 2008. **455**: p. 1061-1068.
4. Parsons, D.W., et al., *An integrated genomic analysis of human glioblastoma multiforme*. Science, 2008. **321**[5897]: p. 1807-12.
5. Schwartzenuber, J., et al., *Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma*. Nature, 2012. **482**[7384]: p. 226-31.
6. Sturm, D., et al., *Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma*. Cancer Cell, 2012. **22**[4]: p. 425-37.
7. Beroukhi, R., et al., *Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma*. Proc Natl Acad Sci U S A, 2007. **104**[50]: p. 20007-12.
8. Tomlins, S.A., et al., *Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer*. Science, 2005. **310**[5748]: p. 644-8.
9. Soda, M., et al., *Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer*. Nature, 2007. **448**[7153]: p. 561-6.
10. Ozawa, T., et al., *PDGFRA gene rearrangements are frequent genetic events in PDGFRA-amplified glioblastomas*. Genes Dev, 2010. **24**[19]: p. 2205-18.
11. Paugh, B.S., et al., *Novel Oncogenic PDGFRA Mutations in Pediatric High-Grade Gliomas*. Cancer Res, 2013.
12. Singh, D., et al., *Transforming fusions of FGFR and TACC genes in human glioblastoma*. Science, 2012. **337**[6099]: p. 1231-5.
13. Frattini, V., et al., *The integrated landscape of driver genomic alterations in glioblastoma*. Nat Genet, 2013. **45**[10]: p. 1141-9.
14. Wiedemeyer, R., et al., *Feedback circuit among INK4 tumor suppressors constrains human glioblastoma development*. Cancer Cell, 2008. **13**[4]: p. 355-64.
15. Olshen, A.B., et al., *Circular binary segmentation for the analysis of array-based DNA copy number data*. Biostatistics, 2004. **5**[4]: p. 557-72.
16. Bax, D.A., et al., *Molecular and phenotypic characterisation of paediatric glioma cell lines as models for preclinical drug development*. PLoS ONE, 2009. **4**[4].

17. Iafrate, A.J., et al., *Detection of large-scale variation in the human genome*. Nat Genet, 2004. **36**[9]: p. 949-51.
18. Jones, C., L. Perryman, and D. Hargrave, *Paediatric and adult malignant glioma: close relatives or distant cousins?* Nat Rev Clin Oncol, 2012. **9**[7]: p. 400-13.
19. Wu, W., et al., *Joint NCCTG and NABTC prognostic factors analysis for high-grade recurrent glioma*. Neuro Oncol, 2010. **12**[2]: p. 164-72.
20. Zheng, S., et al., *A survey of intragenic breakpoints in glioblastoma identifies a distinct subset associated with poor survival*. Genes Dev, 2013. **27**[13]: p. 1462-72.
21. Veeriah, S., et al., *The tyrosine phosphatase PTPRD is a tumor suppressor that is frequently inactivated and mutated in glioblastoma and other human cancers*. Proc Natl Acad Sci U S A, 2009. **106**[23]: p. 9435-40.
22. Solomon, D.A., et al., *Mutational inactivation of PTPRD in glioblastoma multiforme and malignant melanoma*. Cancer Res, 2008. **68**[24]: p. 10300-6.
23. Liu, P., et al., *Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing*. Carcinogenesis, 2012. **33**[7]: p. 1270-6.
24. Neubrand, V.E., et al., *Kidins220/ARMS as a functional mediator of multiple receptor signalling pathways*. J Cell Sci, 2012. **125**[Pt 8]: p. 1845-54.
25. Cesca, F., et al., *Kidins220/ARMS is an essential modulator of cardiovascular and nervous system development*. Cell Death Dis, 2011. **2**: p. e226.
26. Cools, J., et al., *Fusion of a novel gene, BTL, to ETV6 in acute myeloid leukemias with a t[4;12](q11-q12;p13)*. Blood, 1999. **94**[5]: p. 1820-4.
27. Adams, J.P., et al., *The A-type potassium channel Kv4.2 is a substrate for the mitogen-activated protein kinase ERK*. J Neurochem, 2000. **75**[6]: p. 2277-87.
28. Filipek-Gorniok, B., et al., *Expression of chondroitin/dermatan sulfate glycosyltransferases during early zebrafish development*. Dev Dyn, 2013. **242**[8]: p. 964-75.
29. Holmborn, K., et al., *On the roles and regulation of chondroitin sulfate and heparan sulfate in zebrafish pharyngeal cartilage morphogenesis*. J Biol Chem, 2012. **287**[40]: p. 33905-16.
30. Gainor, J.F. and A.T. Shaw, *Novel targets in non-small cell lung cancer: ROS1 and RET fusions*. Oncologist, 2013. **18**[7]: p. 865-75.
31. Fusco, A. and M. Santoro, *20 years of RET/PTC in thyroid cancer: clinico-pathological correlations*. Arq Bras Endocrinol Metabol, 2007. **51**[5]: p. 731-5.
32. Fuller-Pace, F.V., *DEAD box RNA helicase functions in cancer*. RNA Biol, 2013. **10**[1]: p. 121-32.
33. Lam, D., et al., *MAP4K3 modulates cell death via the post-transcriptional regulation of BH3-only proteins*. Proc Natl Acad Sci U S A, 2009. **106**[29]: p. 11978-83.
34. Wright, J.H., et al., *The STE20 kinase HGK is broadly expressed in human tumor cells and can modulate cellular transformation, invasion, and adhesion*. Mol Cell Biol, 2003. **23**[6]: p. 2068-82.
35. Findlay, G.M., et al., *A MAP4 kinase related to Ste20 is a nutrient-sensitive regulator of mTOR signalling*. Biochem J, 2007. **403**[1]: p. 13-20.

5. Integrated whole genome and transcriptome sequencing identifies novel expressed fusion transcripts in paediatric high grade glioma

The computation analysis presented throughout this chapter was performed by Alan Mackay, Glioma Team, Institute of Cancer Research, Sutton, United Kingdom.

The FISH was performed with the help and valuable technical advice of Anna Burford, Glioma Team, Institute of Cancer Research, Sutton, United Kingdom

5. INTEGRATED WHOLE GENOME AND TRANSCRIPTOME SEQUENCING IDENTIFIES NOVEL EXPRESSED FUSION TRANSCRIPTS IN PAEDIATRIC HIGH GRADE GLIOMA

5.1. INTRODUCTION

GBM is a highly aggressive brain tumour with a dismal clinical outcome, and though less common in the paediatric setting than in adults, it remains one of the biggest reasons for loss of life due to cancer in children. Recent genomic studies have profiled large numbers of GBM samples, leading to characterization of core pathogenic pathways and enabling molecular sub-classification [1-5]. Nevertheless, despite this exhaustive characterization, only few fusion genes have been reported in adult GBM and one in paediatric GBM. Singh and colleagues have recently identified tandem duplication on 4p16.3 which creates a fusion transcript, *FGFR3:TACC3* [6]. The fusion occurs in 3% of adult GBM, displays oncogenic activity and is associated with amplification of both genes [6]. In addition, a less frequent fusion involving *PDGFRA*, a gene commonly mutated in GBM, and *KDR*, the gene that encodes for the VEGF receptor, was found 1 out of 215 cases [7]. The *KP* fusion renders a receptor constitutively active and tumourigenic *in vivo* [7].

Over the past few decades, genomic rearrangements and the creation of fusion genes have been reported to contribute to tumourigenesis in various malignancies, especially leukaemias and sarcomas [8, 9]. More recently, several reports have demonstrated that common epithelial cancers, such as carcinoma of the breast, prostate and lung, may also be the target of fusion genes [8, 10]. Gene fusions can lead to the production of oncogenic fusion proteins, such as the first to be discovered, the *BCR:ABL1* kinase in chronic myelogenous leukaemia [11]. Alternatively, fusion genes can lead to an enhanced expression of oncogenes *via* promoter switching, as seen in the *TMPRSS2:ERG* fusion in prostate cancer [12]. In addition to creating fusion genes, genomic rearrangements in these cancers break many other genes, affecting gene function and leading to gene inactivation [13].

The recent development of high throughput sequencing technologies has enabled the detection of structural alterations in cancer DNA in a systematic way [14]. Most solid tumours fusion transcripts, like the recurrent *VT11A:TCF7L2* fusion in colorectal adenocarcinomas, were identified using paired-end whole genome or whole transcriptome sequencing. Moreover, very complex rearrangements like chromothripsis [15] and chromoanasythesis [16] have been identified with these methods. The first is a cellular crisis that leads to chromosome shattering and the second comprises local rearrangements

with altered gene copy numbers produced by serial, microhomology-mediated template switching during DNA replication.

Fusion transcripts found in patient samples have also been observed in several cell lines such as prostate cancer [17, 18], breast cancer [19] and NSCLC [20]. KNS42, SF188 and UW479 are pHGG cell lines extensively used as models to carry out preclinical and functional work. KNS42 and SF188 were derived from WHO grade IV paediatric GBMs in male patients aged 16 and 8 years old, respectively [21, 22]. UW479 was derived from a grade III anaplastic astrocytoma of a 13 year old female [23]. All three high grade lines have highly complex genomic profiles and harbour amplifications and deletions at several known cancer genes dysregulated in paediatric glioma, including *CCND1*, *MYC*, *CDK4*, *PIK3CA*, *CDKN2A/B*, and *RB1* [24]. SF188 harbours an amplifier phenotype, with coordinated amplification of *CCND1* and *CDK4*. By contrast, KNS42 cells achieves a similar end through direct deletion of *RB1* itself. Further disruption to the p53 pathway (through point mutations of *TP53* in both lines), and the RTK/PI3K/AKT pathway via *NF1* deletion in SF188 and *PIK3CA* copy number gain in KNS42. Interestingly, UW479 cells also harbour numerous high level amplifications, including 3p11-p12, 6p21, 18p11 and 19q12 and an homozygous deletion at the *CDKN2A/B* locus [24].

To examine whether SVs that lead to the production of oncogenic fusion proteins are present in pHGG, we used whole genome and transcriptome paired end sequencing to detect novel gene fusions in pHGG model cell lines – KNS42, SF188 and UW479.

5.2. MATERIAL AND METHODS

5.2.1. Cell lines

Paediatric GBM cell line KNS42 was obtained from the JCRB (Japan Cancer Research Resources) cell bank. Paediatric SF188 cells were provided by Dr Daphne Haas-Kogan and UW479 were kindly provided by Dr Michael Bobola. These cell lines have been previously characterised [24]. They were grown as monolayers in DMEM F12-HAM complete media (Sigma-Aldrich) supplemented with 10% foetal bovine serum (FBS), incubated at 37°C with 5% CO₂. DNA was extracted from all the cell lines, and samples screened for mycoplasma at Surrey Diagnostics (Cranleigh, UK).

5.2.2. DNA and RNA extraction

Genomic DNA was extracted from pHGG cell lines using the DNeasy Blood & Tissue Kit (Qiagen, Manchester, UK), and the quantity and quality of the DNA was assessed with

Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies, Paisley, UK) and ND-1000 Nanodrop (Thermo Scientific, Loughborough, UK). Total RNA was extracted using the Absolutely RNA Miniprep Kit (Agilent, Wokingham, UK), and the quantity and quality of the RNA was assessed on the 2100 Bioanalyzer (Agilent, Wokingham, UK), using the Agilent RNA 6000 Nano Kit (Agilent, Wokingham, UK), according to manufacturer instructions. cDNA was synthesized using the SuperScript II system (Life Technologies, Paisley, UK) according to manufacturer's protocol.

5.2.3. Whole genome sequencing

Sequencing was performed at the Tumour Profiling Unit, Institute of Cancer Research, London, UK. KNS42, SF188 and UW479 were sequenced using the Illumina HiSeq2000 sequencer. Briefly, 2.5-5 micrograms of DNA from each cell line were used to prepare the sequencing library through shearing of the DNA using Covaris S2 followed by ligation of sequencing adaptors according to the Illumina protocol. Fragments sized 00-766bp were extracted from the gel using the MinElute Gel Extraction Kit (Qiagen, Manchester, UK) to prepare 500bp insert libraries. Each sample was sequenced on multiple Illumina flow cells using the TruSeq Paired-End Cluster Kit v3-cBot-HS (Illumina, Little Chesterford, UK) with paired 100-bp reads to achieve ~30X coverage.

5.2.4. Transcriptome sequencing

RNA sequencing was performed at Source Bioscience, Nottingham, UK. 3 micrograms of total RNA from each cell line were prepared for high-throughput sequencing using the Illumina TruSeq RNA sample preparation Kit v2, according to manufacturer's protocol. Sequencing produced roughly 74-92 million paired reads,

5.2.5. Copy number analysis

The copy number analysis uses the cell lines Affymetrix 500K SNP array data previously published by our lab [24]. The copy number data was re-processed using the Aroma Affymetrix package in R based upon the log ratio of each sample against a pool of 30 European Hap Map mothers and smoothed by standard circular binary segmentation in the DNA copy package. Gains and losses were defined with a threshold of ± 0.1 and amplifications and deletions with a threshold of ± 0.5 . Only regions of 10 or more contiguous probes were included in copy number analysis.

5.2.6. Bioinformatic analysis

Raw sequencing data was aligned to the human genome (version hg19, Ensembl56) with the BWA v0.5.10 using default parameters [25]. BAM files were analysed by the BreakDancer v1.1.2 [26] with default parameters to characterize the structural rearrangements present in each line – deletions (DEL), insertions (INS), inversions (INV) and intra-chromosomal (ITX) and inter-chromosomal (CTX) translocations. The filtering process was based on sequence depth and retained rearrangements supported by at least ten paired reads. For ITX rearrangements, insert size should be greater than 1Kb, according to the variability in the insert size produced by the technique. As no matched constitutional DNA is available from the original patient from whom these cell lines were derived, we used publicly available normal genomes, and manually checked all regions involved in order to exclude those regions showing sufficient homology to represent a strong possibility of being false positives due to misalignments.

Fusion gene detection was run on RNA Sequencing data using the Chimerascan pipeline v0.4.5 [27]. This relies upon Bowtie alignments [28] of paired end reads and the identification of mismatched pairs surrounding putative breaks encompassing reads containing the breakpoint sequence. Filtering of the Chimerascan output included steps to annotate each breakpoint for the longest Ensembl transcript covering the region and identifying candidate fused exons either side, as well as the predicted phase of the two partner exons (in-house script), intergenic fusions, readthroughs and those nominating SVs breakpoints smaller than 10kb apart. Junctions were filtered to remove highly promiscuous partner reads with nominations to more than two distinct genomic loci and those representing multi-mapping genomic regions.

Overlay of DNA and RNA sequencing levels fusions was based upon the intersection of gene annotations and genomic loci of the two partner genes in each case.

Sequenced Variants were called using GATKv2.3-9 variant calling. Variants were filtered for depth, allele fraction, and minor allele frequency (MAF) in 1000 genomes of less than 5%. Genes which had more than 100 variants per Mb sequenced and which carry more than two coding variants in each individual cell lines were removed

5.2.7. PCR validation

Four PCR primers (IDT Technologies, Coralville, US) were designed using Primer3 [29, 30] for each pair of genes involved in the predicted SV (Table 5.1). These primers were designed ± 200 bp apart from the breakpoints to generate amplicons spanning the breakpoint-

junction-sequences of predicted SVs. In addition to fusion gene amplification, we amplified both partner genes involved. DNA from the cell line carrying the fusion and normal control DNA (Promega, Southampton, UK) were used. PCR was carried out with Platinum® Taq DNA Polymerase High Fidelity (Life Technologies, Paisley, UK) in a 25 µl volume and with 100 nanograms of genomic DNA or cDNA as template. The following program was used: Initial denaturation at 94°C for 3min, followed by a 2-Step-Touchdown: 1. (94°C 30 sec, 68°C 45 sec, 72°C 1 min) 18 cycles, 2. (94°C 30 sec, 50°C 45 sec, 72°C 1 min) 30 cycles; followed by an additional cycle of 72°C 10 min. Fusions PCR products were extracted by gel purification (Qiagen, Manchester, UK) or cleaned up with the Illustra™ ExoStar™ 1-STEP kit (VWR, Lutterworth, UK). Sanger DNA sequencing of the PCR products was performed by DNA Sequencing & Services (MRCPPU, College of Life Sciences, University of Dundee, Scotland, www.dnaseq.co.uk) using Applied Biosystems Big-Dye Ver 3.1 chemistry on an Applied Biosystems model 3730 automated capillary DNA sequencer.

5.2.8. Metaphase slides

KNS42, SF188 and UW479 were grown to 70-80% confluency and incubated with colcemid (Life Technologies, Paisley, UK) (1:100) for 1-3h. pHGG cells were harvested and treated with an hypotonic solution of 0.075M KCl for 5-8min at 37°C. After this treatment, cells were pre-fixed with an ice-cold 3:1 methanol:acetic acid solution. The cells were collected by centrifugation, the supernatant was removed completely and the cell pellet was resuspended in 1 ml of fresh fixative (3:1 methanol:acetic acid solution) and stored at -20C until use. The nuclear suspensions were dropped onto clean slides and dried at room temperature prior to use. Control metaphase slides were kindly provided by John Swansbury (Clinical Cytogenetics, Royal Marsden Hospital, Sutton, UK).

5.2.9. Fluorescent *in-situ* Hybridization

Fusion genes were either detected using break-apart FISH probes for one of the genes involved in the fusion or using fusion probes, where a probe for each gene involved was selected and differentially labelled. Centromere probes were used as chromosome-specific controls. BAC clones were obtained from 32K re-array BAC collection distributed by CHORI and housed by the Breakthrough Microarray Facility (Institute of Cancer Research, London, UK) and FISH-mapped onto control metaphase slides to ensure specificity. BAC clones used are listed in Table 5.2. BAC DNA was amplified from 10 nanograms starting material using the Illustra™ GenomiPhi™ V2 DNA amplification kit (GE Healthcare, Little Chalfont, UK) according to manufacturer's instructions. Probes were labelled with biotin, dioxigenin (DIG) or Aqua using the BioPrime® DNA Labeling System (Life Technologies, Paisley, UK).

Table 5.1. Sequence of primers used for the screen of structural variants involving genes at either side in KNS42, SF188 and UW479 cell lines. The first column corresponds to the fusion ID which contains the cell line in which the fusion is present, followed by the chromosomal location of the partner gene 1 and partner gene 2.

Cell Line Fusion ID	Gene 1	Gene 1 Primer Forward	Gene 1 Primer Reverse	Gene 2	Gene 2 Primer Forward	Gene 2 Primer Reverse
KNS42_1_156051752_3_189542047	MEX3	CTTTGTAGAGAGCGCAGAGCTT	AGATCTGCTTTCTTGGTTTTGC	TP63	CGGATAGAGGCAGCAGGTAT	TTGAAAAAGTCCAGGCTAGG
KNS42_2_171814878_13_49848076	GORASP2	TCTCATCCAGCTCTTTAGTCC	CACAGCTACTTGGTTAGGCTGA	CDADC1	CTTGCATTCTTTTTCCCTCAAG	GATTTCCCTTCTGCCCAAAT
KNS42_1_58802590_10_74176383	DAB1	AGCTGCAAGACTTCAATCCTTT	AACCACCACGTTACTCCAATTT	MICU1	TTATGGCCACATTTACACAG	TGGGAATGCAGTAGTGAGCA
KNS42_2_43437013_6_45891802		CTCCCATGATGGTTAGAACACC	CCAGCTGTCTGTCTCTCTCA	CLIC5	CATCGGATCGTGAAGTATT	AGGGTATCCAGGGAAGTCT
KNS42_6_14261299_8_8718282		CTGAAGCTGACAAACACAAAGC	AACCTCCACTATGATTCTGGA	MFHAS1	AGAACGCCCTCTGTTCTTT	TTACCTCAGCCCAAGAAACG
KNS42_6_15532273_8_8667072	DTNBP1	ATCTCAGTCACTGCAACCTCT	ATTCCACCAGTTTTCTAGCTG	MFHAS1	GCTCCCACTGATTCCACATT	GCCGTCTGCAGTAACCTCTC
KNS42_1_152387895_11_132713260		CACACATGTTTTGTGGTGATG	GTGATGTTTCATAAGGTGCCAAA	OPCML	CATCAGCCTAAGATGCTTGC	GCTCTGTTTCAGATGCCAAG
KNS42_12_96967216_X_100735303	c12orf63	CGAGGAAGGATATGTCTCCAAA	CTAGAGCCTGCATACCAGTCC		TGCCCAAGGATATCCAAAA	AGACAGGGTTTTCGCTGTGTC
KNS42_7_103159742_12_98598388	RELN	GTGTTTCTTTGCTTTCCAGTCC	AGCCTAGGTGACAGAGCAAGAC		CACAGACTGGCAAAATTGGAT	TTCAGGAGCAGGTTGTTCCG
KNS42_3_164722685_5_57093054	SI	CTGCTGAATATGCATCCAAAAG	ATTGACTCACAGTTCTGCATGG		CCCATCTTGTGGCCATTGT	TAACAGCAGGCTTCCAAAA
KNS42_12_96977487_X_98569531	c12orf63	TGATCTTGACCTTGAGACTGA	GTGAACTTCATTTCCCTTCAC		CAAGCCAGTCAAAGCAGACA	TCTGATTCTTTGCCATTGGTC
KNS42_5_64964650_5_65433006	SGTB	ACTTGAGGTCAGGAGTTCCGAGA	GGAGAAAAGGGGCATTAGGTAT		TTCCACATTTTTGGGCATCT	CCACCAGGATTGTGATAGGG
SF188_20_4895215_20_25459154	SLC23A2	ATCCCTGAGAGGCCCTAAGGA	CTGCCTCAGCCTCCTGAGTA		GCCTCCCAAAGTGCTGATTA	TCAAGTCTGGGTTTTCTTGC
SF188_6_158783682_17_78736108	TULP4	TGCGCTTGTGTTTGTTCAT	CAATCTTGGCTCACTGCAAAA	RPTOR	CAATCTTGGCTCACTGCAAAA	AGGGCGGTAGGACAAACTCT
SF188_1_144680231_1_144954567		CAGGATTCAAATGCCACAGA	TCCCCTCTCACCTTAAGCAA	PDE4DIP	CGTGTAGCCAGGATGGTTT	GGGAAAATGGGCACTCTCAT
SF188_11_88909186_15_27561047		TCCGGCATCATCGAAATAGT	TGAGGCCTCCACATCATA	GABRG3	GGCAAATTGGACCCTGAATA	CTCAGCCTCCCAAGTAGCTG
SF188_20_278393_21_17652744		AAAAACCAAACCCACACAGC	ATCTGGATGACCGCTAGAT	c21orf34	TTGTGGAGAGGAATGGGGTA	GATGTTGTTCTGGCCTCTCC
SF188_16_542305_16_542428	RAB11FIP3	GGGTTTGTAAAGGTGGCAGA	AAGTCTCTGCGTGGGAAGAA		GGGTTTGTAAAGGTGGCAGA	AAGTCTCTGCGTGGGAAGAA
SF188_11_57812425_20_3753964	OR9Q1	CGACATCTGTGCAGTCAGGT	TCCCTTATCCGAATTGCTTG		TCAAAGACCACCAGGCAAT	TTGCCTGTAACCCAGCTAC
UW479_8_137751825_17_49361503		CTGCAACATGCAATTCATCA	CTTTCCCCTGCAGAGACTTG	UTP18	CACGCCCAGCTAATTTTTGT	GATCCAGAACACGGGAATTT
UW479_8_117230013_17_49354767		AGCCATGTGGGACTGTGAGT	TACCTTTCCCCTCACAAATGG	UTP18	GGTTGATGGGAAAACAAATCC	TGGGTATGGTAGAGGGCAGA
UW479_8_141832537_15_42650218	PTK2	CATGCAGATTGGACATGGAG	ATGCTCCCTTATCACCAACG	CAPN3	TTTGCGTCAGCGAGTAAATG	GGGATGTCACACCTGGTTTC
UW479_3_47028066_22_19936322	NBEAL2	AGGGTGTGGTTCTAGGGACA	GGAGCCCAAAATACCCAGAT		AAAATTAGCTGGGCATGGTG	TCTCACCACTGCACTCCATC
UW479_3_55024288_19_47734383	CACNA2D3	GGATCAGAAGTCATGCACCA	AGAATGGTCAGGGGAGAGGT		CTGCACTCCTGTCACCTCCT	GTCGGTCCGTCTGTGTACG
UW479_9_32381210_9_33849115		GCGTATAGCAGGTGTGGTGA	TGATGTGGGTGGTCAAAGA	UBE2R2	CATGCATGAGCTACCACACC	ACAGGCGTCCATCCATACTC
UW479_1_113149456_20_3902968	ST7L	TACAGTGGCACCCCACTTCT	GCCCCATATCTAGGCCCTTA		CGAAGTGCTGGTTACCTGT	AAGAGGTGTGGAGAGCAAGG
UW479_14_32066014_14_33251221	NUBPL	TCTGACGCTGAGATGACTGG	GGACGGAGGAAGAACAGGAT	AKAP6	GGCAAGCATTACCCAGATGT	AGCTGTGGATCCCCATTATT
UW479_17_8745523_17_9557484	PIK3R6	AGTTGCTGTGCTGTTGCAGT	TTGGGAAAACACTTGTGTGA		GCTGGGTCCCTCCTTAGGTA	CTCAGCCTCCCAAGTAGCTG
UW479_9_27945060_18_30711907		TTCAGATGCCACAGAATCCA	TTCCACTCTGGTTTTGCTCA	c18orf34	CCCACCAACATTATGCCATT	CATTTCAAAGAGGATATGCAGAGA
UW479_9_102146807_18_30736184		AACGAGCAGGAGACAATGCT	ATGAGTCCCAGGGAAAATCC	c18orf35	ACTGGGGGTATTTTCATGGTG	CCCAGAAGTGTGCATGCAGAA

UW479_9_86638539_18_30729821		GGACCAGATGACCTCCAAGA	AGACCAGTGGTCCCAAGAA	c18or36	GAAGCCTTTCCGTATCCACA	CTACATGTGGCCTCCCAGTT
UW479_1_82756062_14_67423798		AGGACACGACATTGGGTTTC	TGAGCAGCTATTGCAGAGGA	GPHN	TTCTCTGATGGCCAGTGATG	GGGAGGCTGAGGTAGGAGAA
UW479_14_31036241_15_30984655	G2E3	TTGGTGCCAATTTTTCTTCA	GAAATGAAAGTGGGACACTGC		CACCACCACACTTGGCTAAT	AGCCCCAGAAACAGAAGGAC
UW479_12_1695918_12_9246077	A2M	TGAGCTCTGATCATGCCACT	CTCCCAAAGTGCTGGGATTA		TAGAAAGGTGGTTGGGGTTG	GCCTCTCCTGGCTTGTTAAG
UW479_8_47159193_9_137280905	RXRA	CCATGCAGAGCAAAGAAACA	TCATGCCGCTTCTTCTACT		CTTCGATGTCCTTCTCTGC	ATACCCCCAAAGTCCAAAGC
UW479_12_3690145_20_62854780	PRMT8	GAAGGCCATGCTACACTGGT	GTCCTGGCATCTCCAGTCAT	MYT1	GCTGTCTCCAAGGACTGAG	GATCTTCAGGAGAGGCTGGA
UW479_3_114154875_3_168102681	ZBTB20	CAACCTGAAGTTTGGCTTGG	GGTTTTCTTCTGGGTCACCA		ATGGCATCTAAAGCCAGCAC	GTTTCACGCAGACAGCAAAA
UW479_20_15544438_22_35396912	MACROD2	CCTGGTCCATGCCCTTAATA	GGCGACATGAGAAACCATCT		TGCAATCTCAGCCTACATCG	GCCACATGGTCTCTGTAGCA
UW479_5_93556435_22_35366567	c5orf36	ATGACTGGCAGCACACATTT	GCTCTACCCCTGAGACCACA		CCAGAACCTCAGAATGTGACC	CCTGGGGCTGCTATAACAAA
UW479_1_47836334_14_53815391	CMPK1	TAACCCACGTCCTCTCATC	GGAGTACAATGGTGCGATCC		CTGGGTGCTTCTTCCCTTG	TGCAGGACACTTAAGGAGCA
UW479_6_45229405_8_62847098	SUPT3H	CAAAACCTGCCAGAGACACA	TCAGCTTGGGCATTATTGGT		TTCAAGGTCACTCCAAAGCA	TAGGTCATGGCTGCCAAACT
UW479_8_62579337_8_62708134	ASPH	GCAATGCCACATTTAAGCAG	GGGCTTGGCATATTGTTTGT		TGCAAGTTGAGAGGATTCTTTG	GGCAGAGATTGGAAGATTGAA
UW479_8_57896780_19_31716239	IMPAD1	ATTTTGGGAAGAGGGGAAAA	GGTGAGAGATGGTGGCTAA		AAATCAGAAAAGGGGGCATT	ATAAGGCAACAGCAGGAGGA
UW479_7_71425213_7_99282578	CALN1	AATTCCTGGGCAAACCAAAT	AACACGGATGGAAGTGAAG		GCAATGGGCAAAGTCACAGT	TCAACCAGAAAAACCCATTG
UW479_8_40433239_14_55716413	ZMAT4	CCACCCCAAGAAGATACAGC	CATTGCCAAGATTTGTTTGG		GGAGACAAAAGGCCATCAAA	CACATGTTTCCTGCATGTGG
UW479_9_28000942_9_86632969	LINGO2	TCTCTGCATTCTGACACCA	CTTCTGTGGCATGTGGAAGA		CTTTTGCCCGGCTAATTTTT	GGGCAACAGAGTGAGACTCC
UW479_7_88827600_8_143781332	ZNF804B	TCAATTGAGTAGCAAATTGGATG	CAGGTAACCTCAAGGTCACATTATT		GCATGGAAATCCTCCAGTTC	ATCTTCTCCGCTGAAGAGCA

Briefly, the metaphase slides were incubated in 70% acetic acid for 10min and washed in PBS 3x3min and 2XSSC 1x3min, prior to digestion for 5min at 37°C with 0.01M HCl / 0.02mg/ml Pepsin solution (Sigma, Gillingham, UK). The slides were washed in PBS 2x3min and PBS / 0.05M MgCl₂ 1x3min followed by fixation in 1%formaldehyde / PBS / 0.05M MgCl₂ for 5min. Dehydration was carried out with 3 min 70%, 90%, 100% ethanol and slides were air dried for 10min at room temperature. The slides were denatured at 75°C and hybridised with the appropriate biotin and/or DIG and/or Aqua labelled probe overnight at 37°C in a humidified chamber. Removal of the coverslips was carried out by soaking the slides in 2xSSC / 0.1% Igepal for 2 mins at room temperature and immersing in 0.4xSSC / 0.3% Igepal at 73°C for 2min. Slides were rinsed in 2xSSC / 0.1% Igepal for 1min and PBS for 3min. When dual FISH was performed, slides were first incubated in a humidified chamber at 37°C with two layers of anti-DIG-fluorescein (Roche, Welwyn Garden City, UK) for 10min each, washed with PBS 3x3min and incubated with two layers of Streptavidin-cy3 (Sigma, Gillingham, UK) 10min each. Slides were soaked in PBS for 3min, dehydrated with 3min 70%, 90%, 100% ethanol and left to dry at room temperature for 10min. Slides were finally mounted in Vectashield with DAPI (Vector Laboratories, Peterborough, UK), and captured on the Leica Ariol microscope (Wetzlar, Germany) at x20 using filters for DAPI, Cy3, FITC and Aqua.

Table 5.2. List of BAC clones of the 32K re-array BAC collection used for FISH

Probe ID	chrom	start	end	bac.id	Gene/chr/cent
TP1708E04	17	78045988	78182971	RP11-334C17	<i>RPTOR</i>
TP1709F04	17	78559488	78720126	RP11-317F05	<i>RPTOR</i>
TP1707A07	17	78698182	78900648	RP11-812H19	<i>RPTOR</i>
TP0613A07	6	158899201	159089502	RP11-654E18	<i>TULP4</i>
TP0618H04	6	159060790	159272494	RP11-507C10	<i>TULP4</i>
TP0619A11	6	159308756	159421507	RP11-589H21	<i>TULP4</i>
TP0206C02	2	171595376	171751356	RP11-164H08	<i>GORASP2</i>
TP0206A10	2	88991806	89157237	RP11-157D12	cent 2(p)
TP0222G01	2	89159257	89307765	RP11-294I20	cent 2(p)
TP0225B01	2	89264572	89459196	RP11-631H05	cent 2(p)
TP0218B05	2	89389635	89602031	RP11-685C07	cent 2(p)
TP0228B08	2	89526216	89623186	CTD-2063D22	cent 2(p)
TP1308G12	13	49942118	50119412	RP11-767O08	<i>CDADC1</i>
TP1305D11	13	49572364	49735007	RP11-446P06	<i>CDADC1</i>
TP1306C04	13	19398031	19540646	RP11-518D18	cent 13 (q)
TP1306C02	13	19501986	19679218	RP11-521I05	cent 13 (q)
TP1302H02	13	19697044	19857280	RP11-139J24	cent 13 (q)
TP1304B12	13	19845749	20046229	RP11-304L17	cent 13 (q)
TP1306C09	13	19909447	20076150	RP11-535A03	cent 13 (q)
TP1402H07	14	31636363	31795856	RP11-200I02	<i>NUBPL</i>
TP1402F05	14	32290056	32435545	RP11-159D23	<i>NUBPL</i>
TP1103C02	11	55707995	55878656	RP11-129H09	cent 11q
TP1103D04	11	55875895	56040558	RP11-138A07	cent 11q
TP1202E10	12	33739612	33925027	RP11-102G23	cent 12p
TP0806G06	8	47450712	47566922	RP11-445A15	cent 8
TP0403E10	4	54072620	54226790	RP11-98L24	<i>SCFD2</i> (4q2)
TP0416F09	4	54252656	54420846	RP11-809N18	<i>LNX1, FIP1L1</i> (4q12)
TP0402E10	4	54576785	54759122	RP11-61P24	<i>FIP1L1</i> (4q12)
TP0411H12	4	55093276	55273379	RP11-626H04	<i>PDGFRA</i> (4q12)
TP0813F05	8	128690817	128824097	RP11-237F24	<i>MYC</i> (8q24)
TP1112D12	11	48449190	48603500	RP11-657L09	11p11.2
TP1109F12	11	69409496	69572287	RP11-681H17	<i>CCND1</i> (11q13.2)
TP1108E01	11	118313325	118461726	RP11-606K01	<i>MLL</i> (11q23.3)
TP1215E03	12	58002869	58210408	RP11-571M06	<i>CDK4</i> (12q14)

Cent: centromere; chr: chromosome

5.3. RESULTS

To identify potential novel structural variants in pHGG, we performed whole genome sequencing of three cell line models - KNS42, SF188 and UW479, generating paired 100-base reads with an average of 30X sequence coverage. The Breakdancer package was then applied to detect SVs [26], as it is considered to provide accurate and comprehensive predictions of SVs in genomes, and is routinely used to identify fusion genes [31]. It predicts individual genomic breakpoints by searching from clusters of abnormally mapped paired end reads [26]. We used a stringent filtering method that only identified rearrangements supported by at least ten reads, with a Breakdancer score of ≥ 80 , with an even balance between positive and negative strands, and which was filtered against normal genomes. This approach identified 2646 genomic junctions (INS, DEL, INV, ITX, CTX) of which 303 represented genic rearrangements: 76 in KNS42, 93 in SF188 and 134 in UW479 (Appendix V). These events were divided into two groups: CTX (n=216) and ITX (n=87). The genomes of the cell lines are represented as Circos plots in Figure 5.1, also annotated for known cancer-associated single nucleotide variants.

As fusion genes are often associated with specific copy number breakpoints [32, 33], we interpreted our SV analysis in the context of a previously published 500K SNP array study carried out on the cell lines [24] (Appendix VI). 74/303 (24.4%) of the nominated SVs were associated with an intragenic copy number difference in at least one of the two breakpoints. SF188 showed the highest number of copy number-related breakpoints (n=34) due to the occurrence of a complex rearrangement that involves regions of chromosomes 4, 8, 11 and 12 (see below). UW479 was also found to be highly rearranged (n=28), more so than KNS42 (12 breakpoints associated with copy number).

The predicted SVs were validated by PCR using genomic DNA, with 17/74 (23%) confirmed (though additional 27/74 (36.4%) correspond to the SF188 extra-chromosomal rearrangements which were subsequently validated by FISH) (Table 5.3). Sanger sequencing of the detected SVs revealed sequence microhomology at the rearrangement breakpoints in 13/17 (76%) validated SVs (ranging from 1-7 nucleotides). Two SVs in UW479 (*PRMT8:MYT1*, *ASPH:chr8*) were separated by small genomic shards, 1 and 6 nucleotides, respectively (Figure 5.2). A schematic representation of the predicted structures of each fusion validated is presented in Figure 5.3.

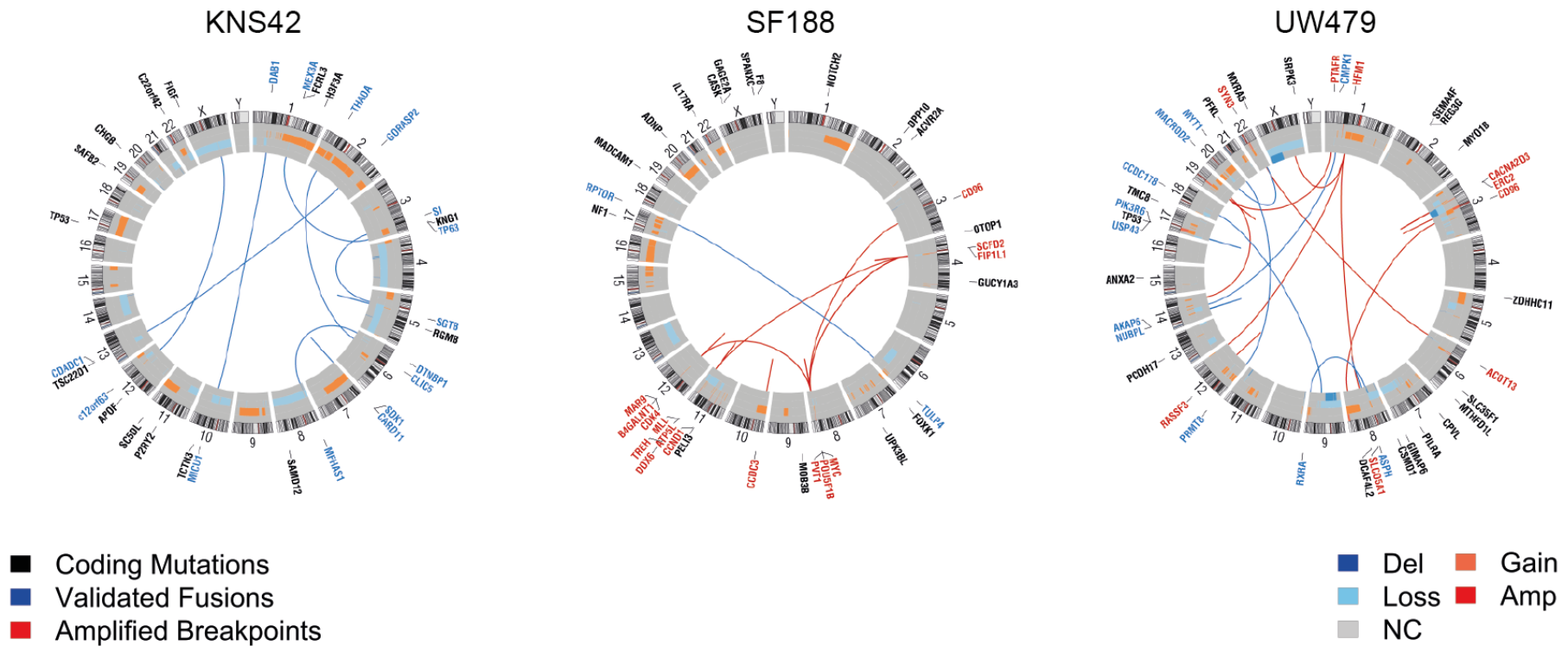


Figure 5.1. DNA structural rearrangements and copy number alterations detected in three paediatric high grade glioma cancer cell lines – KNS42, SF188 and UW479 – displayed as Circos plots. Each Circos plot depicts whole genome sequencing data from each cell line. Outer ring contains chromosomal ideograms, annotated for somatic single nucleotide variants in coding genes (black). Inner ring plots copy number derived from copy number data, dark orange=amplification, orange=gain, dark blue=deletion, light blue=loss. Inside the circle are drawn structural variants, blue=validated fusion genes, red=amplified breakpoints.

Table 5.3. List of fusion genes validated at the genomic level in each cell line.

Cell line Fusion ID	Gene 1	chr 1	Gene 2	chr 2
UW479_14_32066014_14_33251221	<i>NUBPL</i>	14q12	<i>AKAP6</i>	14q12
UW479_17_8745523_17_9557484	<i>PIK3R6</i>	17p13.1	<i>USP43</i>	17p13.1
UW479_12_3690145_20_62854780	<i>PRMT8</i>	12p13.32	<i>MYT1</i>	20q13.33
UW479_20_15544438_22_35396912	<i>MACROD2</i>	20p12.1		22q12.3
UW479_1_47836334_14_53815391	<i>CMPK1</i>	1p33		14q22.1
UW479_8_62579337_8_62708134	<i>ASPH</i>	8q12.3		8q12.3
UW479_9_86638539_18_30729821		9q21.32	<i>CCDC178</i>	18q12.1
UW479_8_47159193_9_137280905		8q11.1	<i>RXRA</i>	9q34.2
KNS42_2_171814878_13_49848076	<i>GORASP2</i>	2q31.1	<i>CDADC1</i>	13q14.2
KNS42_1_58802590_10_74176383	<i>DAB1</i>	1p32.2	<i>MICU1</i>	10q22.1
KNS42_7_3023122_7_3359981	<i>CARD11</i>	7p22.2	<i>SDK1</i>	7p22.2
KNS42_6_15532273_8_8667072	<i>DTNBP1</i>	6p22.3	<i>MFHAS1</i>	8p23.1
KNS42_5_64964650_5_65433006	<i>SGTB</i>	5q12.3		5q12.3
KNS42_12_96977487_X_98569531	<i>c12orf63</i>	12q23.1		Xq22.1
KNS42_3_164722685_5_57093054	<i>SI</i>	3q26.1		5q11.2
KNS42_2_43437013_6_45891802		2p21	<i>CLIC5</i>	6p21.1
SF188_6_158783682_17_78736108	<i>TULP4</i>	6q25.3	<i>RPTOR</i>	17q25.3

To determine whether the validated SVs resulted in expressed fusion transcripts at the mRNA level, we additionally performed paired-end whole transcriptome sequencing of the three cell lines and applied the Chimera Scan software package to identify chimeric gene sequences [27]. The predicted list of fusion transcripts was based on the overlay of DNA and RNA sequencing levels fusions through the intersection of gene annotations and genomic loci of the two partner genes in each case and it is represented as circos plots (Figure 5.4).

Of the validated expressed fusions, we chose to focus on three for further study: *RPTOR:TULP4* (SF188, described in detail in Chapter 6), *NUBPL:AKAP6* (UW479), and *GORASP2:CDADC1* (KNS42).

The *NUBPL:AKAP6* fusion transcript is derived from an intra-chromosomal rearrangement at 14q12. The sequences of the DNA junction exhibited microhomology of two nucleotides, a characteristic of the non-homologous end joining (NHEJ) mechanism (Figure 5.2). Both Nucleotide Binding Protein-Like (*NUBPL*) and A Kinase Anchor-Protein 6 (*AKAP6*) are transcribed in the same direction, with the DNA copy number profile demonstrating amplification of regions preserving the 5' end of *AKAP6* and the 3' end portion of *NUBPL* (Figure 5.5A), similar to the mechanism driving *KIAA1549:BRAF*, a fusion that results from a tandem duplication of 2Mb at 7q34 in LGG [34].

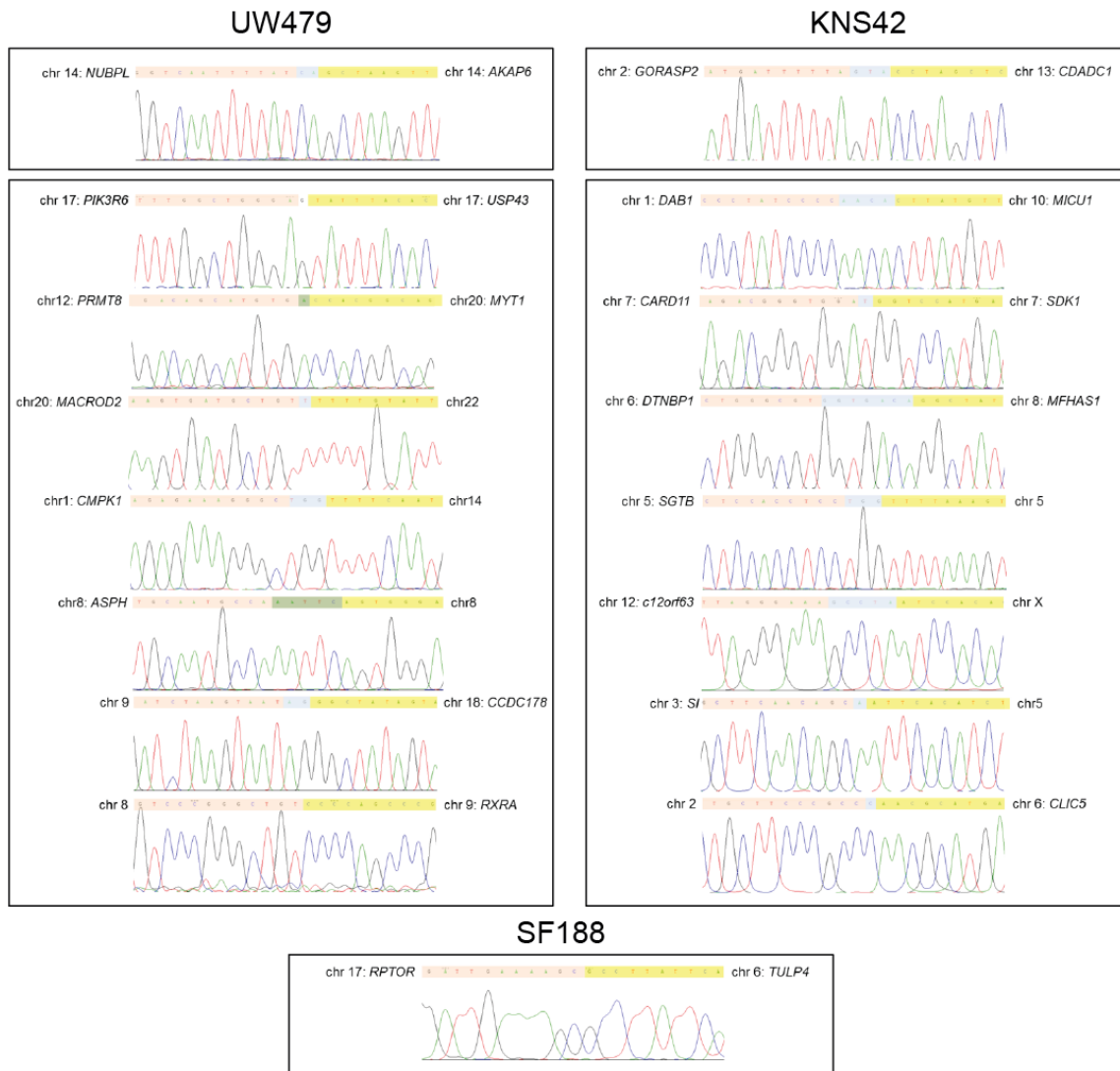


Figure 5.2. Sanger sequencing validation of novel fusion genes in the paediatric high grade glioma cell lines – UW479, KNS42 and SF188. Sequence traces of eight novel fusion genes (two inter-chromosomal and six intra-chromosomal) in UW479, eight inter-chromosomal fusions in KNS42 and one inter-chromosomal translocation in SF188.

In order to determine whether the *NUBPL:AKAP6* rearrangement was the result of a tandem duplication, we designed a break-apart FISH probe for *NUBPL* (Figure 5.6A), however we were unable to detect this using FISH, presumably due to the low resolution of this technique for such a small region (1Mb) (Figure 5.6C). Regardless, the pair of signals of the split *NUBPL* probe are separated at a relatively constant distance from each other Figure 5.6C (yellow arrow). Sanger sequencing of the RT-PCR product showed fusion of *AKAP6* exon 12 with exon 4 of *NUBPL* (Figure 5.5B).

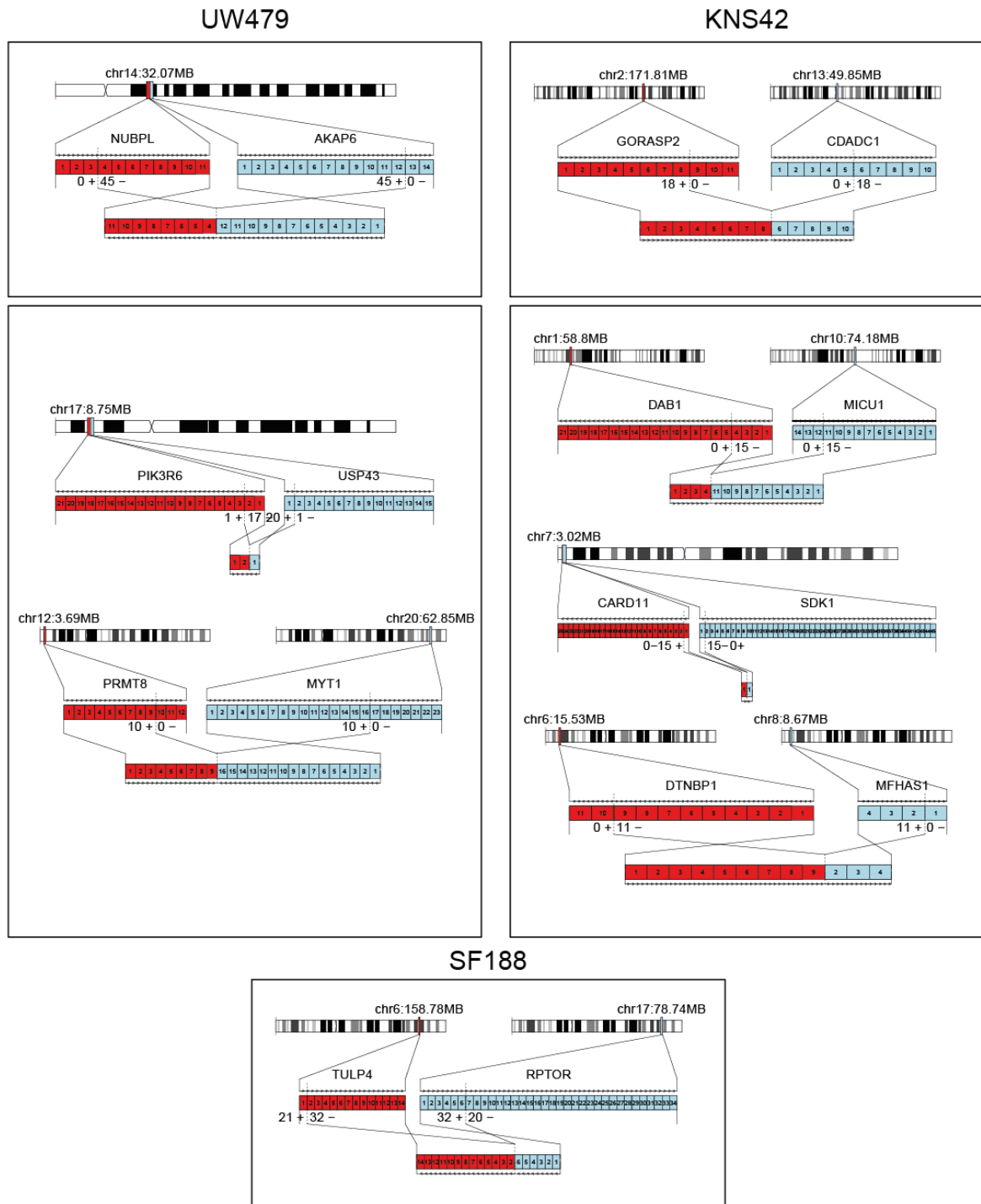


Figure 5.3. Schematic representation of each gene involved in fusion formation as well as the formation of the predicted fusion transcript for each validated (DNA level) translocation in the paediatric high grade glioma cell lines.

The fusion gene identified in KNS42 involved the genes encoding the Golgi reassembly stacking protein 2 (*GORASP2*) and the cytidine and dCMP deaminase domain containing 1 (*CDADC1*) (Figure 5.7C). This rearrangement resulted from an inter-chromosomal translocation between 2q31 and 13q14, with the genes involved being transcribed in the same direction. Sequencing of the genomic breakpoint revealed a similar microhomology region to that described above with *NUBPL:AKAP6* (Figure 5.2). We used two different FISH strategies to detect this rearrangement: a fusion probe for the *GORASP2* and *CDADC1* regions involved in the fusion and a *CDADC1* break-apart probe to determine if the gene was disrupted (Figure 5.8A). FISH confirmed that *CDADC1* and *GORASP2* are fused in KNS42 (Figure 5.8B). The *CDADC1* split apart probe FISH indicated that part of this gene is deleted (5' end), since we can only detect two copies of the normal gene and one red signal for the 3' end region of *CDADC1*, the part involved in the fusion (Figure 5.8C). Additional sequencing of the cDNA revealed the fusion junction to involve exon 7 of *GORASP2* and exon 6 of *CDADC1* exon, joining the 5' end region of *GORASP2* with the 3' end of *CDADC1* (Figure 5.7B). Expression of the fusion gene was confirmed by assessment of exon coverage from the RNA sequencing data, demonstrating the predicted exon-specific expression pattern, restricted to KNS42 (Figure 5.7D).

Paired end reads from the whole genome sequencing analysis suggested a complex rearrangement involving different chromosomes in the pHGG cell line SF188. Intra- and inter-genic breakpoints associated with highly amplified regions in chromosomes 4q12, 8q24, 11p11, 11q13, 11p14, 11q23 and 12q14 (Figure 5.9) were detected by the Breakdancer package. The amplicon structures showed alternation of amplified and non-amplified segments, some with similar copy number, suggesting a bridging mechanism that connects the segments and enables co-amplification. Several oncogenes previously noted to be altered in GBMs were located in these amplicons: *MYC* (8q24), *CCND1* (11q13) and *CDK4* (12q14). Gene amplification at 4q12 has been previously reported in GBM cases and this region spans three neighbouring RTKs - *PDGFRA*, *KIT* and *KDR*. FISH on SF188 revealed that *PDGFRA* was not amplified in SF188 and it is not involved in this complex structure (Figure 5.10A). Instead, the two focal amplifications at 4q12 are located at the *SDCD2* (Figure 5.10A) and *FIP1L1* (Figure 5.10B). loci, upstream of the 4q12 RTKs, and it is these genes which are found in the amplified structure (Figure 5.11). Additionally, the amplicon at 11q23 involves *MLL*, a histone 3 lysine 4 (H3K4) methylase frequently involved in translocation in human acute leukaemias of myeloid and lymphoid lineages [35]. The amplified regions at 11p11 and 11p14 do not include any coding genes.

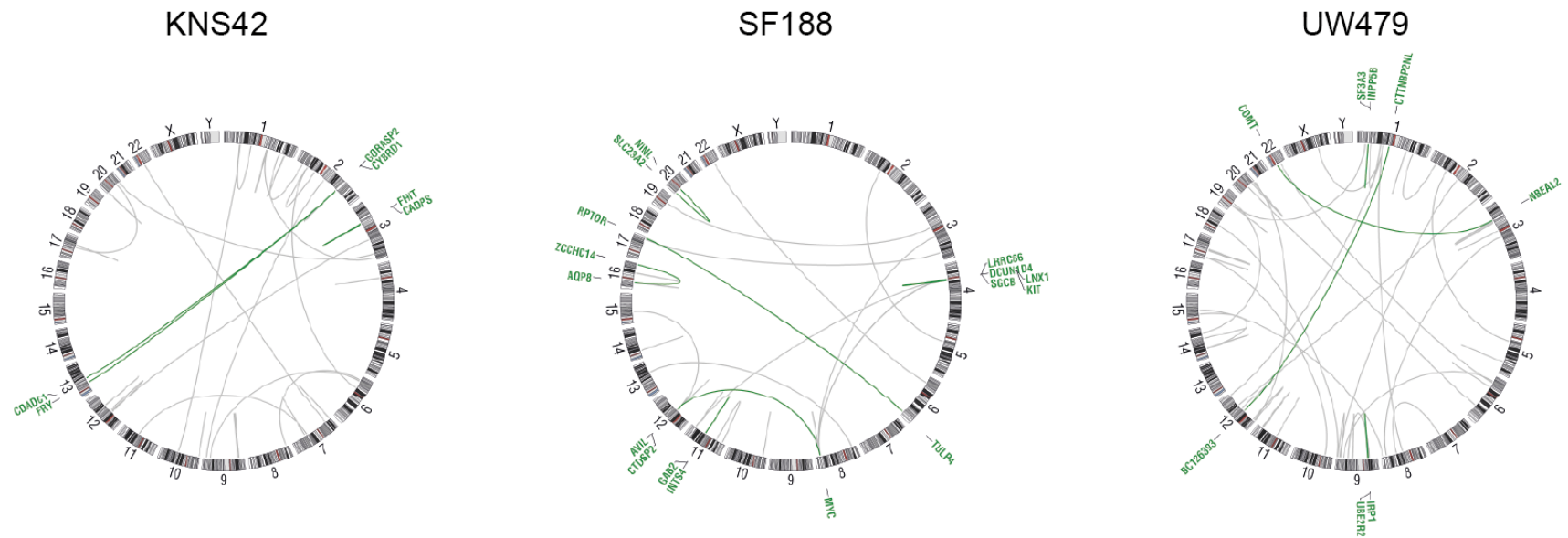
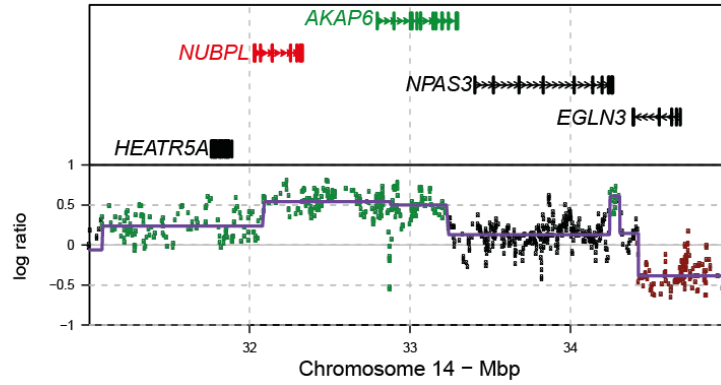
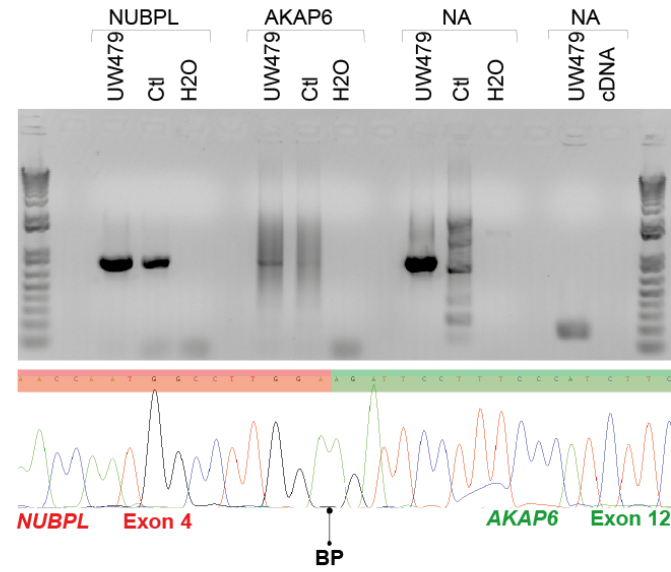


Figure 5.4. RNA structural rearrangements detected in three paediatric high grade glioma cancer cell lines – KNS42, SF188 and UW479 – displayed as Circos plots. Each Circos plot depicts an overlay of transcriptome sequencing with whole genome sequencing for each cell line. Green: expressed fusion genes; Grey: structural variants only detected in RNA sequencing.

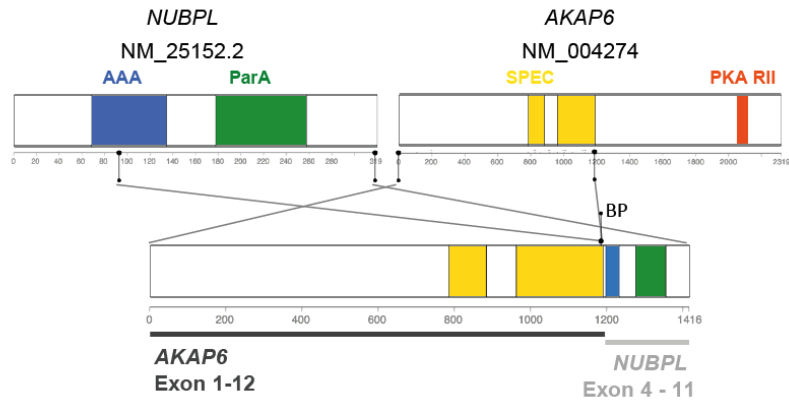
A



B



C



D

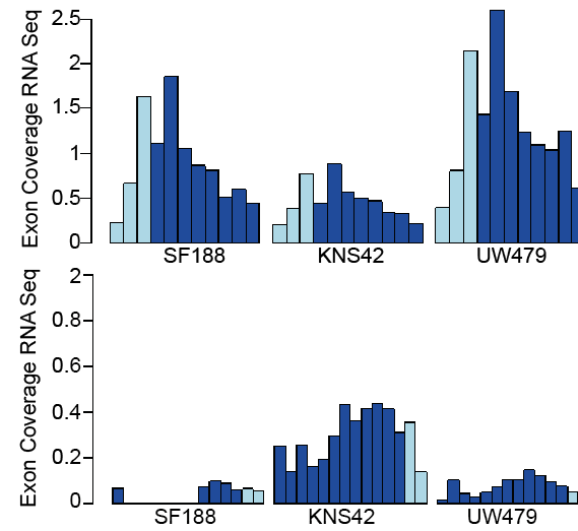


Figure 5.5. Identification of *NUBPL:AKAP6* gene fusion in UW479 paediatric high grade glioma cell line. A) UW479 500K Affymetrix aCGH copy number profile of *NUBPL* and *AKAP6*. B) PCR and RT-PCR detection of *NUBPL:AKAP6* in UW479 genomic DNA and cDNA. Ctl: normal human genomic DNA. Partial sequence of the *NUBPL:AKAP6* fusion transcript identified in UW479. C) Schematic representation of *NUBPL* and *AKAP6* genes as well as the formation of the fusion transcript involving exons 1-12 of *NUBPL* and exons 4-11 of *AKAP6*. D) Expression of *NUBPL* and *AKAP6* exons in UW479 vs KNS42 and SF188 cell lines. NUBPL: Nucleotide Binding Protein-Like; AKAP6: A Kinase Anchor-Protein 6; NA: NUBPL:AKAP6; AAA: AAA ATPases domain used for ATP binding; ParA: ParA/MinD ATPase like domain; SPECT: spectrin repeats; PKA RII: ser/thr protein kinase A RII interacting domain

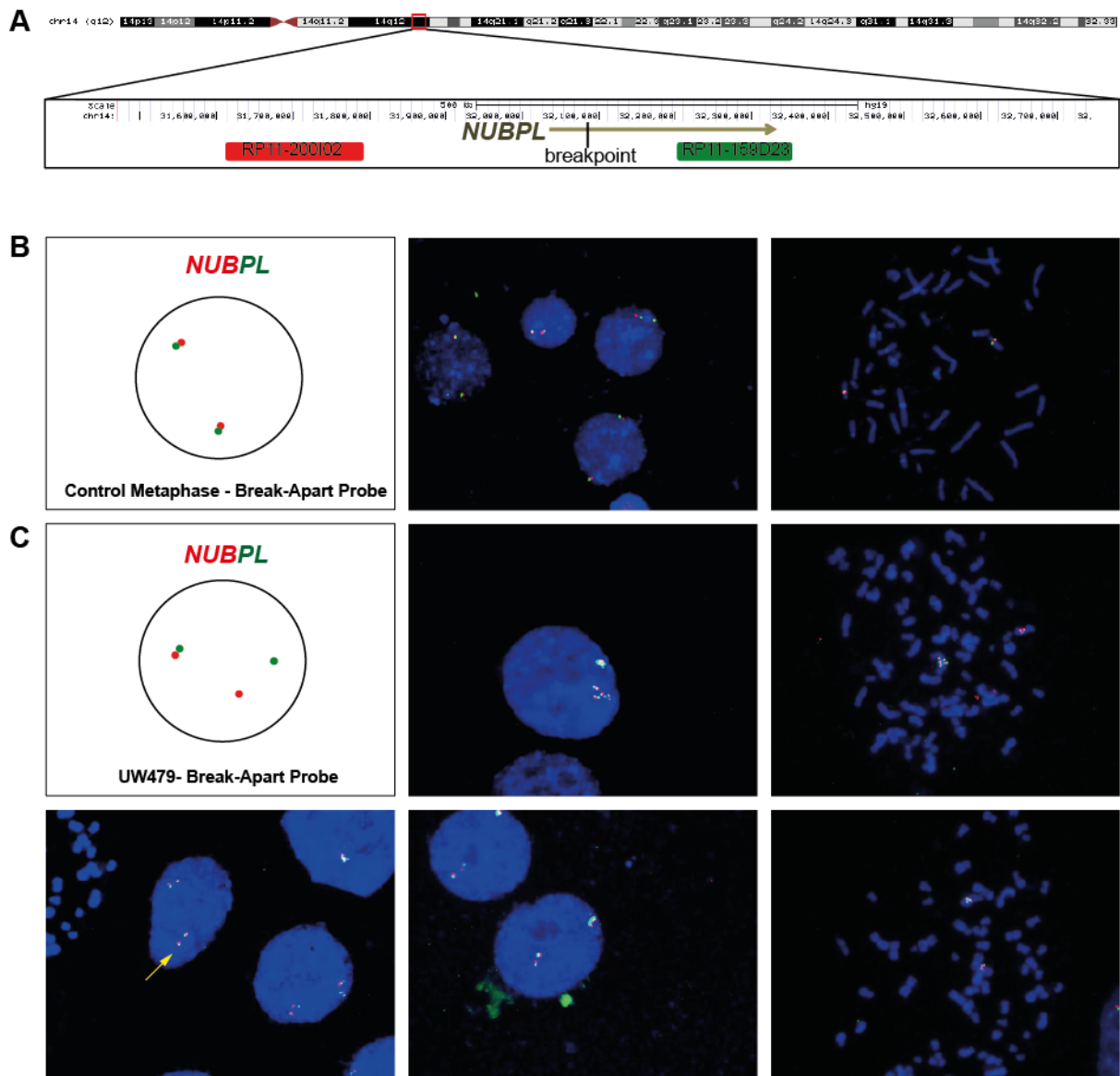


Figure 5.6. FISH of *NUBPL*:*AKAP6* rearrangement in UW479. A) Region specific BAC probes used for break-apart FISH. Probes were selected according to hg19 build of the human genome to overlap with the region of *NUBPL* that is involved in the fusion (RP11-200I02; green bar) with *AKAP6* and to the region of the same gene that it is not disrupted (RP11-159D23; red bar). B) Diagram of the FISH signals expected when using control tissue, followed by interphase and metaphase *NUBPL* FISH in control tissue. C) Representative cartoon of the expected signals in the positive fusion cell line followed by FISH on interphase and metaphase UW479 cells. Original magnification 1000X.

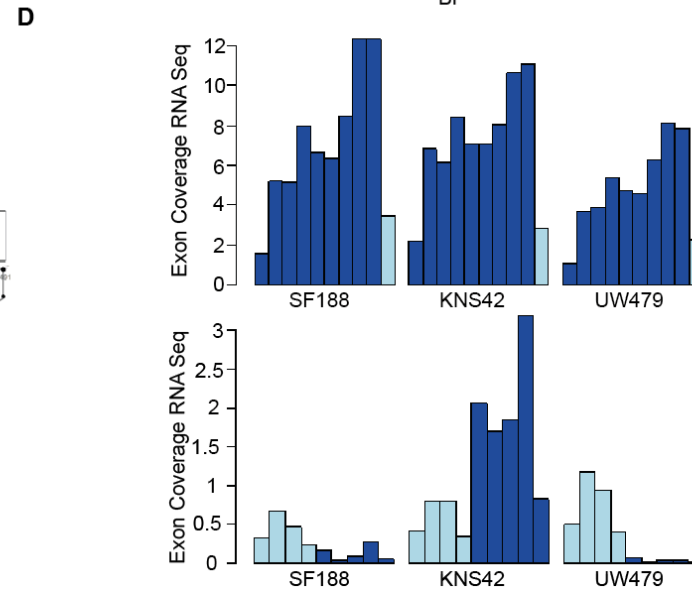
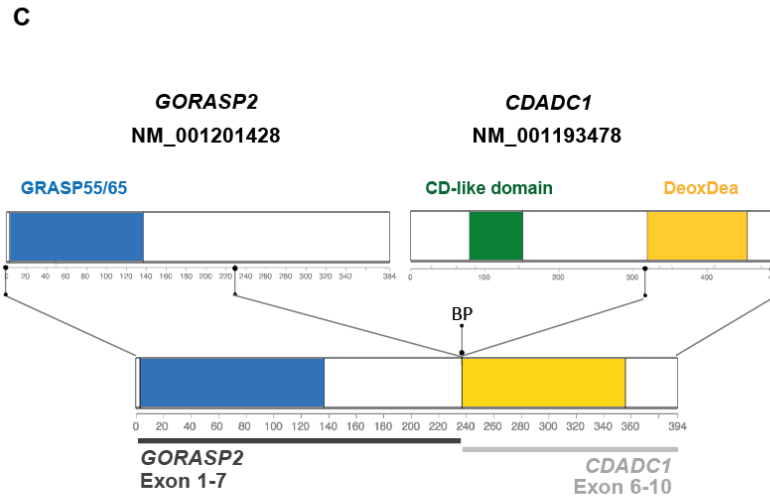
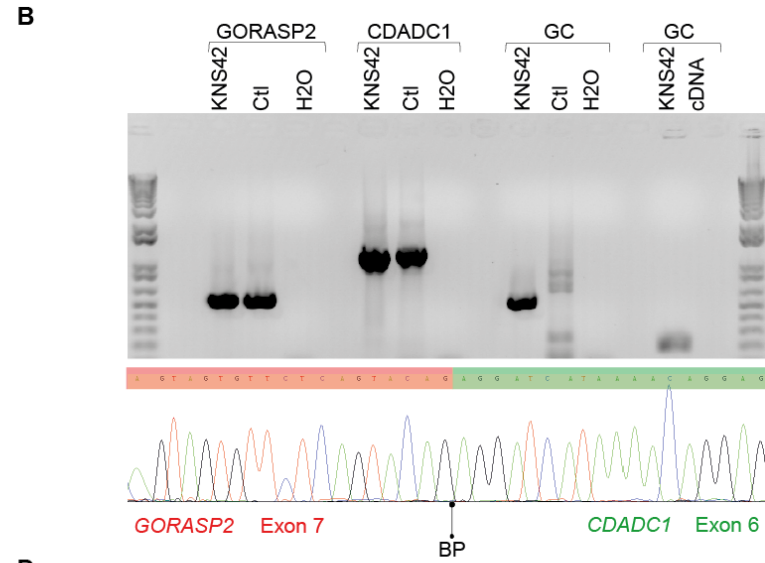
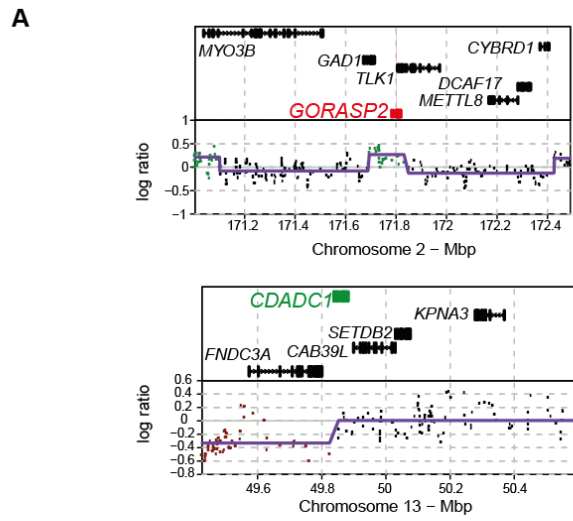


Figure 5.7. Identification of *GORASP2:CDADC1* gene fusion in KNS42 paediatric high grade glioma cell line. A) KNS42 500K Affymetrix aCGH copy number profile of *GORASP2* and *CDADC1*. B) PCR and RT-PCR detection of *GORASP2:CDADC1* in KNS42 genomic DNA and cDNA. Ctl: normal human genomic DNA. Partial sequence of the *GORASP2:CDADC1* fusion transcript identified in KNS42. C) Schematic representation of *GORASP2* and *CDADC1* genes as well as the formation of the fusion transcript involving exons 1-12 of *GORASP2* and exons 4-11 of *CDADC1*. D) Expression of *GORASP2* and *CDADC1* exons in KNS42 vs UW479 and SF188 cell lines. *GORASP2*: golgi reassembly stacking protein 2 cytidine; *CDADC1*: dCMP deaminase domain containing 1; GC: *GORASP2:CDADC1*; GRASP55/65: GRASP55/65 PDZ like domain; CD-like dom: cytidine deaminase like domain; Deoxy Dea: deoxycytidylate deaminase domain

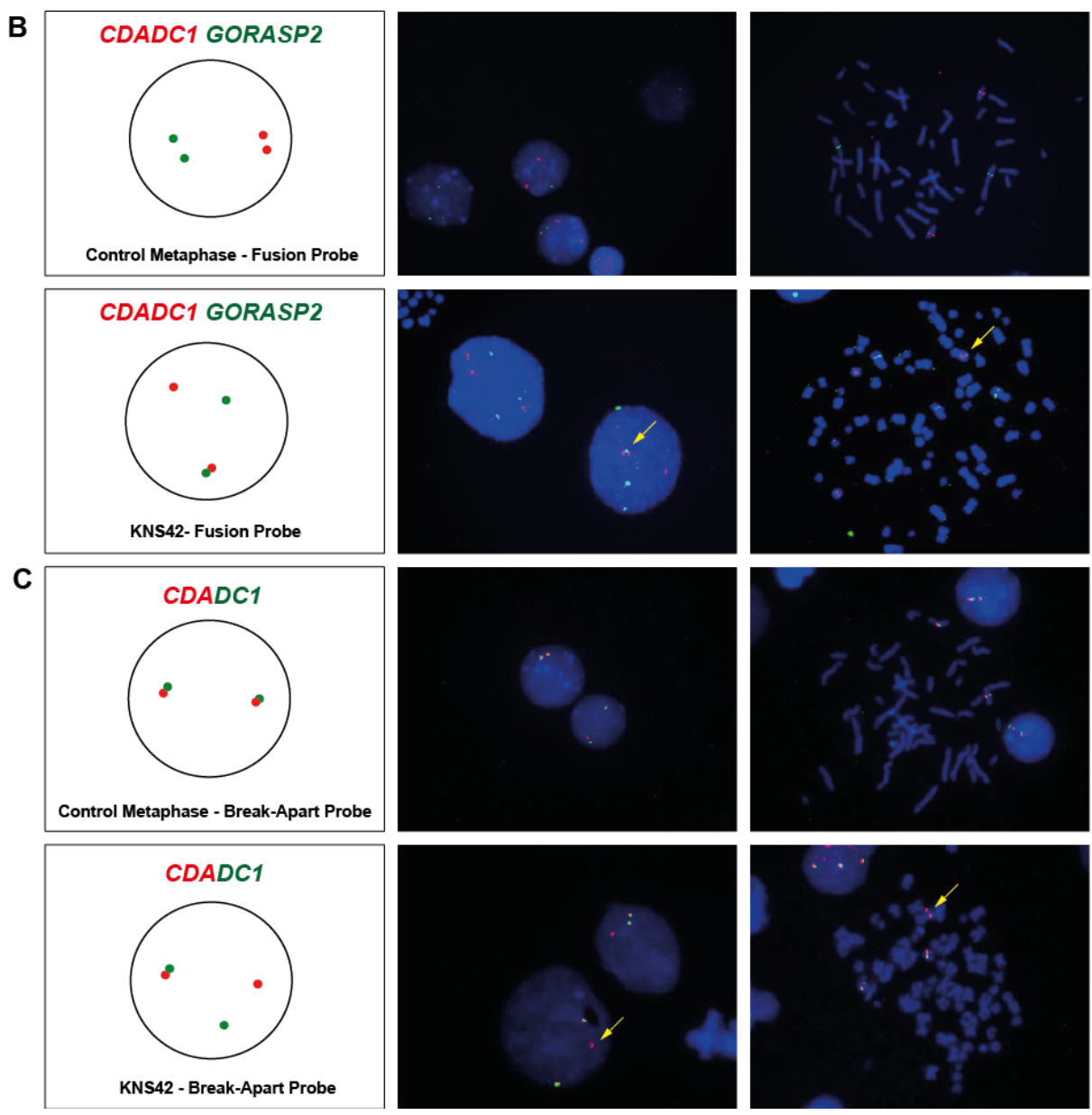
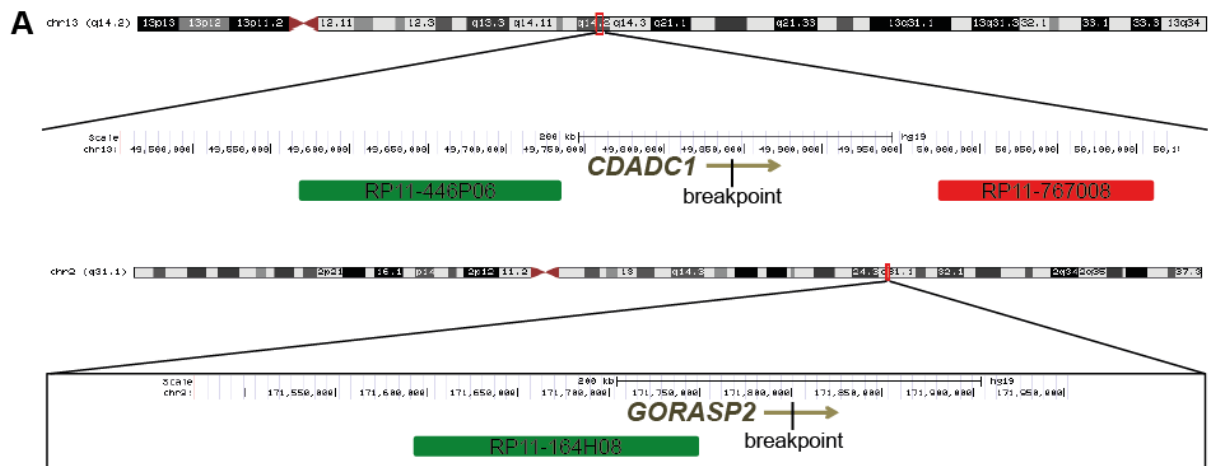


Figure 5.8. FISH of *GORASP2-CDADC1* rearrangement in KNS42. A) Region specific BAC probes used for FISH. Probes were selected according to hg19 build of the human genome to overlap with the region of *CDADC1* that is involved in the fusion (RP11-767008; red bar) and to the region of the same gene that is not disrupted (RP11-446P06; green bar). Another probe covering the region of *GORASP2* involved in the fusion was also selected (RP11-164H08; green bar). B) Diagram of the FISH signals expected when using fusion probes for *GORASP2* and *CDADC1*, followed by interphase and metaphase FISH in control tissue and in the positive fusion cell line, KNS42. The yellow arrows indicate the fusion in KNS42. C) Representative cartoon of the expected signals when using a *CDADC1* break-apart probe. FISH onto interphase and metaphase cells of control tissue and KNS42, showing split of 1 *CDADC1* signal (yellow arrow) which indicates the rearrangement of this gene. Original magnification X1000.

The phenomenon of gene amplification in cancer cells is often manifested either as extra-chromosomal self-replicating double-minutes (DMs) or as uniformly staining, linearly integrated chromosomal segments called homogeneously staining regions (HSRs) [36]. DMs have been previously reported in GBM and other cancer types as a mean of amplifying oncogenes [37-40]. In order to determine if the amplicons in SF188 were present as DMs or HSRs, we performed multi-coloured FISH in SF188 metaphase spreads. As shown in Figure 5.10, all amplicons were present as DMs and all regions in these amplicons showed similar amplification levels as predicted by the aCGH data (Figure 5.9), apart from the 11p11 region where levels of amplification were considerably lower (Figure 5.10, suggesting the presence of different double minute structures. Interestingly, we found that some SF188 cells contained a high number of DMs with *CDK4* (12q14) whereas other cells just harboured few DMs of the same amplicon, a result of an uneven cell division and heterogeneity within the cell line. The same was true for the *CCND1* (11q13) amplicon (Figure 5.10).

In order to resolve the structure of these complex rearrangements and understand if these DMs were present in more than one extra-chromosomal structure, we performed dual FISH, combining probes for each amplicon listed by Breakdancer. We found that *SCFD2* (4q12), *FIP1L1* (4q12), *MYC* (8q24), *MLL* (11q23) and *CDK4* (12q14) were present in the same DM (Figure 5.11). Interestingly, we found that the DMs containing *MYC* (8q24) are sometimes isolated (Figure 5.11B, 5.11D) (yellow arrows) or in the same structure as *SCFD2* (4q12) (Figure 5.11B). or *CDK4* (12q14) (Figure 5.11D). Combination of FISH probes for 11p11 and 11q13 (*CCND1*) with 4q12 (*SCFD2*) revealed that these exist in different DM structures (Figure 5.12A, 5.12B). Moreover, distinct subpopulations of cells were observed when combining *CCND1* (11q13) with *CDK4* (12q14) (Figure 5.12E). A population of cells harboured only *CDK4* amplification (Figure 5.12E green arrow), a second population contained only *CCND1* amplification (Figure 5.12E red arrow), and some cells contained both amplifications in the same DM (Figure 5.12E yellow arrow). To demonstrate which amplicons were present in the second DM structure, we then combined 11q13 (*CCND1*) with either 11p11 or 11p14 and observed that these are present in the same extra-chromosomal structure, as shown by FISH in Figure 5.13. To further confirm that these DMs were separated structures we combined FISH probes for 11p14 with either 4q12 (Figure 5.12C) or *CDK4* (12q14) (Figure 5.12D) or *MYC* (8q24) (Figure 5.12F). This FISH demonstrated once more that the amplicons occur in different DMs. Though the rearrangements involving these four different chromosomes were too complex for all the fragments to be resolved, a possible structure for each DM based on the FISH and Breakdancer results is shown in Figure 5.14.

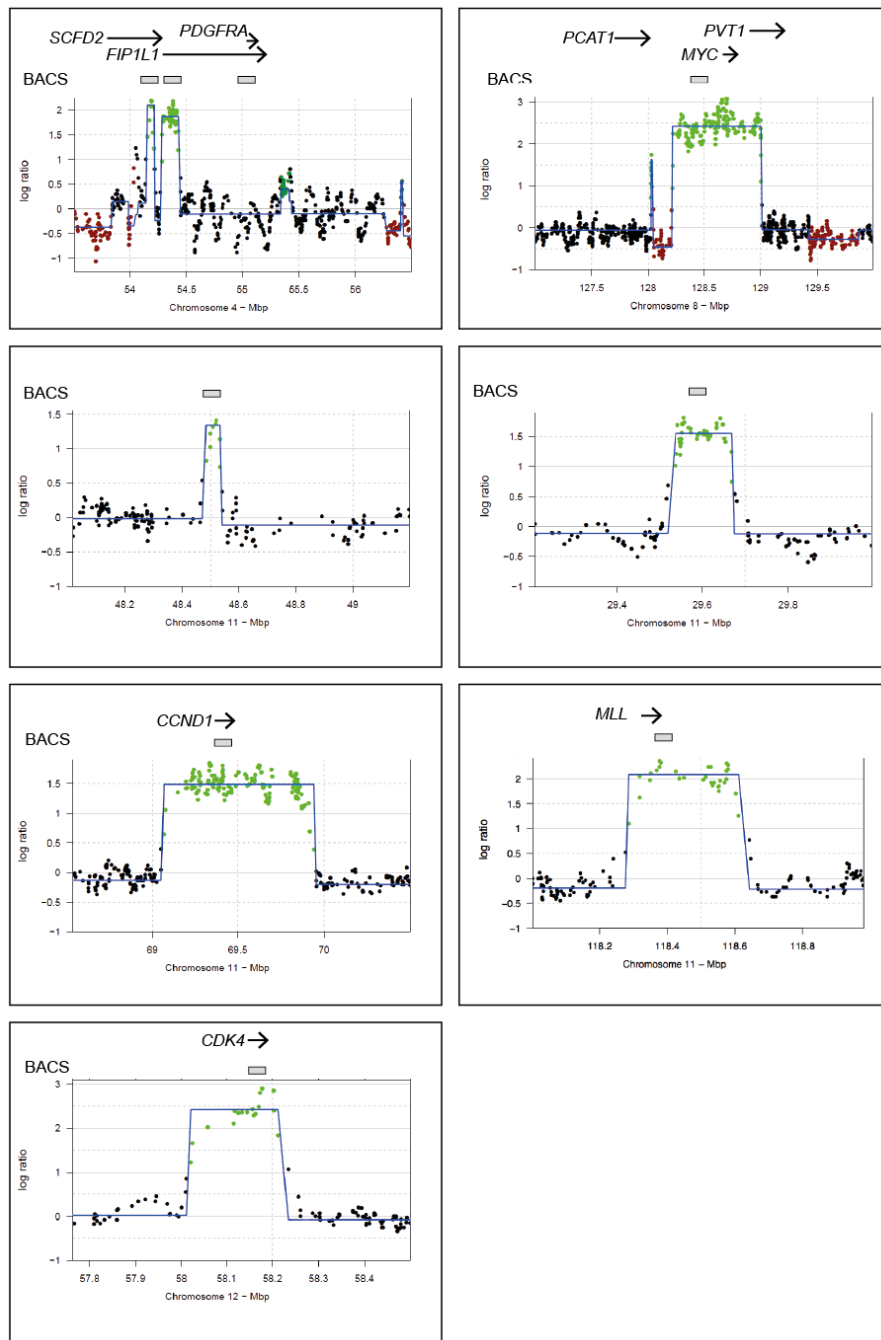


Figure 5.9. SF188 amplicons. Affymetrix 500K SNP array of chromosome 4, 8, 11 and 12 showing the genes involved in each amplicon. BAC clones selected for FISH validation are represented underneath each gene.

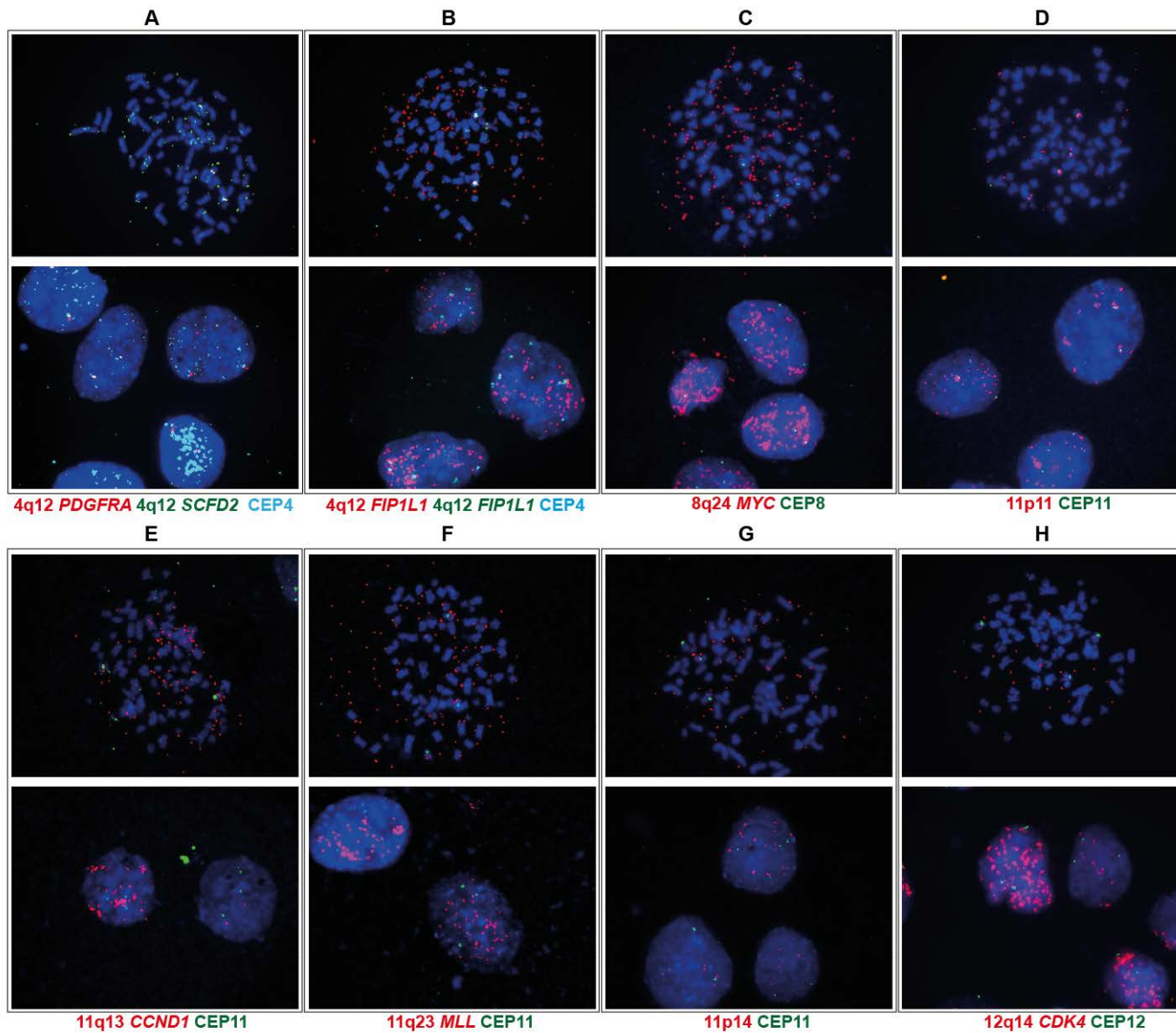


Figure 5.10. FISH validation of the amplification peaks in SF188. Specific BAC probes for 8q24 (*MYC*) (C), 11p11 (D), 11q13 (*CCND1*) (E), 11q23 (*MLL*) (F), 11p14 (G), 12q14 (*CDK4*) (H) were labelled with Cy3 (red) and co-hybridized to SF188 interphase and metaphase cells with chromosome specific control probes labelled with fluorescein (green). Centromere 4 was labelled with Aqua and 4q12 gene *PDGFRA* was labelled with Cy3 (A), *SCFD2* with fluorescein (A) and *FIP1L1* with both (B). FISH shows that all regions amplified are in double minutes structures (B-H). *PDGFRA* is not amplified in SF188 (A) and only one region of the *FIP1L1* gene in the 4q12 region is amplified (B). The levels of amplification of all regions are similar, with the exception of 11p11 (D) and 11p14 (G) which appear to have significantly fewer copies.

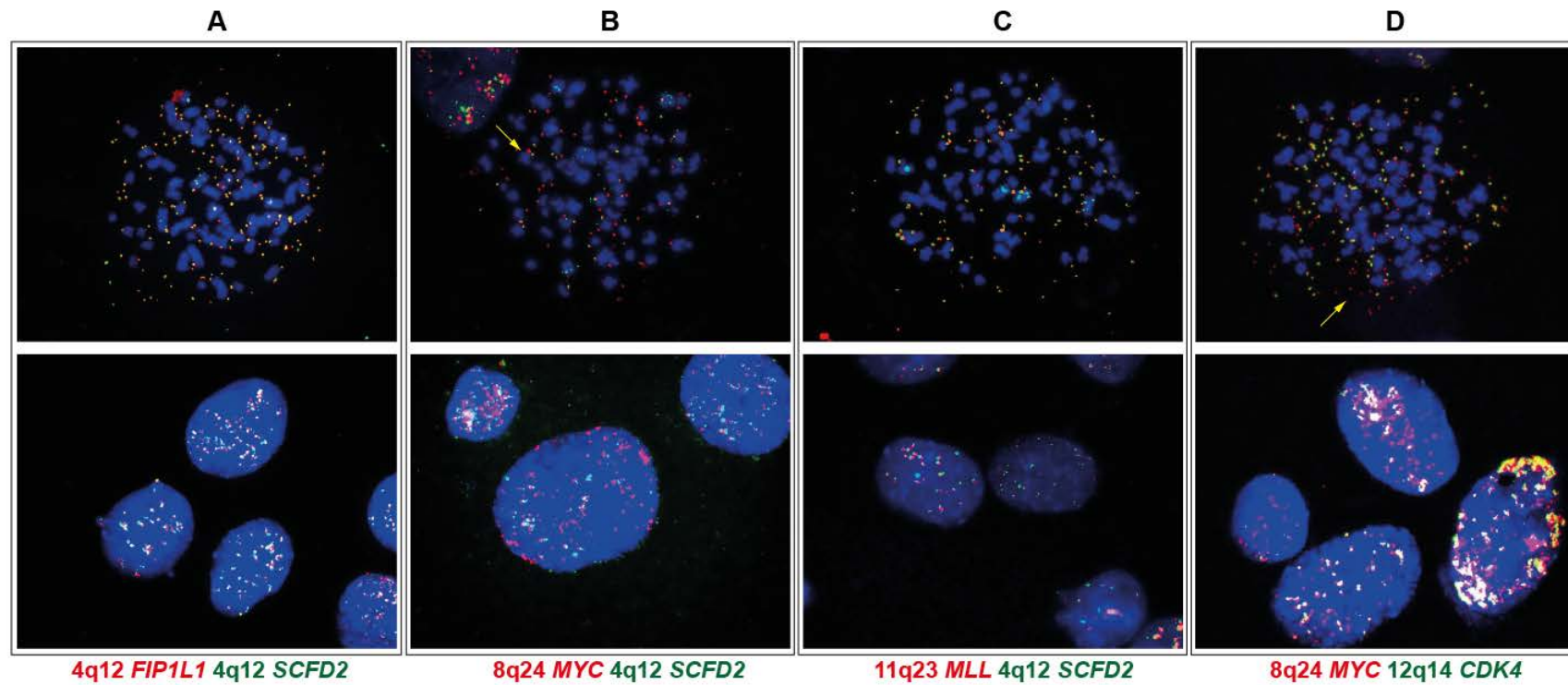


Figure 5.11. Dual FISH for the different amplicons in SF188. *SCFD2* (4q12), *FIP1L1* (4q12) (A), *MYC* (8q24) (B), *MLL* (11q23) (C), *CDK4* (12q14) (D) are present in the same double minute structure. The yellow arrows show that *MYC* is not always amplified in the same double minute with *SCFD2* and *CDK4*, but may be seen by itself.

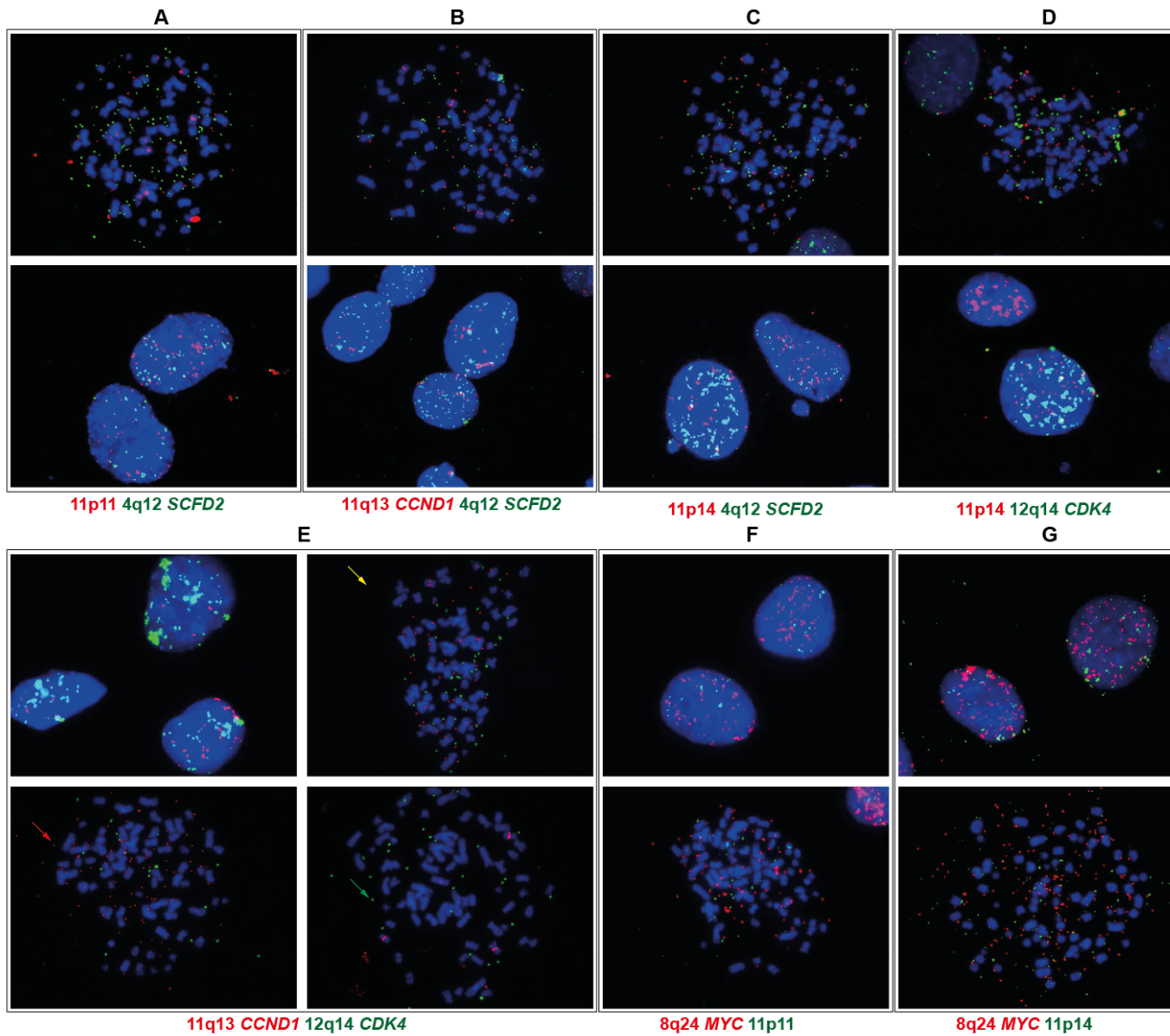


Figure 5.12. Dual FISH on SF188 showing that some regions are amplified in distinct double minute structures. 11p11 (A), 11q13 (*CCND1*) (B) and 11p14 (C) are not amplified with 4q12 (*SCFD2*) and 12q14 (*CDK4*) (D) in the same double minute structure. Dual FISH for 11q13 (*CCND1*) and 12q14 (*CDK4*) (E) show a mixed population with cells harbouring *CCND1* amplification, some only *CDK4*, and others both.

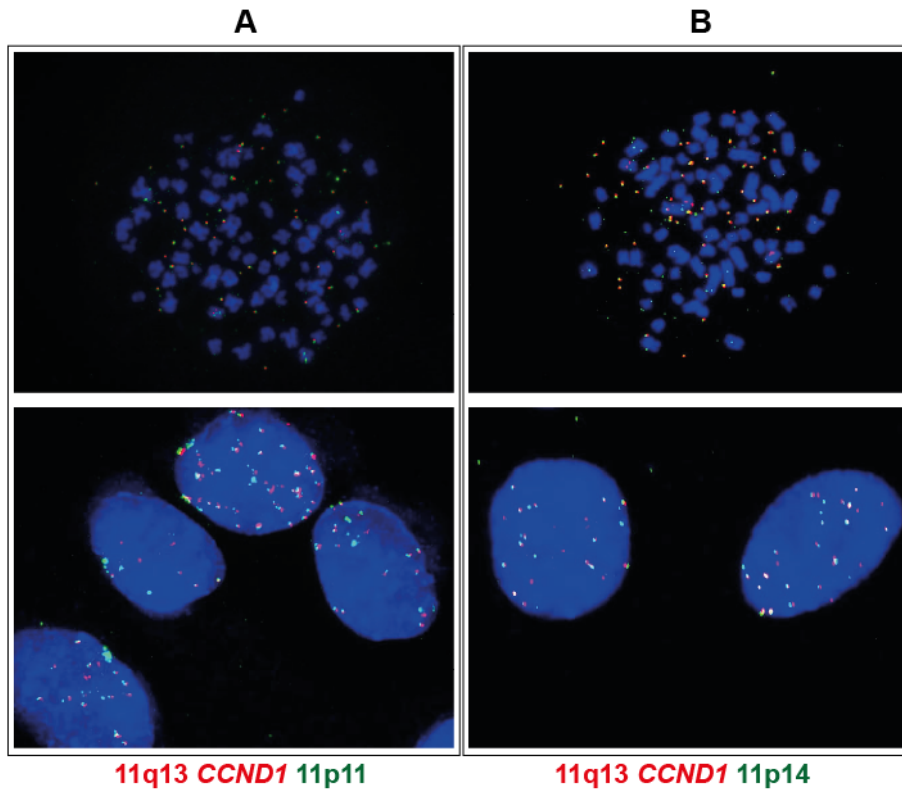


Figure 5.13. Dual FISH on SF188 showing that 11p11 (A), 11q13 (*CCND1*) (A and B) and 11p14 (B) are amplified in the same double minute structure.

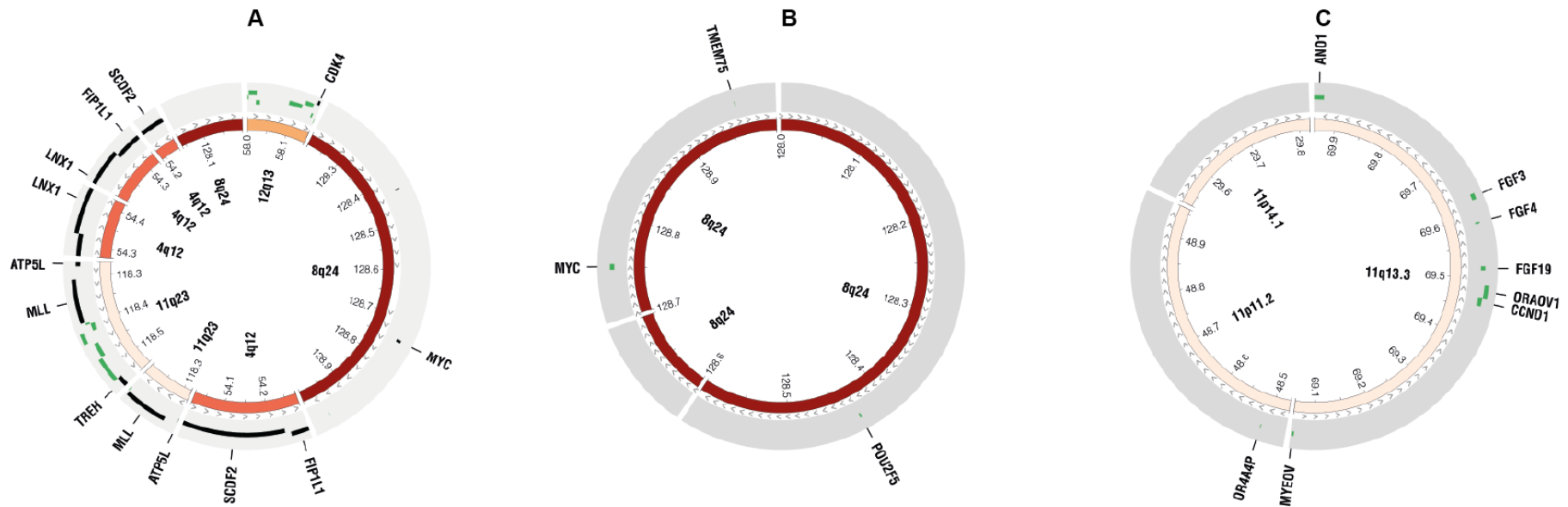


Figure 5.14. Double minutes structures present in SF188 paediatric high grade glioma cell line displayed as Circos plots (predicted structures). Chromosomes 4q12, 8q24, 11q23 and 12q14 are present in one double minute (A). The second double minute is constituted of chromosome 11p11, 11q13 and 11p14 (C). *MYC* is sometimes present as an independent structure (B). Outer ring contains copy number derived from whole genome sequencing data (green=gain), annotated for the genes involved in the amplicons, followed by DNA orientation (grey arrows) and chromosomal location in the inner ring.

5.4 DISCUSSION

In this chapter we report the landscape of gene fusions from three pHGG cell lines analysed by whole genome and transcriptome sequencing. Integrative analysis of a panel of three pHGG cell lines identified 17 novel fusion genes, all of which were localized to regions of amplification. Several distinct types of SVs were identified, including a complex extra-chromosomal rearrangement present in SF188 that leads to amplification of multiple oncogenes through the production of DMs

The pHGG cell lines were derived from patient tumour samples many years previously, and as such lacked matched normal controls that could be used to determine germline variation. Even using a very stringent filtering method to filter Breakdancer output, the number of predicted rearrangements was very high. The predicted rearrangements were spread throughout the genome and therefore reflect the patterns seen in other solid tumours [41-43], rather than as hotspots of rearrangements observed in leukaemia. Only 23% (17/74) of the SVs that coincide with genic regions were validated by PCR, and further 36.4% of these were validated by FISH (SF188 DMs), which accounts for a total of 59.4% validated SVs across the three cell lines. Other studies show that even when patient matched normal samples are sequenced and used to remove false positives, the validation rate is never 100%. Berger and colleagues showed that in prostate cancer 78% of candidate rearrangements were validated using multiplexed PCR followed by massively parallel sequencing, and in this case matched normal samples were used to remove the false positives [44]. Contrary to other cancers, e.g. breast [42] and colorectal cancer [45], the majority of rearrangements validated in pHGG cell lines were inter-chromosomal.

Integrating relative copy number data with the breakpoints is an important approach to assist in filtering putative SVs, as it has been shown that repeated breaks and rejoining of chromosomes during chromosomal amplifications leads to the generation of amplicon-associated fusion genes [33]. Furthermore, other studies have reported that focal amplifications of certain genes resulted in the creation of chimaeras, such as the *FGFR3:TACC3* fusion described in adult GBM [6]. Some of the rearranged genes reported in this study have unknown function and/or have not been related to cancer before (eg. *SI*, *ASPH*). Others have been reported in the literature as harbouring breakpoints in multiple cancer samples, such as the methyltransferase-encoding *PRMT8*, the O-acetyl-ADP-ribose deacetylase *MACROD2*, both disrupted in UW479, and the chloride ion channel *CLIC5* detected in KNS42 [45].

The intra-chromosomal translocation discovered in UW479 that creates the *NUBPL:AKAP6* fusion involves two genes that have been connected to a specific genetic disorder and to cancer, respectively. *NUBPL* is a gene required for the assembly of the respiratory chain NADH dehydrogenase (complex I), an oligomeric enzymatic complex located in the inner mitochondrial membrane. Little is known about its structure, but disruption of this gene through missense mutations leads to mitochondrial complex I deficiency in humans [46]. A deletion of *NUBPL* exons 1-4 with duplication of exon 7, leading to inactive *NUBPL*, has been found in this disorder [46, 47], and corresponds to the same region in UW479. *NUBPL:AKAP6* fusion loses an important functional domain of *AKAP6*, the Ser/Thr protein kinase A (PKA) RII interacting domain. *AKAP6* is part of A-kinase anchoring proteins family, which are key components of signal transduction. AKAP family members are structurally different yet functionally related proteins that bind and anchor the PKA to specific subcellular sites, thereby confining PKA activity to potential substrates, such as cyclic adenosine monophosphate [48]. In addition, there is evidence that AKAP family members expression and gene variation are directly associated with cancer development. *AKAP3* mRNA expression appears to be associated with poor prognosis in epithelial ovarian cancer [49], and genetic alterations of *AKAP9*, *AKAP11*, *AKAP12*, and *AKAP13* are associated with the aetiology of colorectal and lung cancer [50, 51], prostate, and breast cancers [52-54]. Moreover, Ciampi and colleagues reported an in-frame fusion between the C-terminal catalytic domain (exons 9 – 18) of the Ser/Thr kinase BRAF and the N-terminus (exons 1 – 8) of another member of the AKAPs family, *AKAP9*, leading to constitutive BRAF and MAPK pathway activation. *AKAP9* which was normally located in the centrosome and Golgi, lacked the C-terminal centrosomal domain, losing its centrosome location in cancer cells [55]. Although the role of *NUBPL:AKAP6* is unclear, one could speculate that both genes play important physiological functions and their disruption in UW479 can either inactivate a copy of *AKAP6*, by disruption of its important PKA RII interacting domain, or lead to an inactive *NUBPL*, which plays an important role in cell metabolism.

The *GORASP2:CDADC1* chimera is differently expressed in KNS42 compared to the other lines, which argues for a functionally significant role. Though neither gene has previously been associated with cancer, it is known that *GORASP2* plays an important part in the regulation of mitosis through Golgi fragmentation [56]. The *GORASP2* domain involved in this fusion (N-terminal GRASP domain) is essential for protein oligomerization/dimerization and for maintaining the Golgi complex structure. *GORASP2* forms a complex matrix with *GRASP65* and *GM130* which is essential for the structure and function of the Golgi complex. When these proteins are phosphorylated by *ERK1/CDK1/PLK1*, the Golgi is disassembled and cells enter mitosis [56]. Interestingly, it has been shown that the GRASP domain of

GORASP2 is necessary and sufficient for oligomerization, which leads to inhibition of Golgi fragmentation in living cells resulting in a delay of the G2/M transition [57]. There is only one report available on *CDADC1* where this gene was first identified [58], with no further information available on its structure and function. This fusion may activate *CDADC1* or create a functionally active *GORASP2* protein independent of being regulated by mitosis.

Recently, Siyuan Zheng and colleagues identified a complex structural rearrangement that occurred within the same chromosome. A breakpoint-enriched region was identified at 12q14-15 with focal amplifications involving the *CDK4* and *MDM2* oncogenes. This pattern was identified in ~5% of GBM patient samples and is associated with poor survival [58]. It is evident that amplifications do not occur randomly throughout the genome, and regions that are amplified tend not to be deleted, and vice versa. The amplified loci normally contain oncogenes, as it is the case of the genes involved in the extra-chromosomal structures present in SF188: *MYC* on 8q24, *CCND1* on 11q13 and *CDK4* on 12q14. More importantly, even though SF188 is a cell line originated from one single clone, we observed selective heterogeneity within the line. Some cells contained DMs with *CDK4* and other cells contained DMs with *CCND1*, with few cells containing both. This reflects the asymmetric segregation of the DMs during cell division as they do not align along the mitotic spindle. These type of complex co-amplifications have also been described in other cancer types, such as breast cancer samples and cell lines [59, 60].

As well as the novel fusion identification described above, the data provided in this chapter will form a valuable reference for future functional studies of the disease, as these cell lines are part of a small number of available pHGG models, with KNS42 representing the only available model for the study of the *H3F3A* G34V mutation.

5.5. References

1. Parsons, D.W., et al., *An integrated genomic analysis of human glioblastoma multiforme*. Science, 2008. **321**(5897): p. 1807-12.
2. Phillips, H.S., et al., *Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis*. Cancer Cell, 2006. **9**(3): p. 157-73.
3. Paugh, B.S., et al., *Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease*. J Clin Oncol, 2010. **28**(18): p. 3061-8.
4. Bax, D.A., et al., *A distinct spectrum of copy number aberrations in pediatric high-grade gliomas*. Clin Cancer Res, 2010. **16**(13): p. 3368-77.
5. Sturm, D., et al., *Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma*. Cancer Cell, 2012. **22**(4): p. 425-37.
6. Singh, D., et al., *Transforming fusions of FGFR and TACC genes in human glioblastoma*. Science, 2012. **337**(6099): p. 1231-5.
7. Ozawa, T., et al., *PDGFRA gene rearrangements are frequent genetic events in PDGFRA-amplified glioblastomas*. Genes Dev, 2010. **24**(19): p. 2205-18.
8. Mitelman, F., B. Johansson, and F. Mertens, *The impact of translocations and gene fusions on cancer causation*. Nat Rev Cancer, 2007. **7**(4): p. 233-45.
9. Calasanz, M.J. and J.C. Cigudosa, *Molecular cytogenetics in translational oncology: when chromosomes meet genomics*. Clin Transl Oncol, 2008. **10**(1): p. 20-9.
10. Edwards, P.A., *Fusion genes and chromosome translocations in the common epithelial cancers*. J Pathol, 2010. **220**(2): p. 244-54.
11. Shtivelman, E., et al., *Fused transcript of abl and bcr genes in chronic myelogenous leukaemia*. Nature, 1985. **315**(6020): p. 550-4.
12. Tomlins, S.A., et al., *Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer*. Science, 2005. **310**(5748): p. 644-8.
13. Mao, X., et al., *Chromosome rearrangement associated inactivation of tumour suppressor genes in prostate cancer*. Am J Cancer Res, 2011. **1**(5): p. 604-17.
14. Chin, L., et al., *Making sense of cancer genomic data*. Genes Dev, 2011. **25**(6): p. 534-55.
15. Stephens, P.J., et al., *Massive genomic rearrangement acquired in a single catastrophic event during cancer development*. Cell, 2011. **144**(1): p. 27-40.
16. Liu, P., et al., *Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements*. Cell, 2011. **146**(6): p. 889-903.
17. Tomlins, S.A., et al., *Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer*. Nature, 2007. **448**(7153): p. 595-9.

18. Mertz, K.D., et al., *Molecular characterization of TMPRSS2-ERG gene fusion in the NCI-H660 prostate cancer cell line: a new perspective for an old model*. *Neoplasia*, 2007. **9**(3): p. 200-6.
19. Ha, K.C., et al., *Identification of gene fusion transcripts by transcriptome sequencing in BRCA1-mutated breast cancers and cell lines*. *BMC Med Genomics*, 2011. **4**: p. 75.
20. Koivunen, J.P., et al., *EML4-ALK fusion gene and efficacy of an ALK kinase inhibitor in lung cancer*. *Clin Cancer Res*, 2008. **14**(13): p. 4275-83.
21. Rutka, J.T., et al., *Establishment and characterization of five cell lines derived from human malignant gliomas*. *Acta Neuropathol*, 1987. **75**(1): p. 92-103.
22. Takeshita, I., et al., *Characteristics of an established human glioma cell line, KNS-42*. *Neurol Med Chir (Tokyo)*, 1987. **27**(7): p. 581-7.
23. Bobola, M.S., et al., *O6-methylguanine-DNA methyltransferase, O6-benzylguanine, and resistance to clinical alkylators in pediatric primary brain tumor cell lines*. *Clin Cancer Res*, 2005. **11**(7): p. 2747-55.
24. Bax, D.A., et al., *Molecular and phenotypic characterisation of paediatric glioma cell lines as models for preclinical drug development*. *PLoS One*, 2009. **4**(4): p. e5209.
25. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform*. *Bioinformatics*, 2009. **25**(14): p. 1754-60.
26. Chen, K., et al., *BreakDancer: an algorithm for high-resolution mapping of genomic structural variation*. *Nat Methods*, 2009. **6**(9): p. 677-81.
27. Iyer, M.K., A.M. Chinnaiyan, and C.A. Maher, *ChimeraScan: a tool for identifying chimeric transcription in sequencing data*. *Bioinformatics*, 2011. **27**(20): p. 2903-4.
28. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome*. *Genome Biol*, 2009. **10**(3): p. R25.
29. Koressaar, T. and M. Remm, *Enhancements and modifications of primer design program Primer3*. *Bioinformatics*, 2007. **23**(10): p. 1289-91.
30. Untergasser, A., et al., *Primer3--new capabilities and interfaces*. *Nucleic Acids Res*, 2012. **40**(15): p. e115.
31. Wang, Q., et al., *Application of next generation sequencing to human gene fusion detection: computational tools, features and perspectives*. *Brief Bioinform*, 2013. **14**(4): p. 506-19.
32. Hao, J.J., et al., *Characterization of gene rearrangements resulted from genomic structural aberrations in human esophageal squamous cell carcinoma KYSE150 cells*. *Gene*, 2013. **513**(1): p. 196-201.
33. Kalyana-Sundaram, S., et al., *Gene fusions associated with recurrent amplicons represent a class of passenger aberrations in breast cancer*. *Neoplasia*, 2012. **14**(8): p. 702-8.

34. Jones, D.T., et al., *Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas*. *Cancer Res*, 2008. **68**(21): p. 8673-7.
35. Mohan, M., et al., *Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis*. *Nat Rev Cancer*, 2010. **10**(10): p. 721-8.
36. Gebhart, E., *Double minutes, cytogenetic equivalents of gene amplification, in human neoplasia - a review*. *Clin Transl Oncol*, 2005. **7**(11): p. 477-85.
37. Lalic, H., C. Volavsek, and B. Radosevic-Stasic, *Chromosomal instability and double minute chromosomes in a breast cancer patient*. *Acta Med Okayama*, 2004. **58**(1): p. 51-8.
38. Kuttler, F. and S. Mai, *Formation of non-random extrachromosomal elements during development, differentiation and oncogenesis*. *Semin Cancer Biol*, 2007. **17**(1): p. 56-64.
39. Vogt, N., et al., *Molecular structure of double-minute chromosomes bearing amplified copies of the epidermal growth factor receptor gene in gliomas*. *Proc Natl Acad Sci U S A*, 2004. **101**(31): p. 11368-73.
40. Thomas, L., et al., *Double minute chromosomes in monoblastic (M5) and myeloblastic (M2) acute myeloid leukemia: two case reports and a review of literature*. *Am J Hematol*, 2004. **77**(1): p. 55-61.
41. Pleasance, E.D., et al., *A small-cell lung cancer genome with complex signatures of tobacco exposure*. *Nature*, 2010. **463**(7278): p. 184-90.
42. Stephens, P.J., et al., *Complex landscapes of somatic rearrangement in human breast cancer genomes*. *Nature*, 2009. **462**(7276): p. 1005-10.
43. Campbell, P.J., et al., *Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing*. *Nat Genet*, 2008. **40**(6): p. 722-9.
44. Berger, M.F., et al., *The genomic complexity of primary human prostate cancer*. *Nature*, 2011. **470**(7333): p. 214-20.
45. Bass, A.J., et al., *Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion*. *Nat Genet*, 2011. **43**(10): p. 964-8.
46. Tucker, E.J., et al., *Next-generation sequencing in molecular diagnosis: NUBPL mutations highlight the challenges of variant detection and interpretation*. *Hum Mutat*, 2012. **33**(2): p. 411-8.
47. Calvo, S.E., et al., *High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency*. *Nat Genet*, 2010. **42**(10): p. 851-8.
48. Moss, S.B. and G.L. Gerton, *A-kinase anchor proteins in endocrine systems and reproduction*. *Trends Endocrinol Metab*, 2001. **12**(10): p. 434-40.

49. Sharma, S., et al., *A-kinase anchoring protein 3 messenger RNA expression correlates with poor prognosis in epithelial ovarian cancer*. *Gynecol Oncol*, 2005. **99**(1): p. 183-8.
50. Rudd, M.F., et al., *Variants in the GH-IGF axis confer susceptibility to lung cancer*. *Genome Res*, 2006. **16**(6): p. 693-701.
51. Webb, E.L., et al., *Search for low penetrance alleles for colorectal cancer through a scan of 1467 non-synonymous SNPs in 2575 cases and 2707 controls with validation by kin-cohort analysis of 14 704 first-degree relatives*. *Hum Mol Genet*, 2006. **15**(21): p. 3263-71.
52. Lewis, T.E., et al., *Tissue transglutaminase interacts with protein kinase A anchor protein 13 in prostate cancer*. *Urol Oncol*, 2005. **23**(6): p. 407-12.
53. Wirtenberger, M., et al., *Association of genetic variants in the Rho guanine nucleotide exchange factor AKAP13 with familial breast cancer*. *Carcinogenesis*, 2006. **27**(3): p. 593-8.
54. Frank, B., et al., *Association of a common AKAP9 variant with breast cancer risk: a collaborative analysis*. *J Natl Cancer Inst*, 2008. **100**(6): p. 437-42.
55. Ciampi, R., et al., *Oncogenic AKAP9-BRAF fusion is a novel mechanism of MAPK pathway activation in thyroid cancer*. *J Clin Invest*, 2005. **115**(1): p. 94-101.
56. Corda, D., et al., *Golgi complex fragmentation in G2/M transition: An organelle-based cell-cycle checkpoint*. *IUBMB Life*, 2012. **64**(8): p. 661-70.
57. Feinstein, T.N. and A.D. Linstedt, *Mitogen-activated protein kinase kinase 1-dependent Golgi unlinking occurs in G2 phase and promotes the G2/M cell cycle transition*. *Mol Biol Cell*, 2007. **18**(2): p. 594-604.
58. Zheng, S., et al., *A survey of intragenic breakpoints in glioblastoma identifies a distinct subset associated with poor survival*. *Genes Dev*, 2013. **27**(13): p. 1462-72.
59. Hampton, O.A., et al., *A sequence-level map of chromosomal breakpoints in the MCF-7 breast cancer cell line yields insights into the evolution of a cancer genome*. *Genome Res*, 2009. **19**(2): p. 167-77.
60. Hicks, J., et al., *Novel patterns of genome rearrangement and their association with survival in breast cancer*. *Genome Res*, 2006. **16**(12): p. 1465-79.

6. *RPTOR* rearrangements in paediatric high grade glioma

The computation analysis presented throughout this chapter was performed by Alan Mackay, Glioma Team, Institute of Cancer Research, Sutton, United Kingdom.

The FISH was performed with the help and valuable technical advice of Anna Burford, Glioma Team, Institute of Cancer Research, Sutton, United Kingdom

6. *RPTOR* REARRANGEMENTS IN PAEDIATRIC HIGH GRADE GLIOMA

6.1. Introduction

Many of the dysregulated signalling processes identified in both paediatric and adult GBM involve the mammalian target of rapamycin (mTOR), a PI3K-related protein kinase [1]. mTOR controls cell growth in response to energy, nutrients, growth factors and other environmental cues, and is also frequently deregulated in other cancers [2, 3]. In the previously described pHGG cell line sequencing study (chapter 5), we identified a novel gene fusion involving *RPTOR* in SF188, a cell line derived from an 8 year old boy with GBM.

The mTOR kinase is the catalytic subunit of two multiprotein complexes, mTORC1 and mTORC2 (Figure 6.1A). mTORC1 is defined by its regulatory associated protein of mTOR (RPTOR) subunit [4-6], which is replaced by RICTOR (rapamycin-insensitive RPTOR independent companion of mTOR) in mTORC2 [7]. Both complexes contain mammalian lethal with sec-13 (mLST8; also known as G β L) [6] and the recently identified DEP domain-containing mTOR-interacting protein (DEPTOR) [8], but differ in several subunits that interact with RPTOR and RICTOR [2]. Further unique components of mTORC1 include a negative regulator proline-rich AKT substrate 40 kDa (PRAS40) [9], whereas mTORC2 contains protein observed with rictor-1 (Protor-1) and Protor-2, which probably help complex assembly [10], and mammalian stress-activated protein kinase interacting protein 1 (mSIN1), which may target mTORC2 to membranes [11, 12]. These complexes perform non-overlapping functions within the cell. mTORC2 primarily responds to growth factors, promoting cell cycle entry, cell survival, actin cytoskeleton polarization and anabolic output, contrary to mTORC1, which regulates cell growth by prompting translation, ribosome biogenesis and autophagy [13].

mTORC1 drives growth through two downstream substrates, eukaryotic initiation factor 4E binding protein 1 (4E-BP1) and the S6 kinase 1 (S6K1) (Figure 6.1B) [14, 15], which control cap-dependent translation initiation and elongation, respectively. The regulation of the activity of mTORC1 towards these downstream targets and yet unidentified substrates appears to be complex and is likely to be dependent on the organization of the various subunits in the mTORC1 complex. RPTOR functions as an essential scaffolding protein and it is indispensable for mTOR to phosphorylate 4E-BP1 and S6K1, and for mTORC1 complex formation and dimerization [4]. RPTOR consists of a conserved N-terminal region followed by three HEAT repeats and seven WD40 repeats.

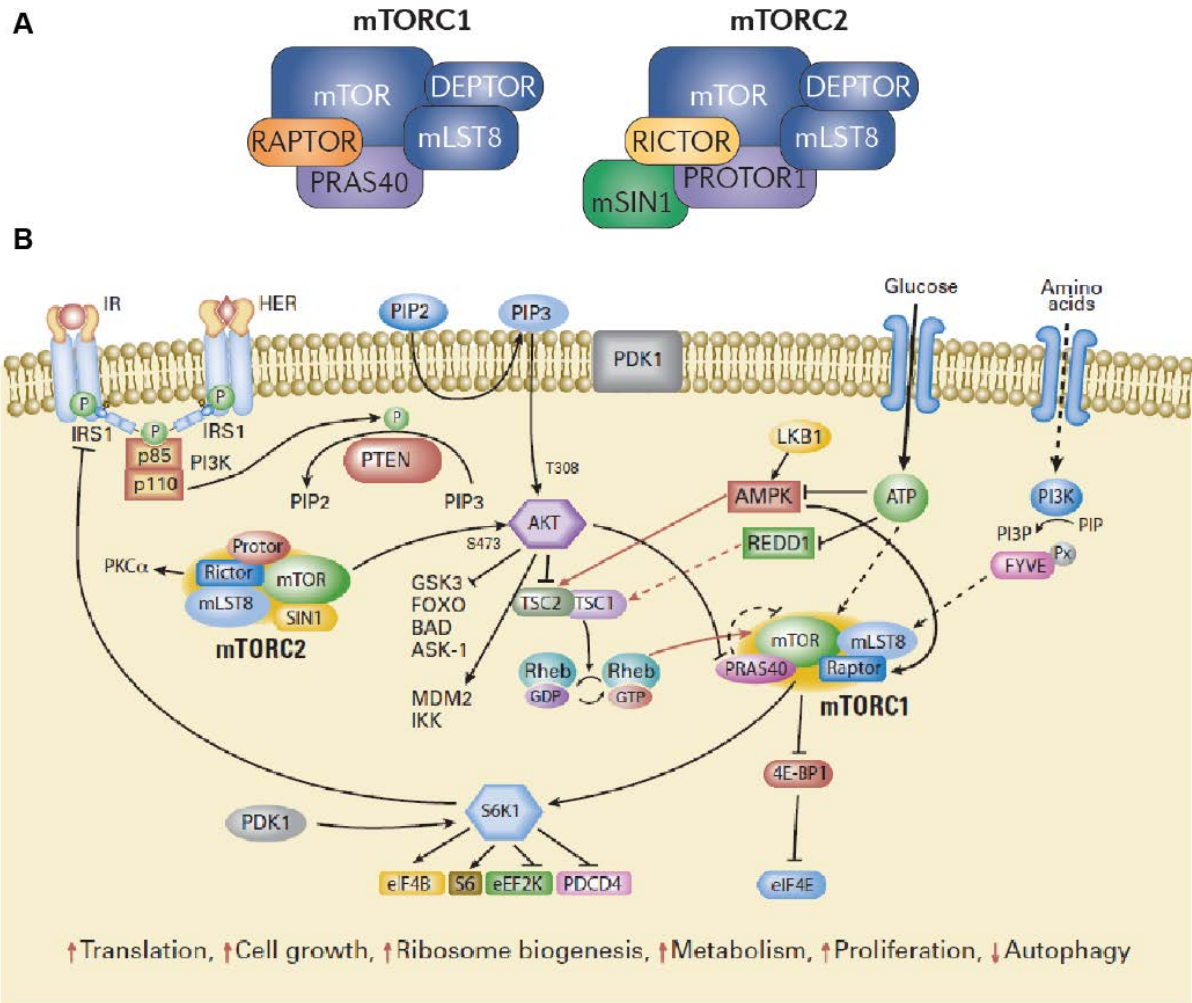


Figure 6.1. (A), Mammalian target of rapamycin (mTOR) complexes (mTORC1 and mTORC2) contain in addition to mTOR, DEPTOR (DEP-domain-containing mTOR-interacting protein) and mLST8 (mammalian lethal with SEC13 protein 8; also known as G β L). Other protein partners differ between the two complexes. Whereas mTORC1 contains RAPTOR (regulatory-associated protein of mTOR) and PRAS40 (proline-rich AKT substrate 40 kDa), mTORC2 contains RICTOR (rapamycin-insensitive companion of mTOR), mSIN1 (mammalian stress activated MAP kinase-interacting protein 1) and PROTOR (protein observed with RICTOR). **(B)** The mTOR signalling network. Arrows represent activation, bars represent inhibition. mTOR signalling regulates multiple critical cellular processes by integrating energy and nutrient status and PI3K/Akt signalling induced by growth factors and insulin. Adapted from [16, 17].

The RPTOR-mTOR interaction is very dynamic, and is thought to require the HEAT repeats of mTOR. The strength of interaction between mTOR and RPTOR can be modified by nutrients and other signals that regulate the mTORC1 pathway, but how this translates into regulation of the pathway remains vague. It has been shown that when in the presence of nutrients and amino acids, mTORC1 is recruited to lysosomes and late endosomes, a

process mediated by RPTOR, and co-localizes with its activator, the small GTPase Ras homolog enriched in brain (RHEB). RHEB is predicted to bind to and activate the mTOR kinase domain, and displace the mTORC1 inhibitor PRAS40 from RPTOR, leading to phosphorylation of 4E-BP1 and S6K, and mRNA translation. RPTOR phosphorylation on S722/792 is mediated by AMP-activated protein kinase (AMPK) and is required for the inhibition of mTORC1 activity induced by energy stress, whereas phosphorylation on S719/721/722 is mediated by p90 ribosomal S6 kinases (RSKs) and contributes to the activation of mTORC1 by mitogen stimulation.

Given our identification of the novel *RPTOR:TULP4* fusion in SF188, we sought to determine if SVs frequently involve *RPTOR* in paediatric and adult HGG samples, and begin to unravel the biological consequences of such *RPTOR* disruptions.

6.2. MATERIAL AND METHODS

6.2.1. Patient samples

A series of 213 (53 paediatric samples and 160 adult cases) (Appendix I) formalin-fixed, paraffin- embedded (FFPE) samples were retrieved after approval from Wandsworth Research Ethics Committee approval from 213 consecutive patients diagnosed with GBM (WHO grade IV), anaplastic astrocytoma (WHO grade III) or anaplastic oligodendroglioma (WHO grade III) within the last 5 years from the archives of King's College Hospital, with diagnosis confirmed by re-review (S. Popov and S. Al-Sarraj).

1 mm diameter cores from each case were taken and constructed in 11 tissue microarrays (TMAs). In addition, a series of control tissues were included, comprising normal human brain, normal mouse brain, placenta and colon.

6.2.2 Fluorescent in-situ hybridization

FISH based on the break-apart or fusion probe strategy was used to validate the *RPTOR* gene aberrations. For the *RPTOR:TULP4* fusion probe, the following BAC clones were used: RP11-317F05 (red) and RP11-819H19 (green). For the break-apart probe the following BAC clones were used: RP11-317F05 (red), RP11-334C17 (red), RP11-819H19 (green) and RP11-387C10 (green). The cell line FISH protocol is described in chapter 5. The following protocol was used for FISH in TMAs: BAC DNA was amplified from 10 nanograms starting material using the Illustra™ GenomiPhi™ V2 DNA amplification kit (GE Healthcare, Little Chalfont, UK) according to manufacturer's instructions. Probes were labelled with biotin and dioxigenin (DIG) using the BioPrime® DNA Labeling System (Life Technologies, Paisley,

UK). Briefly, the FFPE tissue sections were deparaffinised in 3 changes of xylene, 10min each and hydrated with 3 min 100%, 90%, 70% and deionised H₂O (dH₂O). The slides were pre-treated with 0.2M HCl for 20min, washed in dH₂O for 3min and incubated in 8% (w/v) sodium thiocyanate at 80°C for 30min. Slides were washed in 2xSSC for 3min, prior to digestion for 20min at 37°C with 0.01M HCl/ 0.025% (w/v) pepsin (Sigma, Gillingham, UK) solution. Slides were washed in 2xSSC for 5min, dehydrated with 3min 70%, 90%, 100% ethanol and left to dry at room temperature for 10min. Denaturation was carried out for 5min at 75°C and hybridization with the appropriate biotin- and DIG-labelled probe was done overnight at 37°C in a humidified chamber. Removal of the coverslips was carried out by soaking the slides in 2xSSC/0.1% Igepal for 2 mins at room temperature and immersing in 0.4xSSC/0.3% Igepal at 73°C for 2min. Slides were rinsed in 2xSSC/0.1% Igepal for 1min and PBS for 3min. Slides were first incubated in a humidified chamber at 37°C with two layers of anti-DIG-fluorescein (Roche, Welwyn Garden City, UK) for 10min each, washed with PBS 3x3min and incubated with two layers of Streptavidin-cy3 (Sigma, Gillingham, UK) 10min each. Slides were soaked in PBS for 3min, dehydrated with 3min 70%, 90%, 100% ethanol and left to dry at room temperature for 10min. TMA slides were finally mounted in Vectashield with DAPI (Vector Laboratories, Peterborough, UK), and captured on the Ariol System (Leica Microsystems, Wetzlar, Germany) using filters for DAPI, Cy3 and FITC.

6.2.3. Cell lines

The cell line SF188 was previously described in chapter 5. Briefly, cells were grown in DMEM F12-HAM complete media (Sigma-Aldrich) supplemented with 10% foetal bovine serum (FBS), incubated at 37°C with 5%CO₂. DNA was extracted from the cell line and samples screened for mycoplasma at Surrey Diagnostics (Cranleigh, UK).

6.2.4. DNA and RNA extraction

Genomic DNA was extracted from SF188 pHGG cell line using the DNeasy Blood & Tissue Kit (Qiagen, Manchester, UK), and the quantity and quality of the DNA was assessed with Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies, Paisley, UK) and ND-1000 Nanodrop (Thermo Scientific, Loughborough, UK). Total RNA was extracted using the Absolutely RNA Miniprep Kit (Agilent, Wokingham, UK), and the quantity and quality of the RNA was assessed on the 2100 Bioanalyzer (Agilent, Wokingham, UK), using the Agilent RNA 6000 Nano Kit (Agilent, Wokingham, UK), according to manufacturer instructions. cDNA was synthesized using the SuperScript II system (Life Technologies, Paisley, UK) according to manufacturer's protocol.

6.2.5. PCR validation

Four PCR primers (IDT Technologies, Coralville, US) were designed using Primer3 [18, 19] for *RPTOR* and *TULP4*. These primers were designed ± 200 bp apart from the breakpoints to generate amplicons spanning the breakpoint-junction-sequences of predicted SV. In addition to fusion gene amplification, we amplified both partner genes involved. The following primer combinations were performed for DNA: *RPTOR* Fwd [5'- AGTTTGTCTACCGCCCTTG-3'] x *RPTOR* Rev [5'-CCTTTGGTGGGAAGCTCCTC-3'], *TULP4* Fwd [5'-GTAGCATCAGCTCTTGCCCA-3'] x *TULP4* Rev [5'-GCTGGACAATACTCCCTGGC-3'], *RPTOR* Fwd x *TULP4* Rev, and for cDNA *RPTOR* Exon6 Fwd [5'-GTAGCTGCAATCAACCCAAATC-3'] x *TULP4* Exon2 Rev [5'-AGTTTCTGGTAGGGCTCATTCC - 3']. DNA from SF188 and normal control DNA (Promega, Southampton, UK) were used. PCR was carried out with Platinum® Taq DNA Polymerase High Fidelity (Life Technologies, Paisley, UK) in a 25 μ l volume and with 100 nanograms of genomic DNA or cDNA as template. The following program was used: Initial denaturation at 94°C for 3min, followed by a 2-Step-Touchdown: 1. (94°C 30 sec, 68°C 45 sec, 72°C 1 min) 18 cycles, 2. (94°C 30 sec, 50°C 45 sec, 72°C 1 min) 30 cycles; followed by an additional cycle of 72°C 10 min. The fusion PCR product was cleaned up with the Illustra™ ExoStar™ 1-STEP kit (VWR, Lutterworth, UK). Sanger DNA sequencing of the PCR products was performed by DNA Sequencing & Services (MRCPPU, College of Life Sciences, University of Dundee, Scotland, www.dnaseq.co.uk) using Applied Biosystems Big-Dye Ver 3.1 chemistry on an Applied Biosystems model 3730 automated capillary DNA sequencer.

6.3. RESULTS

Analysis of the whole genome sequence and DNA copy number profile of the pHGG cell line SF188 identified a novel fusion preserving the 5' end of *RPTOR* and the 3' end portion of *TULP4* (Figure 6.2A). This rearrangement results from an inter-chromosomal translocation involving chromosome 6q25 and 17q25, with only one of the alleles of each gene being disrupted (Figure 6.2B). The genomic breakpoint junction shows blunt ends, contrary to the majority of the other fusions described in the previous chapter, which revealed microhomology at the breakpoint site (Chapter 5 - Figure 5.1). The only double-strand break repair pathway that can join DNA ends with no homology at the repair site is the classical NHEJ [20]. Using two different FISH strategies we were able to validate in SF188 the *RPTOR:TULP4* fusion (Figure 6.3B) and also the disrupted *RPTOR* gene (Figure 6.3C).

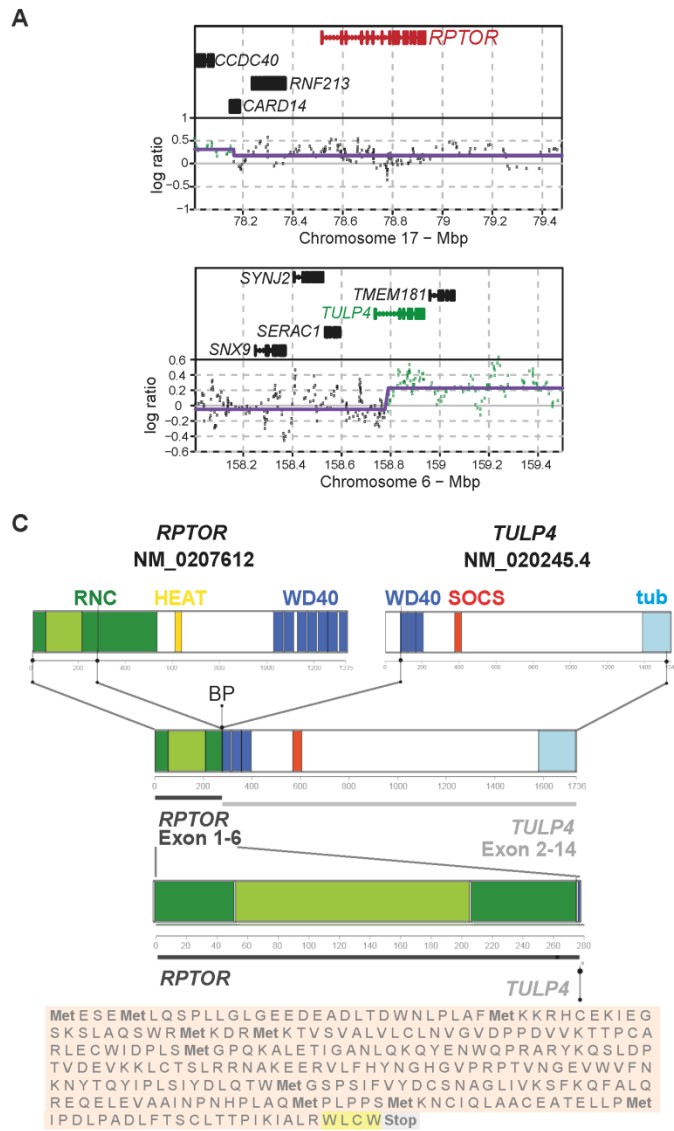


Figure 6.2. Identification of *RPTOR:TULP4* gene fusion in SF188 paediatric high grade glioma cell line. **A)** SF188 500K Affymetrix aCGH copy number profile of *RPTOR* and *TULP4*. **B)** PCR and RT-PCR detection of *RPTOR:TULP4* in SF188 genomic DNA and cDNA. Partial sequence of the *RPTOR:TULP4* fusion transcript identified in SF188. **C)** Schematic representation of *RPTOR* and *TULP4* genes as well as the formation of the fusion transcript involving exons 1-6 of *RPTOR* and exons 2-14 of *TULP4*, that leads to a truncation of *RPTOR* containing exon 1-6 plus four amino acids of *TULP4*. The translated amino acid sequence is shown below the cartoon, with *RPTOR* in pink. *TULP4* in yellow and the STOP codon in blue. **D)** Expression of *RPTOR* and *TULP4* exons in SF188 vs KNS42 and UW479 cell lines. *RPTOR*: regulatory associated protein of mTOR; *TULP4*: tubby like protein 4; RT: *RPTOR:TULP4*; RNC: *RPTOR* N-terminal conserved domain; HEAT: Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and the yeast kinase TOR1 repeats domain; WD40: ~40 amino acid motifs, often terminating in a Trp-Asp (W-D) dipeptide; SOCS: suppressor of cytokines signalling box; tub: tubby proteins domain.

Sanger sequencing of the RT-PCR product confirmed the fusion of *RPTOR* exon 6 with exon 2 of *TULP4* (Figure 6.2B), which is predicted to lead to the expression of a truncated *RPTOR* (Figure 6.2C). The truncated form of *RPTOR* preserves the conserved N-terminal region and loses the domains that bind to mTOR (HEAT domain and WD40 repeats) (Figure 6.2C). When expression of the fusion gene was assessed by exon coverage from the RNA sequencing data, there was no exon-specific expression pattern restricted to SF188 (Figure 6.2D).

We next sought to screen for *RPTOR* rearrangements in our paediatric and adult HGG series using specific break-apart FISH probes (Figure 6.3A). We screened 11 adult and pHGG TMAs, of which only cores with 50 scorable cells were screened. The *RPTOR* FISH probe was also assessed in control tissue, and on the basis of non-specific signal patterns observed, a threshold of 10% positive cells was set for assessment of the glioma samples. For the first TMAs screened we used a break-apart probe containing only one BAC clone at either side of *RPTOR* (RP11-317F05 and RP11-819H19) in order to specifically detect the *RPTOR* breakpoint seen in SF188 (Figure 6.4). Using this probe we found an adult case (1/160), RMH5987 where 14% cells contained a disrupted *RPTOR* (Figure 6.4). We then decided to use a break-apart FISH probe containing two BAC clones at either side of the *RPTOR* breakpoint (RP11-317F05/RP11-334C17 and RP11-819H19/RP11-387L10) in order to cover a larger region of the gene in case the breakpoint is different from the one occurring in SF188, and also in order to have a stronger probe. (Figure 6.4A). We found 2/53 pHGG cases harboured a disrupted *RPTOR* (Figure 6.4).

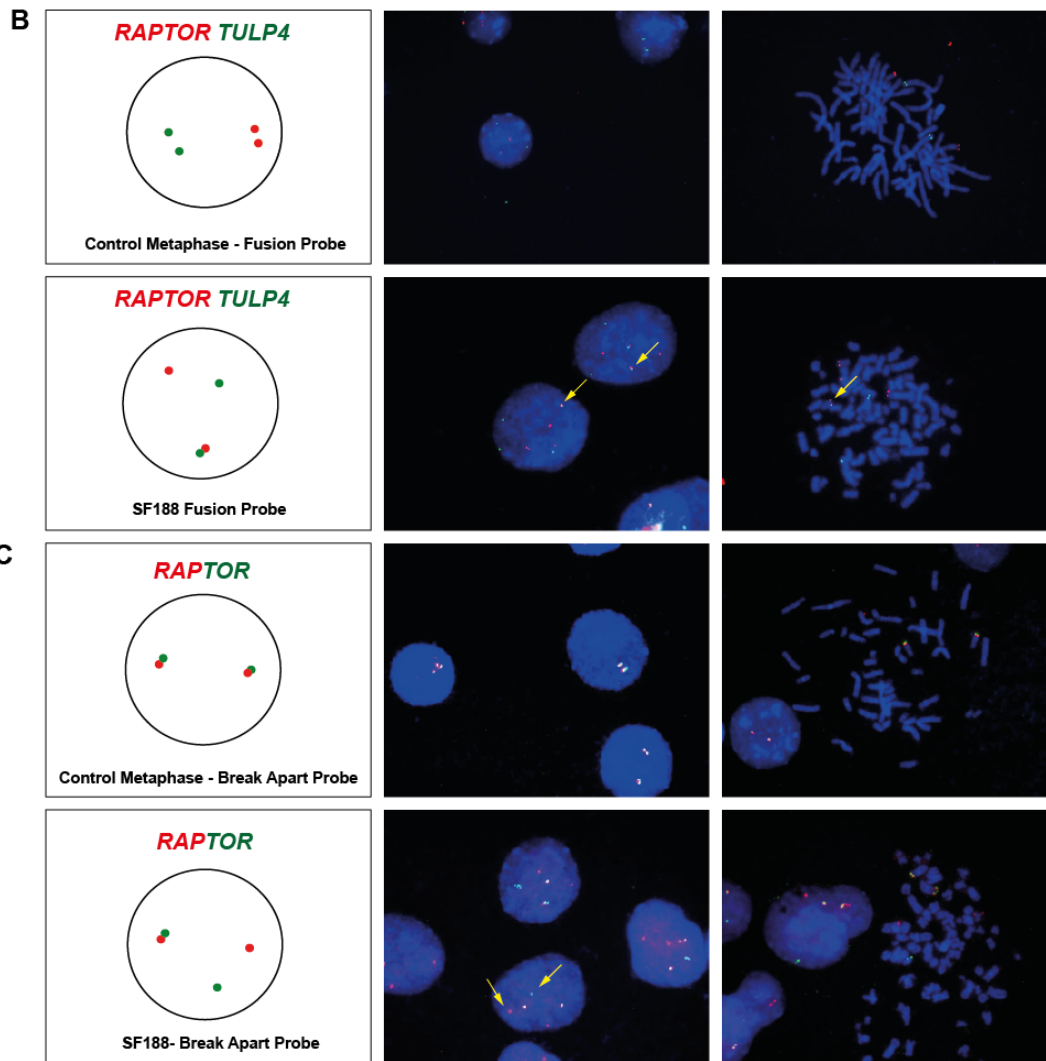
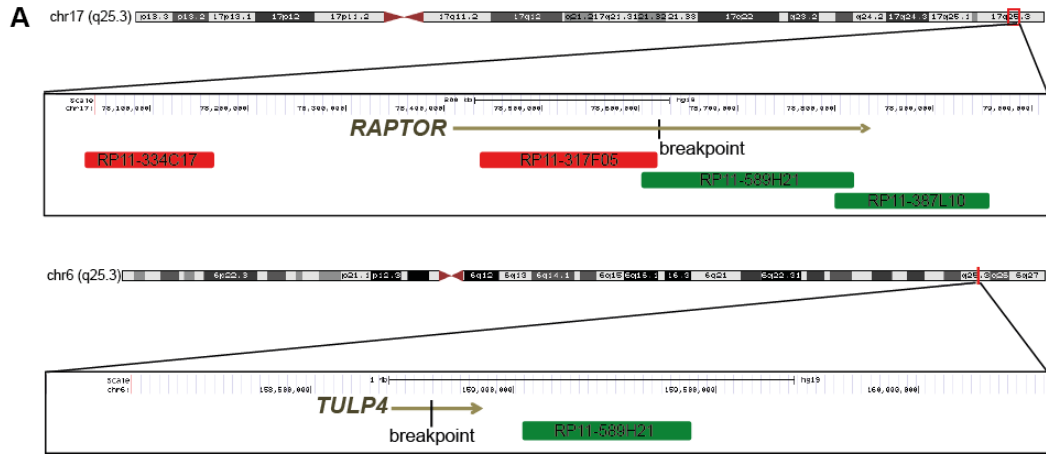


Figure 6.3. FISH of *RAPTOR:TULP4* rearrangement in SF188. A) Region specific BAC probes used for FISH. Probes were selected according to hg19 build of the human genome to overlap with the region of *RAPTOR* that is involved in the fusion (RP11-317F05; red bar) and to the region of the same gene that is not disrupted (RP11-819H19; green bar). A further probe covering the region of *TULP4* involved in the fusion was also selected (RP11-589H21; green bar). B) Cartoon of the FISH signals expected when using fusion probes for *RAPTOR* and *TULP4*, followed by interphase and metaphase FISH in control tissue and in the positive fusion cell line, SF188. The yellow arrows indicate the fusion in SF188. C) Representative cartoon of the expected signals when using a *RAPTOR* break-apart probe. FISH on interphase and metaphase cells of control tissue and SF188, showing split of the *RAPTOR* signal (yellow arrows) which indicates the rearrangement of this gene. Original magnification X1000.

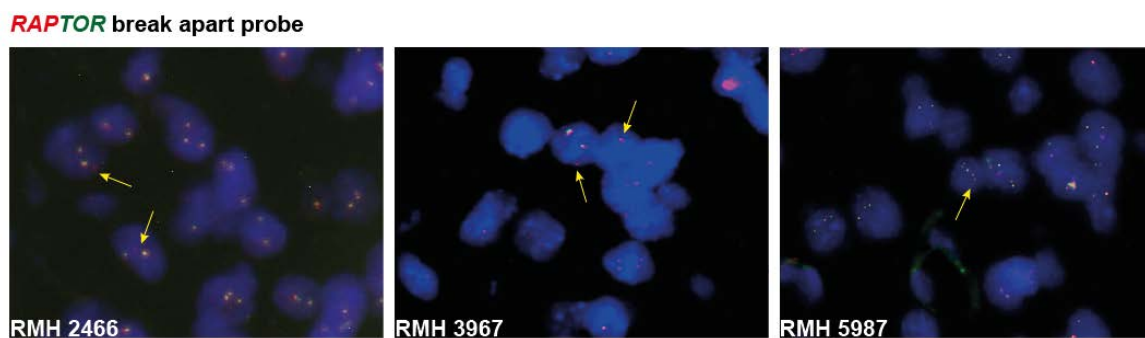


Figure 6.4. Representative FISH images showing disrupted *RAPTOR* in high grade glioma cases RMH2466, RMH3967 and RMH5987. The yellow arrows indicate the split of the *RAPTOR* signal which shows the rearrangement of this gene. A *RAPTOR* break-apart probe was used. Original magnification X1000.

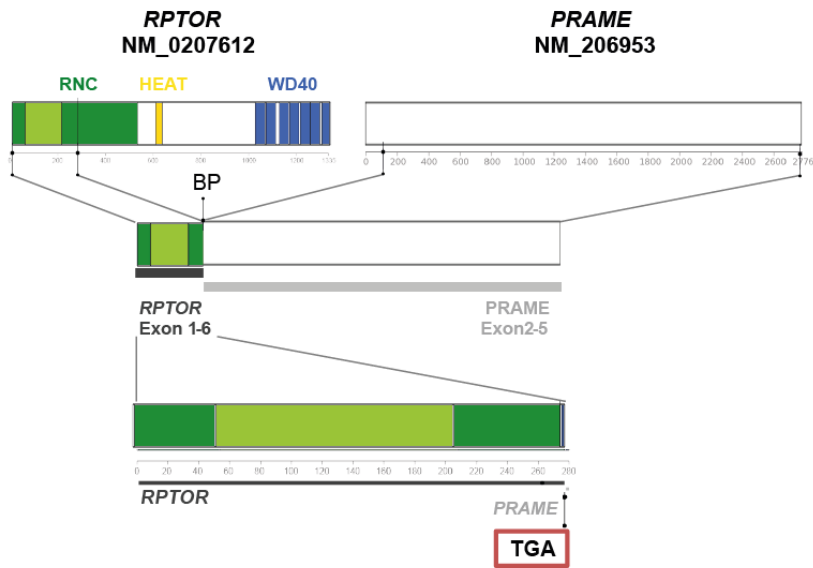
Two cores were available for each case so a mean of the percentage of fusion positive cells screened was calculated. RMH3967, an anaplastic astrocytoma of 9 years old female contained a disrupted *RPTOR* in 36.5% (31% and 42% in each core) of cells. In RMH2466, a GBM of a 10 years old male, the percentage was lower, with only 11% of cells positive (12% and 20% in each core).

We were next able to assess a series of pHGG samples for whole genome and RNA sequencing data was available [21], and identified a further 2/127 (1.6%) cases containing a rearranged *RPTOR* (Figure 6.5). In this study, structural variation in whole genome sequencing was analysed using CREST ('Clipping Reveals Structure'), an algorithm that uses next-generation sequencing reads with partial alignments to a reference genome, to directly map SVs at the nucleotide level of resolution [21, 22]. RNA sequencing SVs were

detected using CICERO ('Cicero Is Crest Extended for RNA Optimizations') (unpublished), a novel algorithm that utilizes *de novo* assembly to identify SVs [21]. SJHGG137, a GBM from a 6 year old boy, harbours a translocation that fuses the first 6 exons of *RPTOR* (1-277 amino acids) to the second exon of *PRAME* (preferentially expressed antigen in melanoma), an inter-chromosomal translocation involving chromosomes 17q25.3 and 22q11.22, respectively. The same region of *RPTOR* is present in the SF188 fusion (Figure 6.2C), and the net result is also predicted to be the truncation of *RPTOR*, with only 6 amino acids of *PRAME* being expressed (Figure 6.5). Secondly, case SJHGG102, a DIPG from a 5 year old girl, was found to harbour a truncation more proximal in the gene when compared to SF188 and SJHGG137 (Figure 6.5). Here, the first four exons of *RPTOR* are fused to an intergenic region located at chromosome 15q23.

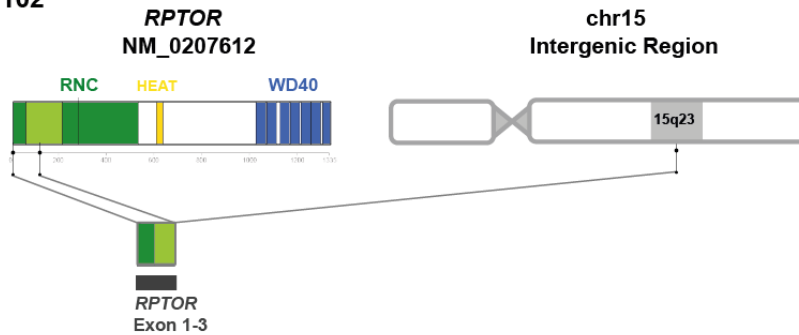
In total, we identified 4/180 (2.2%) paediatric and 1/160 (0.6%) adult HGG to harbour similar *RPTOR* truncations to that identified in SF188 cells.

SJHGG137



Met ESE Met LQSPLLGLGEEDEADLTDWNLPLAF Met KKRHCEKIEGSKSLAQSWR Met K
 DR Met KTVSVALVLCNLVGVDPDPPDVVKTTPCARLECWIDPLS Met GPQKALETIGANLQ
 KQYENWQPRARYKQSLDPTVDEVKKLCTSLRRNAKEERVLFHYNGHGVPRTVNGE
 VVVFNKNYTQYIPLSIYDLQTW Met GSPSIFVYDCSNAGLIVKSFKQFALQREQELEVA
 AINPNHPLAQ Met PLPPS Met KNCIQLAACATELLP Met IPDLPADLFTSCLTTPIKIALR
 CNFAVW Stop

SJHGG102



Met ESE Met LQSPLLGLGEEDEADLTDWNLPLAF Met KKRHCEKIEGSKSLAQSWR Met K
 DR Met KTVSVALVLCNLVGVDPDPPDVVKTTPCARLECWIDPLS Met GPQKALETIGANLQ
 KQYENWQPR

Figure 6.5. Schematic representations of the *RPTOR* rearrangements present in two paediatric high grade glioma cases. In SJHGG137 patient sample, the formation of the fusion transcript involves exons 1-6 of *RPTOR* and exons 2-5 of *PRAME* (Preferentially expressed antigen in melanoma), which leads to a truncation of *RPTOR* containing exon 1-6. In case SJHGG102, exons 1-3 of *RPTOR* are fused with an intergenic region of chromosome 15, that leads to the expression of a truncated form of *RPTOR*. The translated amino acid sequence is shown below the cartoon, with *RPTOR* in pink. *TULP4* in yellow and the STOP codon in blue.

6.4. Discussion

Using the sequencing strategy outlined in Chapter 5, we identified a novel fusion in SF188 cells involving the mTORC1 complex member RPTOR. Whole genome sequencing revealed a t(6;17) resulting in the generation of the novel expressed fusion gene *RPTOR:TULP4*. To search for the relevance of this fusion in patient samples, we utilised published whole genome and transcriptome sequencing data along with FISH screening of HGG TMAs to further identify 4/180 (2.2%) paediatric and 1/160 (0.6%) adult HGG to harbour similar *RPTOR* truncations to that identified in the reference cell line.

RPTOR encodes an important component of the mTORC1 complex which is essential for normal function and pathway activation. The discovery of *RPTOR* rearrangements is of particular interest, since oncogenic activation of the mTOR pathway has been described in different tumour types and also in HGG [2]. However, this usually results from point mutations rather than gene rearrangements, either in the mTOR complexes components, or upstream in the PI3K members which activate mTOR.

In HGG, *RPTOR* is a promiscuous fusion partner found fused with *TULP4*, *PRAME* and an intergenic region in chromosome 15. Due to this, rather than focus on finding the same fusions in our TMA series, we designed break-apart probes to identify a further 3 cases with *RPTOR* truncation instead, regardless of fusion partner. Though the fusion partners of the positive cases found by FISH are unknown, it is likely that they also have different partners, as the fusions with both *TULP4* and *PRAME* obtain a stop codon early in their sequence and are not expressed. Genes with different fusion partners are commonly seen in cancer, such as the recurrent fusion of *TMPRSS2* with ETS transcription factor genes in prostate cancer [23]. This promiscuity indicates that the critical gene in this rearrangement is likely to be *RPTOR*.

This structural rearrangement leads to the fusion of the N-terminal domain of *RPTOR* with the 3' end region of another genomic region, resulting in the expression of a shorter form of *RPTOR*. Do-Hyung Kim and colleagues expressed the first 300 amino acids of RPTOR (N-terminal domain) in HEK293 cells and observed that this isoform was unable to interact with endogenous mTOR [5], predicting for a loss of function, although activation of downstream targets were not evaluated.

Given the necessity for RPTOR in oncogenic signalling through the mTORC1 complex, it is not immediately apparent how a loss of RPTOR function would play a tumourigenic role. An alternative view of mTOR complex function and formation which is dependent on cell type

(reviewed in [24]) in the CNS has recently been proposed. *NF1* is a common tumour suppressor in glioma, which primarily functions as a negative regulator of the Ras proto-oncogene and when lost results in increased Ras activity due to the absent or decreased function of neurofibromin's GTPase-activating protein. Da Yong Lee and colleagues showed that *NF1* gene loss leads to increased neural stem cell proliferation and gliogenesis in the brainstem, but not in the cortex. RPTOR and RICTOR are responsible for this difference, as they are differently expressed in both regions of the brain. This results in unbalanced levels of mTOR/RPTOR (mTORC1) and mTOR/RICTOR (mTORC2) complexes, which lead to either a negative or positive regulation of Akt activity, depending on which complex is higher [25]. In the same way, the rearranged RPTOR in the HGG samples might yield a dominant-negative action in mTORC1 complex, contributing to unbalanced levels of the two mTOR complexes, therefore contributing to tumourigenesis through RICTOR-activated phospho-Akt signalling. Moreover, AKT pathway activation is seen in all paediatric cell lines SF188, KNS42 and UW479 [26]. Additionally, when compared with KNS42 and UW479, SF188 shows low levels of phosphorylated S6 [26]. This protein is a downstream target of the mTORC1 complex which contains RPTOR, therefore *RPTOR* truncation might contribute to this downregulation via mTORC1 and also to the upregulation of activated AKT via mTORC2. Interestingly, both the RMH2466 paediatric case and SF188 cell line harbour *NF1* loss.

Our results provide evidence for a low-frequency recurrent translocation involving *RPTOR*, such as those seen in adult GBM (*FGFR:TACC* (3-8%) and *EGFR:SEPT14* (3.2%)) and in pHGG (*NTRK* fusions (7%)) [21, 27-29]. The discovery of novel *RPTOR* rearrangements in 2.2% of pHGG shows that functionally important fusion events occur in this disease and suggest further structural characterization will likely identify additionally new recurrent rearrangements. More importantly, *RPTOR* SVs constitute an alternative way of activating PI3K/mTOR pathway, which is commonly deregulated in HGG through other mechanism [30-34]. The function of the *RPTOR* rearrangements is still unclear and other studies will be needed to determine whether and how the fusion interacts with the function of mTORC1 and its pathway. Nevertheless, this study generated new insights into the molecular characterization of pHGG, and highlighted a potential therapeutic target for this disease.

6.5. REFERENCES

1. Keith, C.T. and S.L. Schreiber, *PIK-related kinases: DNA repair, recombination, and cell cycle checkpoints*. Science, 1995. **270**(5233): p. 50-1.
2. Zoncu, R., A. Efeyan, and D.M. Sabatini, *mTOR: from growth signal integration to cancer, diabetes and ageing*. Nat Rev Mol Cell Biol, 2011. **12**(1): p. 21-35.
3. Shaw, R.J. and L.C. Cantley, *Ras, PI(3)K and mTOR signalling controls tumour cell growth*. Nature, 2006. **441**(7092): p. 424-30.
4. Hara, K., et al., *Raptor, a binding partner of target of rapamycin (TOR), mediates TOR action*. Cell, 2002. **110**(2): p. 177-89.
5. Kim, D.H., et al., *mTOR interacts with raptor to form a nutrient-sensitive complex that signals to the cell growth machinery*. Cell, 2002. **110**(2): p. 163-75.
6. Loewith, R., et al., *Two TOR complexes, only one of which is rapamycin sensitive, have distinct roles in cell growth control*. Mol Cell, 2002. **10**(3): p. 457-68.
7. Sarbassov, D.D., et al., *Rictor, a novel binding partner of mTOR, defines a rapamycin-insensitive and raptor-independent pathway that regulates the cytoskeleton*. Curr Biol, 2004. **14**(14): p. 1296-302.
8. Peterson, T.R., et al., *DEPTOR is an mTOR inhibitor frequently overexpressed in multiple myeloma cells and required for their survival*. Cell, 2009. **137**(5): p. 873-86.
9. Sancak, Y., et al., *PRAS40 is an insulin-regulated inhibitor of the mTORC1 protein kinase*. Mol Cell, 2007. **25**(6): p. 903-15.
10. Pearce, L.R., et al., *Identification of Protor as a novel Rictor-binding component of mTOR complex-2*. Biochem J, 2007. **405**(3): p. 513-22.
11. Frias, M.A., et al., *mSin1 is necessary for Akt/PKB phosphorylation, and its isoforms define three distinct mTORC2s*. Curr Biol, 2006. **16**(18): p. 1865-70.
12. Yang, Q., et al., *Identification of Sin1 as an essential TORC2 component required for complex formation and kinase activity*. Genes Dev, 2006. **20**(20): p. 2820-32.
13. Laplante, M. and D.M. Sabatini, *mTOR signaling at a glance*. J Cell Sci, 2009. **122**(Pt 20): p. 3589-94.
14. Richter, J.D. and N. Sonenberg, *Regulation of cap-dependent translation by eIF4E inhibitory proteins*. Nature, 2005. **433**(7025): p. 477-80.
15. Ma, X.M. and J. Blenis, *Molecular mechanisms of mTOR-mediated translational control*. Nat Rev Mol Cell Biol, 2009. **10**(5): p. 307-18.
16. Bove, J., M. Martinez-Vicente, and M. Vila, *Fighting neurodegeneration with rapamycin: mechanistic insights*. Nat Rev Neurosci, 2011. **12**(8): p. 437-52.
17. Meric-Bernstam, F. and A.M. Gonzalez-Angulo, *Targeting the mTOR signaling network for cancer therapy*. J Clin Oncol, 2009. **27**(13): p. 2278-87.

18. Koressaar, T. and M. Remm, *Enhancements and modifications of primer design program Primer3*. Bioinformatics, 2007. **23**(10): p. 1289-91.
19. Untergasser, A., et al., *Primer3--new capabilities and interfaces*. Nucleic Acids Res, 2012. **40**(15): p. e115.
20. Bunting, S.F. and A. Nussenzweig, *End-joining, translocations and cancer*. Nat Rev Cancer, 2013. **13**(7): p. 443-54.
21. Wu, G., et al., *The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma*. Nat Genet, 2014. **46**(5): p. 444-50.
22. Wang, J., et al., *CREST maps somatic structural variation in cancer genomes with base-pair resolution*. Nat Methods, 2011. **8**(8): p. 652-4.
23. Tomlins, S.A., et al., *Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer*. Science, 2005. **310**(5748): p. 644-8.
24. Weber, J.D. and D.H. Gutmann, *Deconvoluting mTOR biology*. Cell Cycle, 2012. **11**(2): p. 236-48.
25. Lee da, Y., et al., *Neurofibromatosis-1 regulates neuroglial progenitor proliferation and glial differentiation in a brain region-specific manner*. Genes Dev, 2010. **24**(20): p. 2317-29.
26. Bax, D.A., et al., *Molecular and phenotypic characterisation of paediatric glioma cell lines as models for preclinical drug development*. PLoS One, 2009. **4**(4): p. e5209.
27. Frattini, V., et al., *The integrated landscape of driver genomic alterations in glioblastoma*. Nat Genet, 2013. **45**(10): p. 1141-9.
28. Singh, D., et al., *Transforming fusions of FGFR and TACC genes in human glioblastoma*. Science, 2012. **337**(6099): p. 1231-5.
29. Parker, B.C., et al., *The tumorigenic FGFR3-TACC3 gene fusion escapes miR-99a regulation in glioblastoma*. J Clin Invest, 2013. **123**(2): p. 855-65.
30. Brennan, C.W., et al., *The somatic genomic landscape of glioblastoma*. Cell, 2013. **155**(2): p. 462-77.
31. Paugh, B.S., et al., *Integrated Molecular Genetic Profiling of Pediatric High-Grade Gliomas Reveals Key Differences With the Adult Disease*. J Clin Oncol, 2010.
32. Bax, D.A., et al., *A distinct spectrum of copy number aberrations in pediatric high-grade gliomas*. Clin Cancer Res, 2010. **16**(13): p. 3368-77.
33. TCGA, *Comprehensive genomic characterization defines human glioblastoma genes and core pathways*. Nature, 2008. **455**(7216): p. 1061-8.
34. Parsons, D.W., et al., *An integrated genomic analysis of human glioblastoma multiforme*. Science, 2008. **321**(5897): p. 1807-12.

7. Conclusions and future perspectives

7. CONCLUSIONS AND FUTURE PERSPECTIVES

Although childhood and adult HGG share related histopathological features, recent molecular profiling studies have clearly demonstrated that they are distinct biological diseases [1-6]. The work described throughout this thesis was set out to explore the molecular mechanisms underpinning pHGG, and to further delineate the distinguishing features of the adult and paediatric diseases, with a view to identify novel targets for therapeutic intervention. The presence of structural rearrangements and fusion genes was chosen to be investigated in pHGG, as almost nothing was known at the start of this work. We sought to characterize the presence and frequency of *PDGFRA* molecular alterations – mutations and structural rearrangements, we explored novel structural rearrangements associated with copy number alterations, and finally, we utilised whole genome sequencing to identify novel structural rearrangements that lead to the occurrence of fusion genes in childhood HGG.

7.1. *PDGFRA* alterations in paediatric high grade glioma

PDGFRA is a receptor tyrosine kinase that triggers essential cellular responses such as proliferation, migration, and survival [7]. This gene is commonly altered in cancer, by mutation [8, 9] and/or amplification [5, 6]. More importantly, it is the most frequent target of focal amplification and overexpression in pHGG, including DIPG, playing an important role in this disease [1, 3, 5, 6, 10, 11]. We thus explored the possibility that *PDGFRA* might be a preferential target of other mechanisms of genetic alteration in pHGG: somatic mutations and structural variation. In order to investigate this, we screened a large series of cases for single base changes and small indels, as well as the *PDGFRA* Δ 8,9 deletion and the *KP* fusion gene, previously reported in aHGG [12-14]. A summary of the main results described in this chapter are shown in Table 7.1. The results confirmed that childhood HGG harbour a distinct set of *PDGFRA* alterations when compared with their adult counterparts (*PDGFRA* Δ 89, *KP*), and this is in agreement with previous reports [13, 15-17].

In addition, our study revealed that *PDGFRA* is the target of several novel *in-frame* insertions and deletions, and mutations across all functional domains of the gene, contrary to what is observed in other cancers, such as GIST and NSCLC, where mutations occur in hotspot locations like the TK domain [8, 18]. Interestingly, in our studies *PDGFRA* alterations (mutations and *indels*) are more common in pHGG arising outside the brainstem (14% vs 6%), and 8/18 (44%) cases have concomitant amplification and mutation of *PDGFRA*. Recently, in a study by G. Wu *et al*, 16/112 (14.3%) pHGG analysed by next generation

sequencing harboured both *PDGFRA* alterations (amplification and/or mutation) and histone *H3.3* mutations. More importantly, 12/16 cases were associated with H3.3 (*H3F3A*) K27M mutations, 3/16 with H3.3 (*H3F3A*) G34R and only 1/16 with H3.1 (*HIST1H3B*) K27M [2], which is in accordance with unpublished data by J. Grill *et al.* In the same study, additional alterations in *PDGFRA* were discovered, small *indels* and also a novel translocation involving chromosomes 4 and 10, which leads to *DIP2C:PDGFRA* fusion gene in one of the DIPG cases [2]. In addition, A. Fontebasso *et al* study revealed that 7.1% (3/40) paediatric midline high grade astrocytoma (mHGA) carry mutations in *PDGFRA* tyrosine kinase domain (c.983A>G, c.1021G>A, c.1630_1631_insCAG), and the same cases also contain H3.3 K27M mutation. These studies confirmed the results developed in this thesis, that *PDGFRA* is commonly altered in pHGG, and that these changes occur in all functional domains of the gene.

Table 7.1. – Summary of the findings from the studies of *PDGFRA* in paediatric high grade glioma.

Tumour type	<i>PDGFRA</i> Alteration	<i>PDGFRA</i> Mutation	<i>PDGFRA</i> SV	Altered cases with <i>PDGFRA</i> Amplification
NB-pHGG	13/90 (14%)*	6/13 (46%)	7/13 (54%)	6/13 (46%)
DIPG	5/77 (6%)	4/5 (80%)	1/5 (20%)	2/5 (40%)
Total	18/167 (11%)	10/18 (56%)	8/18 (44%)	8/18 (44%)

NB-pHGG: non brainstem paediatric high grade glioma; DIPG: Diffuse Intrinsic Pontine Glioma; SV: Structural Variant; * 3 Anaplastic Astrocytomas, 1 Anaplastic Oligodendroglioma and 9 Grade IV Glioblastomas

In our study, we further explored the role of six *PDGFRA* mutants *in vitro* and *in vivo*. All the mutations taken for functional analysis were located at the extracellular domain of *PDGFRA*, and were found to be constitutively active and tumourigenic. The same phenomenon has been observed in other TK receptors, such as *EGFR* and *ERBB2*, where mutations in the extracellular domain besides being tumourigenic, are also cause resistance to TK small molecule inhibitors [19, 20]. In contrast to *PDGFRA* D842V mutated GIST tumours [21], pHGG *PDGFRA* alterations (V544ins, C450ins, E10del, E10del2 and E7del) did not confer resistance to dasatinib, a *PDGFRA* small molecule inhibitor.

Furthermore, the identification of the novel *PDGFRA* mutations contributed to the characterization of a subgroup of DIPG with oligodendroglial features, which is largely driven by both mutations and amplifications in *PDGFRA*. It is now known that the different gene expression profile is driven by mutations in the histone genes, *H3.3* and *H3.1* which are

unique to brainstem tumours [1, 3]. Given the clear importance of PDGFRA signalling on pHGG, it is unsurprising that there are multiple mechanisms active in driving tumourgenesis through a common pathway. Even though early-phase clinical trial observed thus far with PDGFRA inhibitors have been disappointing, the fact that PDGFRA represents one of the most frequently altered molecules in pHGG renders it an outstanding therapeutic target, not as single agent but as an addition other therapeutic approaches. Moreover, the molecular characterization of *PDGFRA* alterations presented in this thesis provide further evidence that *PDGFRA* mutated pHGG are characterized by altered RTK modifications.

7.2. Intragenic copy number aberrations and structural variants in paediatric high grade glioma

Comprehensive copy number profiling of adult and pHGG was among the first data to demonstrate the biological differences between these two similar diseases, but focuses only on amplification/deletion of genes in their entirety [5, 6, 22]. This method largely ignores the breakpoints that occur within the gene, which themselves may be functionally significant. Thus, we leveraged DNA copy number profiles to characterize the landscape of intragenic breakpoints in childhood malignant gliomas (summarized in chapter 4). This approach has been used before and it successfully identified a novel fusion in adult GBM, the *KDR:PDGFRA* fusion gene [13]. We analysed a large copy number dataset of 100 pHGG tumours, previously published by our group [6], and demonstrated the extent of iCNAs to be associated with age, type of histone mutation, and prognosis. Notably, in our study the presence of iCNAs was positively correlated with older paediatric patients, and younger patients (infant group >3 years) harboured less iCNAs. Other studies showed that children younger than 3 years of age, which have a paucity of chromosome instability and a lack of mutations, seem to be driven by a single potent driver oncogene recently discovered – *NTRK* fusion genes [2]. This fusion gene was not detected in our study, because it is a copy number silent event, highlighting one of the inherent limitations. Interestingly, a high number of iCNAs was found in tumours carrying the *H3F3A* K27M mutation. This mutation occurs in brainstem and other midline gliomas, and leads to a reduced trimethylation of the K27 residue (H3K27me3), consequently less heterochromatin is formed, which may lead to increased genomic instability [23, 24].

In our series of pHGG, we further confirmed genes that are known to be deleted in GBM (*NF1* and *Rb1*), but we also identified potential tumour suppressor genes in pHGG, such as *FAF1*, *PTPRE*, *PTPRD* and *CSMD3*. Most importantly, we discovered and validated a novel complex gene fusion that involved three genes, *DHX57:TMEM178:MAP4K3*. After the *PDGFRA* rearrangements, this was the first chimaera to be described in a pHGG. Although

no functional assays were performed, this fusion gene might play an important role in childhood malignant gliomas, since pHGG cell lines were highly sensitive to *MAP4K3* siRNA knockdown. We also identified a case that harboured a *CSGALNACT2:RET* fusion, although due to limited amount of material validation was not possible.

In order to further explore the occurrence of fusion genes in copy number altered loci, we carried out whole genome and transcriptome sequencing of pHGG cell line models. We used a similar approach to nominate novel SVs in pHGG by integrating algorithm predicted SVs (Breakdancer [25] and Chimera Scan [26]) identified by both WGS and whole transcriptome sequencing, with copy number data. We reasoned that the integration of these data would uncover novel pHGG fusion genes and would complement our study done in pHGG samples (chapter 4), as other fusion genes identified in brain tumours have been found to be associated with a copy number alteration, eg. *BRAF* rearrangements described in pilocytic astrocytomas [27], the *KDR:PDGFRA* [13] and the *FGFR:TACC* [28] seen in adult GBM. This work is described in chapter 5.

At the start of this thesis next generation sequencing had not been yet performed in pHGG samples and cell line models, and most of the data available was based on aCGH and expression arrays [5, 6, 29]. Our study showed for the first time that glioma cell lines KNS42, SF188 and UW479 are highly rearranged and that they are characterized by several fusion genes. We validated 59.4% of the algorithm-predicted SVs, using both PCR and FISH, which is similar to other studies where matched normal samples were also used [30]. Besides identifying novel fusion genes, we also discovered three extra-chromosomal structural rearrangement structures in SF188, which involve chromosomes 4q12 (*SCFD2*, *FIP1L1*), 8q24 (*MYC*), 11p11, 11q13 (*CCND1*), 11q14, 11q23 (*MLL*) and 12q14 (*CDK4*). The amplified loci represented two very complex structures, which were present in the form of three DMs. Although the exact arrangement of the genes in the DMs was not possible to reconstruct, based on the FISH and sequencing results we were able to draw a predicted structure.

We further characterized a fusion gene in each cell line: *GORASP2:CDADC1* in KNS42, *NUBPL:AKAP6* in UW479 and *RPTOR:TULP4* in SF188. Both *GORASP2* and *CDADC1* have not been reported in cancer studies, so their role in KNS42 is unclear. Nevertheless, *GORASP2* is involved in mitosis control [31], suggesting that this chimera might help cells enter mitosis, therefore promoting tumorigenesis. Although no clear role has been given to *NUBPL:AKAP6* fusion gene, *AKAP* family members have been directly associated with cancer development [32-34]. Recently, a study by Brett E. Johnson et al. on clonal evolution

in recurrent glioma showed that a late recurrence adult anaplastic astrocytoma harboured 39 mutations in *AKAP6*, connecting for the first time this gene to glioma [35].

The *RPTOR:TULP4* fusion gene described in SF188 involves *RPTOR*, an important component of mTOR signalling. This fusion leads to the expression of a truncated form of RPTOR in SF188. We discovered that *RPTOR* was disrupted 2.2% of pHGG cases (described in chapter 6) and that patient samples harbour similar truncations to the SF188 cell line. This finding suggests that structural variation is a different mechanism that pHGG use to alter the PI3K/mTOR pathway, which is commonly deregulated in glioma samples [36]. This has important implications for functional and preclinical studies, as it may underlie the high degree of sensitivity to PI3K/mTOR inhibitors seen with SF188 *in vitro*.

As the roles of these fusion genes are still unclear, further studies need to be performed in order to better understand their role in pHGG. Future experiments concerning *RPTOR* rearrangements should aim at the molecular mechanism of truncated *RPTOR* in mTORC1 complex. In order to do so, a truncated *RPTOR in vitro* cell model could be used to assess protein expression, both upstream and downstream of RPTOR, using both western blot and immunoprecipitation. In parallel, *RPTOR* siRNA knockdowns could be performed in SF188 in order to know if this fusion gene affects cell viability. Genetically engineered mouse models of RPTOR fusions would provide the ultimate validation of a key role in glioma development.

In summary, our results establish that pHGG harbour unique SVs, many of which may play important roles in the disease. In our study, we generated a detailed catalogue of genomic abnormalities occurring in pHGG and uncovered novel potential therapeutic targets.

7.3. References

1. Wu, G., et al., *Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas*. Nat Genet, 2012. **44**(3): p. 251-3.
2. Wu, G., et al., *The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma*. Nat Genet, 2014. **46**(5): p. 444-50.
3. Schwartzenuber, J., et al., *Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma*. Nature, 2012. **482**(7384): p. 226-31.
4. Bjerke, L., et al., *Histone H3.3 Mutations Drive Pediatric Glioblastoma through Upregulation of MYCN*. Cancer Discov, 2013.
5. Bax, D.A., et al., *A distinct spectrum of copy number aberrations in pediatric high-grade gliomas*. Clin Cancer Res, 2010. **16**(13): p. 3368-77.
6. Paugh, B.S., et al., *Integrated Molecular Genetic Profiling of Pediatric High-Grade Gliomas Reveals Key Differences With the Adult Disease*. J Clin Oncol, 2010.
7. Heldin, C.H. and B. Westermark, *Mechanism of action and in vivo role of platelet-derived growth factor*. Physiol Rev, 1999. **79**(4): p. 1283-316.
8. Burger, H., et al., *Activating mutations in c-KIT and PDGFRalpha are exclusively found in gastrointestinal stromal tumors and not in other tumors overexpressing these imatinib mesylate target genes*. Cancer Biol Ther, 2005. **4**(11): p. 1270-4.
9. Corless, C.L., et al., *PDGFRA mutations in gastrointestinal stromal tumors: frequency, spectrum and in vitro sensitivity to imatinib*. J Clin Oncol, 2005. **23**(23): p. 5357-64.
10. Paugh, B.S., et al., *Genome-wide analyses identify recurrent amplifications of receptor tyrosine kinases and cell-cycle regulatory genes in diffuse intrinsic pontine glioma*. J Clin Oncol, 2011. **29**(30): p. 3999-4006.
11. Zarghooni, M., et al., *Whole-genome profiling of pediatric diffuse intrinsic pontine gliomas highlights platelet-derived growth factor receptor alpha and poly (ADP-ribose) polymerase as potential therapeutic targets*. J Clin Oncol, 2010. **28**(8): p. 1337-44.
12. Kumabe, T., et al., *Amplification of alpha-platelet-derived growth factor receptor gene lacking an exon coding for a portion of the extracellular region in a primary brain tumor of glial origin*. Oncogene, 1992. **7**(4): p. 627-33.
13. Ozawa, T., et al., *PDGFRA gene rearrangements are frequent genetic events in PDGFRA-amplified glioblastomas*. Genes Dev, 2010. **24**(19): p. 2205-18.
14. Clarke, I.D. and P.B. Dirks, *A human brain tumor-derived PDGFR-alpha deletion mutant is transforming*. Oncogene, 2003. **22**(5): p. 722-33.
15. TCGA, *Comprehensive genomic characterization defines human glioblastoma genes and core pathways*. Nature, 2008. **455**(7216): p. 1061-8.
16. Parsons, D.W., et al., *An integrated genomic analysis of human glioblastoma multiforme*. Science, 2008. **321**(5897): p. 1807-12.

17. Verhaak, R.G., et al., *Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1*. *Cancer Cell*, 2010. **17**(1): p. 98-110.
18. McDermott, U., et al., *Ligand-dependent platelet-derived growth factor receptor (PDGFR)-alpha activation sensitizes rare lung cancer and sarcoma cells to PDGFR kinase inhibitors*. *Cancer Res*, 2009. **69**(9): p. 3937-46.
19. Greulich, H., et al., *Functional analysis of receptor tyrosine kinase mutations in lung cancer identifies oncogenic extracellular domain mutations of ERBB2*. *Proc Natl Acad Sci U S A*, 2012. **109**(36): p. 14476-81.
20. Lee, J.C., et al., *Epidermal growth factor receptor activation in glioblastoma through novel missense mutations in the extracellular domain*. *PLoS Med*, 2006. **3**(12): p. e485.
21. Heinrich, M.C., et al., *Crenolanib inhibits the drug-resistant PDGFRA D842V mutation associated with imatinib-resistant gastrointestinal stromal tumors*. *Clin Cancer Res*, 2012. **18**(16): p. 4375-84.
22. Jones, C., L. Perryman, and D. Hargrave, *Paediatric and adult malignant glioma: close relatives or distant cousins?* *Nat Rev Clin Oncol*, 2012. **9**(7): p. 400-13.
23. Lewis, P.W., et al., *Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma*. *Science*, 2013. **340**(6134): p. 857-61.
24. Chan, K.M., et al., *The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression*. *Genes Dev*, 2013. **27**(9): p. 985-90.
25. Chen, K., et al., *BreakDancer: an algorithm for high-resolution mapping of genomic structural variation*. *Nat Methods*, 2009. **6**(9): p. 677-81.
26. Iyer, M.K., A.M. Chinnaiyan, and C.A. Maher, *ChimeraScan: a tool for identifying chimeric transcription in sequencing data*. *Bioinformatics*, 2011. **27**(20): p. 2903-4.
27. Jones, D.T., et al., *Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas*. *Cancer Res*, 2008. **68**(21): p. 8673-7.
28. Singh, D., et al., *Transforming fusions of FGFR and TACC genes in human glioblastoma*. *Science*, 2012. **337**(6099): p. 1231-5.
29. Bax, D.A., et al., *Molecular and phenotypic characterisation of paediatric glioma cell lines as models for preclinical drug development*. *PLoS One*, 2009. **4**(4): p. e5209.
30. Berger, M.F., et al., *The genomic complexity of primary human prostate cancer*. *Nature*, 2011. **470**(7333): p. 214-20.
31. Corda, D., et al., *Golgi complex fragmentation in G2/M transition: An organelle-based cell-cycle checkpoint*. *IUBMB Life*, 2012. **64**(8): p. 661-70.
32. Sharma, S., et al., *A-kinase anchoring protein 3 messenger RNA expression correlates with poor prognosis in epithelial ovarian cancer*. *Gynecol Oncol*, 2005. **99**(1): p. 183-8.

33. Rudd, M.F., et al., *Variants in the GH-IGF axis confer susceptibility to lung cancer*. *Genome Res*, 2006. **16**(6): p. 693-701.
34. Lewis, T.E., et al., *Tissue transglutaminase interacts with protein kinase A anchor protein 13 in prostate cancer*. *Urol Oncol*, 2005. **23**(6): p. 407-12.
35. Johnson, B.E., et al., *Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma*. *Science*, 2014. **343**(6167): p. 189-93.
36. Brennan, C.W., et al., *The somatic genomic landscape of glioblastoma*. *Cell*, 2013. **155**(2): p. 462-77.

8. Appendix

8. APPENDIX

Appendix I.

Table S3.1. Clinical sample information for cases used for PDGFRA and RPTOR screens

Sample ID	Cohort	WHO grade	Diagnosis	Age at diagnosis (years)	PDGFRA amplification*	PDGFRA alteration
2HGG169T	Cohort 2	III	AA	8		
2HGG170T	Cohort 2	III	AA	8		
2HGG171T	Cohort 2	III	AA	12		N468S/E10del2
2HGG173T	Cohort 2	III	AA	15		
2HGG178T	Cohort 2	III	AA	13		
2HGG179T	Cohort 2	III	AA	17		
HGG024	Cohort 2	III	AA	7.5		
HGG079	Cohort 2	III	AA	22.75	Yes	Y288C
HGG082	Cohort 2	III	AA	5		
HGG097	Cohort 2	III	AA	9		
HGG156	Cohort 2	III	AA	13.2		
RMH2460	Cohort 2	III	AA	16.4		
RMH2480	Cohort 2	III	AA	15.8		
RMH3552	Cohort 2	III	AA	5.0		
RMH3558	Cohort 2	III	AA	10.0		
RMH3565	Cohort 2	III	AA	8.8		
RMH3568	Cohort 2	III	AA	17.7		
RMH3570	Cohort 2	III	AA	12.7		
RMH3943	Cohort 2	III	AA	10.9		
RMH3945	Cohort 2	III	AA	4.1		
RMH3946	Cohort 2	III	AA	24.9		
RMH3950	Cohort 2	III	AA	16.4		
RMH3951	Cohort 2	III	AA	14.1		
RMH3968	Cohort 2	III	AA	21.2		
HGG007	Cohort 2	III	AO	16.7	Yes	A491ins
RMH2447	Cohort 2	III	AO	13.8		
RMH3492	Cohort 2	III	AO	21.1		
RMH3500	Cohort 2	III	AO	9.0		C235R
RMH3942	Cohort 2	III	AO	11.3		
RMH3963	Cohort 2	III	AO	21.5		
RMH2444	Cohort 2	III	AOA	19.1		
2HGG172T	Cohort 2	IV	GB	3		

2HGG174T	Cohort 2	IV	GB	16	Yes	
2HGG175T	Cohort 2	IV	GBM	4		
2HGG176T	Cohort 2	IV	GBM	4		N659K
2HGG177T	Cohort 2	IV	GBM	0.8		
2HGG180T	Cohort 2	IV	GBM	9		
2HGG181T	Cohort 2	IV	GBM	3		
2HGG182T	Cohort 2	IV	GBM	1.4		
HGG001	Cohort 2	IV	GBM	5.8		
HGG006	Cohort 2	IV	GBM	6.3	Yes	
HGG008	Cohort 2	IV	GBM	16.5		
HGG010	Cohort 2	IV	GBM	17.8	Yes	
HGG014	Cohort 2	IV	GBM	7		C450ins
HGG019	Cohort 2	IV	GBM	14.4		E7del
HGG020	Cohort 2	IV	GBM	13.8		
HGG022	Cohort 2	IV	GBM	14.1		
HGG023	Cohort 2	IV	GBM	9.9		
HGG026	Cohort 2	IV	GBM	19.8		D842V
HGG028	Cohort 2	IV	GBM	19.8	Yes	N468S/E10del2
HGG086	Cohort 2	IV	GBM	14		
HGG093	Cohort 2	IV	GBM	11.3		
HGG152	Cohort 2	IV	GBM	22.7		
HGG154	Cohort 2	IV	GBM	3.8		
HGG157	Cohort 2	IV	GBM	14.8	Yes	E7del
HGG158	Cohort 2	IV	GBM	10.7		
HGG159	Cohort 2	IV	GBM	13.1	Yes	E10del
HGG162	Cohort 2	IV	GBM	13.5		
RMH2449	Cohort 2	IV	GBM	11.0		
RMH2450	Cohort 2	IV	GBM	0.5		
RMH2452	Cohort 2	IV	GBM	3.8		
RMH2453	Cohort 2	IV	GBM	11.9		
RMH2455	Cohort 2	IV	GBM	17.0		
RMH2456	Cohort 2	IV	GBM	13.3	Yes	
RMH2463	Cohort 2	IV	GBM	12.0		
RMH2464	Cohort 2	IV	GBM	9.2		
RMH2466	Cohort 2	IV	GBM	9.7		
RMH2467	Cohort 2	IV	GBM	10.3		
RMH2470	Cohort 2	IV	GBM	14.3		
RMH2471	Cohort 2	IV	GBM	12.4	Yes	
RMH2474	Cohort 2	IV	GBM	15.4		
RMH2475	Cohort 2	IV	GBM	15.2	Yes	

RMH2477	Cohort 2	IV	GBM	4.7		
RMH2478	Cohort 2	IV	GBM	11.5		
RMH2479	Cohort 2	IV	GBM	8.7		
RMH3491	Cohort 2	IV	GBM	9.8		
<i>RMH3498</i>	<i>Cohort 2</i>	<i>IV</i>	<i>GBM</i>	<i>12.5</i>		<i>C290R</i>
<i>RMH3499</i>	<i>Cohort 2</i>	<i>IV</i>	<i>GBM</i>	<i>11.5</i>	Yes	<i>E229K</i>
RMH3549	Cohort 2	IV	GBM	6.3		
RMH3550	Cohort 2	IV	GBM	10.0		
RMH3557	Cohort 2	IV	GBM	5.8		
RMH3944	Cohort 2	IV	GBM	10.4		
RMH3947	Cohort 2	IV	GBM	6.7		
RMH3949	Cohort 2	IV	GBM	5.5	Yes	
RMH3952	Cohort 2	IV	GBM	14.5	Yes	
RMH3953	Cohort 2	IV	GBM	21.4		
RMH3954	Cohort 2	IV	GBM	20.2	Yes	
RMH3955	Cohort 2	IV	GBM	19.8		
RMH3956	Cohort 2	IV	GBM	6.1		
RMH3960	Cohort 2	IV	GBM	11.3		
BSG001T	Cohort 2	IV	DIPG	11.8		
BSG003T(CB)	Cohort 2	IV	DIPG	7.9	Yes	
BSG004TD	Cohort 2	IV	DIPG	5.5		
BSG005T	Cohort 2	IV	DIPG	4.3		
BSG009T	Cohort 2	IV	DIPG	6.1	Yes	
BSG010T	Cohort 2	IV	DIPG	13.7		
BSG012T	Cohort 2	IV	DIPG	5.5		
BSG015T	Cohort 2	IV	DIPG	5.2		
BSG017T	Cohort 2	IV	DIPG	5.7		
BSG019T	Cohort 2	IV	DIPG	15.7	Yes	
BSG020T	Cohort 2	IV	DIPG	5.9		
BSG021T	Cohort 2	IV	DIPG	7.3		
<i>BSG022TD</i>	<i>Cohort 2</i>	<i>IV</i>	<i>DIPG</i>	<i>5.3</i>		<i>V544ins</i>
BSG023T	Cohort 2	IV	DIPG	3.6	Yes	
BSG024TD	Cohort 2	IV	DIPG	14.9		
BSG025T	Cohort 2	IV	DIPG	3.9	Yes	
BSG026T	Cohort 2	IV	DIPG	6.4		
BSG027TD	Cohort 2	IV	DIPG	5.2		
BSG029T	Cohort 2	IV	DIPG	5.2		
BSG030T	Cohort 2	IV	DIPG	12.3		
BSG031T	Cohort 2	IV	DIPG	5.2		
BSG032T	Cohort 2	IV	DIPG	10.9		

BSG033T	Cohort 2	IV	DIPG	3.7		
BSG034T	Cohort 2	IV	DIPG	8.9		
BSG035T	Cohort 2	IV	DIPG	13.2	Yes	
BSG037T	Cohort 2	IV	DIPG	15.7	Yes	
BSG038T	Cohort 2	IV	DIPG	6.9		
BSG039T	Cohort 2	IV	DIPG	6		
<i>BSG040T</i>	<i>Cohort 2</i>	<i>IV</i>	<i>DIPG</i>	12.5		<i>N659K</i>
BSG042T	Cohort 2	IV	DIPG	10.2		
BSG044T	Cohort 2	IV	DIPG	9	Yes	
BSG045T	Cohort 2	IV	DIPG	5.6	Yes	
BSG046T	Cohort 2	IV	DIPG	3		
BSG047T	Cohort 2	IV	DIPG	8.4	Yes	
BSG509T	Cohort 2	IV	DIPG	5.6	Yes	
BSG529T	Cohort 2	IV	DIPG	6.1		
BSG900TD	Cohort 2	IV	DIPG	6.4		
BSG901TD	Cohort 2	IV	DIPG	N/A		
BSG902TD	Cohort 2	IV	DIPG	9.8		
DIPG-WES001	Cohort 2	IV	DIPG	N/A		
DIPG-WES002	Cohort 2	IV	DIPG	N/A		
DIPG-WES003	Cohort 2	IV	DIPG	N/A		
HGG095	Cohort 2	IV	DIPG	5		
RMH 3560	Cohort 1	N/A	DIPG	6.9		
RMH 3964	Cohort 1	N/A	DIPG	4.3		
<i>RMH 3957</i>	<i>Cohort 1</i>	<i>N/A</i>	<i>DIPG</i>	4.8	Yes	<i>W449C</i>
RMH 6962	Cohort 1	N/A	DIPG	8.03		
RMH 6963	Cohort 1	N/A	DIPG	4.46		
RMH 6964	Cohort 1	N/A	DIPG	NA		
RMH 6965	Cohort 1	N/A	DIPG	NA	Yes	
RMH 6966	Cohort 1	N/A	DIPG	3.39		
RMH 6967	Cohort 1	N/A	DIPG	13.55		
RMH 6968	Cohort 1	N/A	DIPG	12.13		
RMH 6969	Cohort 1	N/A	DIPG	NA		
RMH 6970	Cohort 1	N/A	DIPG	NA		
RMH 6971	Cohort 1	N/A	DIPG	5.08		
RMH 6972	Cohort 1	N/A	DIPG	4.49		
RMH 6973	Cohort 1	N/A	DIPG	7.55		
<i>RMH 6974</i>	<i>Cohort 1</i>	<i>N/A</i>	<i>DIPG</i>	4.72	Yes	<i>T276P</i>
RMH 6975	Cohort 1	N/A	DIPG	6.26		
RMH 6976	Cohort 1	N/A	DIPG	4.75		
RMH 6977	Cohort 1	N/A	DIPG	9.41		

RMH 6978	Cohort 1	N/A	DIPG	6.37		
RMH 6979	Cohort 1	N/A	DIPG	NA		
RMH 6980	Cohort 1	N/A	DIPG	11.9		
RMH 6981	Cohort 1	N/A	DIPG	NA		
RMH 6982	Cohort 1	N/A	DIPG	5.27		
RMH 6983	Cohort 1	N/A	DIPG	10.22		
RMH 6984	Cohort 1	N/A	DIPG	6.76		
RMH 6985	Cohort 1	N/A	DIPG	13.49		
RMH 6986	Cohort 1	N/A	DIPG	7.41	Yes	
<i>RMH 6987</i>	<i>Cohort 1</i>	<i>N/A</i>	<i>DIPG</i>	<i>6.76</i>		<i>R479Q</i>
RMH 6988	Cohort 1	N/A	DIPG	NA		
RMH 6989	Cohort 1	N/A	DIPG	11.03		
RMH 6990	Cohort 1	N/A	DIPG	NA		
RMH 6991	Cohort 1	N/A	DIPG	NA		
RMH 6992	Cohort 1	N/A	DIPG	6.22		
RMH5684	adult cohort	IV	GBM	41.97	Yes	
<i>RMH5685</i>	<i>adult cohort</i>	<i>IV</i>	<i>GBM</i>	<i>76.49</i>	<i>Yes</i>	<i>PDGFRA del8,9</i>
RMH5686	adult cohort	IV	GBM	63.17	Yes	
RMH5687	adult cohort	III	AA	65.18	Yes	
RMH 5688	adult cohort	same case as above			Yes	
RMH 5689	adult cohort	same case as above			Yes	
RMH5690	adult cohort	IV	GBM	70.49	Yes	
RMH5696	adult cohort	IV	GBM	57.1	Yes	
RMH5697	adult cohort	IV	GBM	75.96	Yes	
RMH5698	adult cohort	same case as above			Yes	
RMH5701	adult cohort	IV	GBM	54.66	Yes	
RMH5707	adult cohort	III	AO	62.67	Yes	
RMH5714	adult cohort	IV	GBM	61.25	Yes	
RMH5716	adult cohort	IV	GBM	79.08	Yes	
RMH5724	adult cohort	IV	GBM	53.34	Yes	
RMH5725	adult cohort	IV	GBM	54.47	Yes	
RMH5966	adult cohort	IV	GBM	79.65	Yes	
RMH5978	adult cohort	IV	GBM	46.6	Yes	
RMH5981	adult cohort	III	AA	75.25	Yes	
RMH5982	adult cohort	IV	GBM	75.5	Yes	
RMH5984	adult cohort	IV	GBM	33.72	Yes	
RMH5986	adult cohort	III	AO	42.95	Yes	
RMH5989	adult cohort	IV	GBM	69.18	Yes	
RMH5991	adult cohort	IV	GBM	69.11	Yes	
RMH5998	adult cohort	III	AO	26.68	Yes	

RMH6004	adult cohort	IV	GBM	79.96	Yes
RMH6006	adult cohort	IV	GBM	38.15	Yes
RMH6010	adult cohort	IV	GBM	62.88	Yes
RMH6357	adult cohort	IV	GBM	59.71	Yes
RMH 6375	adult cohort	IV	GBM	55.79	Yes
RMH 6382	adult cohort	IV	GBM	61.47	Yes
RMH 6388	adult cohort	IV	GBM	55.91	Yes
RMH 6400	adult cohort	IV	GBM	49.87	Yes
RMH 6416	adult cohort	IV	GBM	73.68	Yes
RMH6417	adult cohort	IV	GBM	55.79	Yes
RMH6427	adult cohort	IV	GBM	62.79	Yes
RMH6433	adult cohort	IV	GBM	64.95	Yes
RMH6440	adult cohort	III	AO	40.24	Yes
RMH6620	adult cohort	IV	GBM	68.16	Yes
RMH6621	adult cohort	III	AO	45.16	Yes
RMH6625	adult cohort	III	AO	25.84	Yes
RMH6628	adult cohort	III	AO	46.29	Yes
RMH6634	adult cohort	III	AA	42.02	Yes
RMH6636	adult cohort	IV	GBM	65.22	Yes
RMH6641	adult cohort	III	AO	53.92	Yes
RMH6645	adult cohort	IV	GBM	52.15	Yes
RMH6646	adult cohort	IV	GBM	47.25	Yes
RMH6653	adult cohort	IV	GBM	50.3	Yes
RMH6665	adult cohort	IV	GBM	48.24	Yes
RMH6666	adult cohort	same case as above			Yes
RMH6667	adult cohort	IV	GBM	55.98	Yes
RMH6668	adult cohort	IV	GBM	55.96	Yes
<i>RMH6674</i>	<i>adult cohort</i>	<i>IV</i>	<i>GBM</i>	<i>78.13</i>	<i>Yes</i>
RMH6677	adult cohort	IV	GBM	48.88	Yes
RMH6685	adult cohort	IV	GBM	45.78	Yes
RMH6689	adult cohort	IV	GBM	51.85	Yes
RMH6862	adult cohort	IV	GBM	67.45	Yes
RMH6863	adult cohort	same case as above			Yes
RMH6872	adult cohort	III	AO	50.17	Yes
RMH6873	adult cohort	IV	GBM	65.5	Yes
RMH6874	adult cohort	III	AO	32.12	Yes
RMH6881	adult cohort	IV	GBM	44.6	Yes
RMH6887	adult cohort	IV	GBM	61.1	Yes
RMH6888	adult cohort	same case as above			Yes
RMH6889	adult cohort	IV	GBM	51.83	Yes

PDGFRA del8,9

RMH6891	adult cohort	III	AO	69.15	Yes
RMH6892	adult cohort	IV	GBM	76.04	Yes
RMH6893	adult cohort	IV	GBM	NA	Yes
RMH6905	adult cohort	IV	GBM	43.68	Yes
RMH6924	adult cohort	IV	GBM	52.02	Yes
RMH6925	adult cohort	IV	GBM	54.91	Yes
RMH6926	adult cohort	IV	GBM	76.07	Yes
RMH6927	adult cohort	IV	GBM	66.83	Yes
RMH6935	adult cohort	IV	GBM	45.36	Yes
RMH6936	adult cohort	same case as above			Yes

Sample ID	Cohort	WHO grade	Diagnosis	Age at diagnosis (years)	RPTOR Structural Variant
RMH2445	paeds TMA	III	OAD	19.1	
RMH2452	paeds TMA	IV	GBM	3.8	
RMH2459	paeds TMA	III	AA	16.4	
RMH2462	paeds TMA	same case as above			
RMH2465	paeds TMA	IV	GBM	13.9	
<i>RMH2466</i>	<i>paeds TMA</i>	<i>IV</i>	<i>GBM</i>	9.7	Yes
RMH2468	paeds TMA	IV	GBM	10.3	
RMH2469	paeds TMA	III	GC	14.5	
RMH2473	paeds TMA	IV	GBM	15.4	
RMH3490	paeds TMA	IV	GBM	11.3	
RMH3492	paeds TMA	II	AO	21.1	
RMH3947	paeds TMA	IV	GBM	6.7	
RMH3952	paeds TMA	IV	GBM	14.5	
RMH3954	paeds TMA	IV	GBM	20.2	
RMH3955	paeds TMA	IV	GBM	19.8	
RMH3957	paeds TMA	IV	DIPG	4.8	
RMH3962	paeds TMA	III	AA	0.4	
RMH3963	paeds TMA	III	AO	21.5	
RMH3965	paeds TMA	IV	GBM	12.5	
<i>RMH3967</i>	<i>paeds TMA</i>	<i>III</i>	<i>AA</i>	9	Yes
RMH4830	paeds TMA	IV	GBM	13	
RMH4833	paeds TMA	IV	GBM	18	
RMH4835	paeds TMA	IV	GBM	6	
RMH4839	paeds TMA	IV	GBM	16	
RMH4840	paeds TMA	IV	GBM	7	
RMH7062	paeds TMA	IV	GBM	13.9	
RMH7063	paeds TMA	III	PXA	13	
RMH7064	paeds TMA	III	AA	9	

RMH7068	paeds TMA	IV	GBM	19.3
RMH7069	paeds TMA	IV	GBM	20.3
RMH7070	paeds TMA	IV	GBM	1.3
RMH7072	paeds TMA	IV	GBM	12.9
RMH7074	paeds TMA	IV	GBM	10
RMH7077	paeds TMA	III	PXA	15
RMH7342	paeds TMA	III	AO	13.2
RMH7345	paeds TMA	III	AA	18
RMH7351	paeds TMA	IV	GBM	9
RMH7352	paeds TMA	IV	GBM	15
RMH7358	paeds TMA	III	PXA	18
RMH7359	paeds TMA	same case as above		
RMH7472	paeds TMA	N/A	N/A	N/A
RMH7479	paeds TMA	N/A	N/A	N/A
RMH7501	paeds TMA	N/A	N/A	N/A
RMH7502	paeds TMA	N/A	N/A	N/A
RMH7503	paeds TMA	N/A	N/A	N/A
RMH7505	paeds TMA	N/A	N/A	N/A
RMH7721	paeds TMA	N/A	N/A	N/A
RMH7767	paeds TMA	N/A	N/A	N/A
RMH7768	paeds TMA	N/A	N/A	N/A
RMH7771	paeds TMA	N/A	N/A	N/A
RMH5684	adult TMA	IV	GBM	69.23
RMH5687	adult TMA	III	AA	62.91
RMH5688	adult TMA	same case as above		
RMH5689	adult TMA	same case as above		
RMH5690	adult TMA	IV	GBM	47.67
RMH5691	adult TMA	IV	GBM	57.17
RMH5692	adult TMA	same case as above		
RMH5695	adult TMA	III	AO	60.28
RMH5697	adult TMA	IV	GBM	68.16
RMH5698	adult TMA	same case as above		
RMH5699	adult TMA	IV	GBM	45.16
RMH5700	adult TMA	same case as above		
RMH5701	adult TMA	IV	GBM	63.12
RMH5704	adult TMA	IV	GBM	72.65
RMH5706	adult TMA	IV	GBM	52.63
RMH5708	adult TMA	IV	GBM	66.46
RMH5709	adult TMA	IV	GBM	58.34
RMH5710	adult TMA	III	AA	56.76

RMH5713	adult TMA	IV	GBM	31.94	
RMH5714	adult TMA	IV	GBM	53.34	
RMH5715	adult TMA	IV	GBM	59.64	
RMH5716	adult TMA	IV	GBM	42.47	
RMH5718	adult TMA	IV	GBM	57.56	
RMH5719	adult TMA	III	AA	46.6	
RMH5720	adult TMA	IV	GBM	57.32	
RMH5722	adult TMA	same case as above			
RMH5723	adult TMA	IV	GBM	45.49	
RMH5724	adult TMA	IV	GBM	47.09	
RMH5966	adult TMA	IV	GBM	25.84	
RMH5967	adult TMA	IV	GBM	47.25	
RMH5968	adult TMA	same case as above			
RMH5969	adult TMA	IV	GBM	75.25	
TMH5970	adult TMA	IV	GBM	70.04	
RMH5972	adult TMA	IV	GBM	55.91	
RMH5973	adult TMA	IV	GBM	43.7	
RMH5974	adult TMA	IV	GBM	46.29	
RMH5976	adult TMA	IV	GBM	36.2	
RMH5978	adult TMA	IV	GBM	51.99	
RMH5980	adult TMA	IV	GBM	65.55	
RMH5981	adult TMA	III	AA	52.14	
RMH5982	adult TMA	IV	GBM	33.72	
RMH5983	adult TMA	IV	GBM	59.88	
RMH5984	adult TMA	IV	GBM	69.16	
RMH5985	adult TMA	IV	GBM	54.47	
RMH5986	adult TMA	III	AO	71.06	
<i>RMH5987</i>	<i>adult TMA</i>	<i>III</i>	<i>AO</i>	<i>56.31</i>	Yes
RMH5988	adult TMA	III	AA	61.49	
RMH5989	adult TMA	IV	GBM	72.68	
RMH5990	adult TMA	same case as above			
RMH5992	adult TMA	III	AA	42.09	
RMH5994	adult TMA	IV	GBM	67.45	
RMH5996	adult TMA	IV	GBM	65.22	
RMH5998	adult TMA	III	AO	42.95	
RMH5999	adult TMA	IV	GBM	60.18	
RMH6003	adult TMA	IV	GBM	42.97	
RMH6004	adult TMA	IV	GBM	51.86	
RMH6006	adult TMA	IV	GBM	66.06	
RMH6007	adult TMA	IV	GBM	49.13	

RMH6010	adult TMA	IV	GBM	70.45
RMH6011	adult TMA	IV	GBM	45.27
RMH6341	adult TMA	III	AO	58.34
RMH6343	adult TMA	IV	GBM	35.88
RMH6345	adult TMA	III	AO	66.63
RMH6346	adult TMA	III	AO	50.52
RMH6347	adult TMA	III	AO	30.57
RMH6354	adult TMA	IV	GBM	49.46
RMH6356	adult TMA	III	AO	31.38
RMH6362	adult TMA	III	AA	60.11
RMH6365	adult TMA	IV	GBM	69.11
RMH6370	adult TMA	IV	GBM	66.15
RMH6372	adult TMA	IV	GBM	77.73
RMH6373	adult TMA	same case as above		
RMH6374	adult TMA	IV	GBM	58.62
RMH6375	adult TMA	IV	GBM	51.35
RMH6377	adult TMA	IV	GBM	60.13
RMH6379	adult TMA	III	AO	62.67
RMH6380	adult TMA	IV	GBM	39.01
RMH6381	adult TMA	III	AO	63.09
RMH6384	adult TMA	IV	GBM	62.65
RMH6385	adult TMA	IV	GBM	26.68
RMH6386	adult TMA	IV	GBM	70.82
RMH6387	adult TMA	IV	GBM	46.8
RMH6388	adult TMA	IV	GBM	68.75
RMH6390	adult TMA	III	AO	50.11
RMH6391	adult TMA	IV	GBM	40.53
RMH6393	adult TMA	IV	GBM	67.07
RMH6394	adult TMA	III	AO	70.4
RMH6395	adult TMA	IV	GBM	64.27
RMH6396	adult TMA	IV	GBM	65.7
RMH6397	adult TMA	IV	GBM	69.78
RMH6398	adult TMA	III	AO	80.83
RMH6399	adult TMA	III	AO	60.49
RMH6400	adult TMA	IV	GBM	45.83
RMH6401	adult TMA	IV	GBM	43.07
RMH6402	adult TMA	III	AO	50.17
RMH6403	adult TMA	IV	GBM	69.45
RMH6404	adult TMA	IV	GBM	82.42
RMH6405	adult TMA	IV	GBM	51.6

RMH6409	adult TMA	IV	GBM	56.99
RMH6411	adult TMA	IV	GBM	41.78
RMH6412	adult TMA	same case as above		
RMH6413	adult TMA	IV	GBM	65.95
RMH6414	adult TMA	IV	GBM	45.76
RMH6417	adult TMA	IV	GBM	36.04
RMH6418	adult TMA	III	AO	57.4
RMH6419	adult TMA	IV	GBM	32.12
RMH6420	adult TMA	IV	GBM	71.32
RMH6421	adult TMA	IV	GBM	56.86
RMH6424	adult TMA	IV	GBM	46.44
RMH6425	adult TMA	IV	GBM	67.56
RMH6427	adult TMA	IV	GBM	55.96
RMH6429	adult TMA	IV	GBM	71.91
RMH6433	adult TMA	IV	GBM	41.91
RMH6434	adult TMA	IV	GBM	59.89
RMH6435	adult TMA	IV	GBM	62.15
RMH6639	adult TMA	IV	GBM	70.92
RMH6641	adult TMA	same case as above		
RMH6650	adult TMA	IV	GBM	63.17
RMH6651	adult TMA	IV	GBM	69.19
RMH6652	adult TMA	same case as above		
RMH6653	adult TMA	IV	GBM	65.18
RMH6665	adult TMA	IV	GBM	51.15
RMH6666	adult TMA	same case as above		
RMH6676	adult TMA	III	AO	57.45
RMH6685	adult TMA	IV	GBM	57.1
RMH6687	adult TMA	IV	GBM	73.38
RMH6689	adult TMA	IV	GBM	55.2
RMH6691	adult TMA	same case as above		
RMH6703	adult TMA	III	AA	69.21
RMH6862	adult TMA	IV	GBM	28.66
RMH6866	adult TMA	IV	GBM	56.01
RMH6867	adult TMA	III	AO	47.38
RMH6871	adult TMA	III	AO	65.91
RMH6874	adult TMA	III	AO	69.32
RMH6877	adult TMA	IV	GBM	76.04
RMH6880	adult TMA	same case as above		
RMH6881	adult TMA	IV	GBM	51.81
RMH6884	adult TMA	IV	GBM	38.49

RMH6887	adult TMA	IV	GBM	37.02
RMH6891	adult TMA	III	AO	73.74
RMH6892	adult TMA	IV	GBM	37.78
RMH6895	adult TMA	IV	GBM	72.62
RMH6897	adult TMA	III	AO	41.84
RMH6898	adult TMA	III	AO	49.43
RMH6899	adult TMA	IV	GBM	38.42
RMH6900	adult TMA	IV	GBM	67.21
RMH6902	adult TMA	IV	GBM	37.33
RMH6905	adult TMA	IV	GBM	54.85
RMH6906	adult TMA	IV	GBM	43.68
RMH6909	adult TMA	III	AA	67.69
RMH6910	adult TMA	IV	GBM	46.97
RMH6913	adult TMA	III	AA	58.56
RMH6918	adult TMA	IV	GBM	55.88
RMH6926	adult TMA	IV	GBM	39.3
RMH6933	adult TMA	IV	GBM	66.83
RMH6934	adult TMA	III	AA	57.12
RMH6935	adult TMA	IV	GBM	60.87
RMH6939	adult TMA	III	AO	60.68
RMH6952	adult TMA	III	AO	N/A
RMH6959	adult TMA	IV	GBM	N/A

AA - non-brainstem pediatric Anaplastic Astrocytoma

GBM - non-brainstem glioblastoma

DIPG - diffuse intrinsic pontine glioma (glioblastoma)

GC - Gliomatosis cerebri

PXA - Pleomorphic anaplastic xanthoastrocytoma

N/A - data not available

*accessed by FISH or aCGH

Mesenchymal Transition and PDGFRA Amplification/Mutation Are Key Distinct Oncogenic Events in Pediatric Diffuse Intrinsic Pontine Gliomas

Stephanie Puget^{1,2,3}, Cathy Philippe^{2,3}, Dorine A. Bax³, Bastien Job⁴, Pascale Varlet⁵, Marie-Pierre Junier⁵, Felipe Andreiuolo², Dina Carvalho^{3,6,7}, Ricardo Reis⁵, Lea Guerrini-Rousseau², Thomas Roujeau¹, Philippe Dessen⁴, Catherine Richon⁸, Vladimir Lazar⁸, Gwenael Le Teuff⁹, Christian Sainte-Rose¹, Birgit Geoger^{2,10}, Gilles Vassal², Chris Jones³, Jacques Grill^{2,10*}

1 Department of Neurosurgery, Necker-Sick Children Hospital, University Paris V Descartes, Paris, France, **2** Unite Mixte de Recherche 8203 du Centre National de la Recherche Scientifique «Virology and Anticancer Therapeutics», Gustave Roussy Cancer Institute, University Paris XI, Villejuif, France, **3** Section of Pediatric Oncology, The Institute of Cancer Research/Royal Marsden Hospital, Sutton, Surrey, United Kingdom, **4** Formation de Recherche en Evolution 2939 du Centre National de la Recherche Scientifique, Integrated Research Cancer Institute in Villejuif, University Paris XI, Villejuif, France, **5** Team Gial Plasticity, Unite Mixte de Recherche 894 de l'Institut National de la Santé et de la Recherche Medicale and Department of Neuropathology, Sainte-Anne Hospital, University Paris V Descartes, Paris, France, **6** Life and Health Sciences Research Institute, University Do Minho, Braga, Portugal, **7** Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal, **8** Functional Genomics Unit, Gustave Roussy Cancer Institute, University Paris XI, Villejuif, France, **9** Department of Biostatistics and Epidemiology, Gustave Roussy Cancer Institute, University Paris XI, Villejuif, France, **10** Department of Pediatric and Adolescent Oncology, Gustave Roussy Cancer Institute, University Paris XI, Villejuif, France

Abstract

Diffuse intrinsic pontine glioma (DIPG) is one of the most frequent malignant pediatric brain tumor and its prognosis is universally fatal. No significant improvement has been made in last thirty years over the standard treatment with radiotherapy. To address the paucity of understanding of DIPGs, we have carried out integrated molecular profiling of a large series of samples obtained with stereotactic biopsy at diagnosis. While chromosomal imbalances did not distinguish DIPG and supratentorial tumors on CGH arrays, gene expression profiling revealed clear differences between them, with brainstem gliomas resembling midline/thalamic tumours, indicating a closely-related origin. Two distinct subgroups of DIPG were identified. The first subgroup displayed mesenchymal and pro-angiogenic characteristics, with stem cell markers enrichment consistent with the possibility to grow tumor stem cells from these biopsies. The other subgroup displayed oligodendroglial features, and appeared largely driven by PDGFRA, in particular through amplification and/or novel missense mutations in the extracellular domain. Patients in this later group had a significantly worse outcome with a hazard ratio for early deaths, ie before 10 months, 8 fold greater than the ones in the other subgroup ($p=0.041$, Cox regression model). The worse outcome of patients with the oligodendroglial type of tumors was confirmed on a series of 55 paraffin-embedded biopsy samples at diagnosis (median OS of 7.73 versus 12.37 months, $p=0.045$, log-rank test). Two distinct transcriptional subclasses of DIPG with specific genomic alterations can be defined at diagnosis by oligodendroglial differentiation or mesenchymal transition, respectively. Classifying these tumors by signal transduction pathway activation and by mutation in pathway member genes may be particularly valuable for the development of targeted therapies.

Citation: Puget S, Philippe C, Bax DA, Job B, Varlet P, et al. (2012) Mesenchymal Transition and PDGFRA Amplification/Mutation Are Key Distinct Oncogenic Events in Pediatric Diffuse Intrinsic Pontine Gliomas. PLoS ONE 7(2): e30313. doi:10.1371/journal.pone.0030313

Editor: Maciej S. Lesniak, The University of Chicago, United States of America

Received: July 1, 2011; **Accepted:** December 15, 2011; **Published:** February 28, 2012

Copyright: © 2012 Puget et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The sponsors of the study were the Canceropole Ile de France – Institut National de Cancer (INCa), the LEM-Recherche (Les Entreprises du Médicament), the association "L'Etoile de Martin" and the Association pour la Recherche en Neurochirurgie Pédiatrique (ARNP). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: grill@gr.fr

These authors contributed equally to this work.

Introduction

Brain tumors are the leading cause of cancer-related morbidity and mortality in children and adolescents, malignant gliomas carrying the worst prognosis among them [1]. Malignant gliomas that diffusely infiltrate the brainstem appear almost exclusively during childhood and adolescence and have a relatively homogeneous presentation and dismal prognosis. DIPG represent the biggest therapeutic challenge in pediatric neuro-oncology with a median survival of 9 months despite collaborative efforts to

improve treatment [2]. The vast majority of children succumb to their disease within 2 years of diagnosis. These tumors are unresectable and radiotherapy is the only treatment offering a significant but transient improvement. The addition of chemotherapy has not shown any benefit over the use of irradiation only [2,3]. The development of targeted therapies for DIPG has been hampered by the lack of knowledge of the biology of this devastating disease. Trials have been implemented so far based on the assumption that biologic properties of these brainstem gliomas of children are identical to cerebral high-grade gliomas of adults

[4,5]. Recent data suggest however that pediatric high-grade gliomas differ from their adult counterparts [6–9], and that there may be biological distinctions between childhood gliomas presenting in the brainstem compared with supratentorial ones [10].

Comprehensive genomic studies of a substantial number of DIPG at diagnosis have not yet been undertaken due to the lack of available tumor material. Indeed, diagnosis is usually based on the association of specific neurological signs, short clinical history with a typical radiological appearance on MRI [11]. A biopsy is not needed for diagnosis in most of the cases [12,13]. In addition, most of these lesions are infiltrating and grading according to the WHO classification does not correlate with outcome. Accordingly and despite the reported safety of the procedure [14], most of the neurosurgical teams limit the use of stereotactic biopsies to the lesions with unusual clinical or radiological characteristics. Therefore, only very limited data on true DIPG is available in the literature and confounded by the inclusion of autopsy or post-radiotherapy cases [10,15–18].

Recently, our group started to use stereotactic biopsies of DIPG to obtain both pathological confirmation and immunohistochemical assessment of some specific biomarkers before the inclusion of patients in trials of targeted agents [19–21]. In this study, we sought to comprehensively define genetic alterations in DIPG at diagnosis by performing genome-wide array CGH and gene expression studies from frozen samples obtained by stereotactic biopsies. This study is the first to comprehensively define the biological alterations of DIPG at diagnosis, allowing the discovery of novel therapeutic targets directed specifically at these poor prognosis brain neoplasms.

Results

DIPG Biopsy Material

Over the 5 years of the study, 61 patients underwent stereotactic biopsies taking from one to eight tumor samples (median 3) in the Neurosurgery Department of Necker Sick Children's Hospital in Paris. In most instances, one or two biopsies were used for histological diagnosis and immunohistochemistry (Figure S1A). The remaining biopsies were snap-frozen with cytological control smears directly in the operating room, and nucleic acids extracted from representative samples. A median of 3.325 microg of DNA (range 0.805 to 21.5 microg) and 2.332 microg of RNA (range 0.048 to 15.84 microg) could be extracted from the biopsies, resulting in a total of 32 and 23 patients with sufficient quality and quantity of DNA and RNA, respectively, for microarray analyses without any amplification step.

A second set of surgical samples from pediatric non-brainstem high-grade gliomas of various histologies with arrayCGH ($n = 34$) and gene expression ($n = 53$) data acquired simultaneously on the same platform was used for comparative studies. Age distribution at diagnosis was similar in DIPG and in HGG.

DIPG Differ from Supratentorial High-grade Gliomas but Co-segregate with a Subgroup of Midline/thalamic Tumors

We first performed array CGH on the 32 frozen biopsies of newly diagnosed DIPG, and compared the high resolution DNA copy number profiles with a series of 34 pediatric supratentorial high grade gliomas. Unsupervised hierarchical clustering of the DIPG samples using the Euclidian distance defined two distinct subgroups, the first characterized by gain of chromosome 1q, and the second by numerous copy number losses and structural rearrangements (Figure S1B). There were no associations between

array CGH subgroup and survival, age at onset, duration of symptoms before diagnosis, radiological characteristics or WHO grade according to the 2007 revision.

Amplifications at specific loci were detected by CGHarray for the oncogenes *HRAS* (5), *PDGFRA* (4), *PDGFB* (2), *CAVI/2* (2), *PTPRN2* (2), *KDM5A* (2), *ETS1* (1), *MYCN* (1), *WNT2* (1), *RAB31* (1). Deletions were detected for *PTEN* (1), *CDKN2A/B* (1) and *FAS* (1). The oncogene *H-RAS* was gained or amplified in 7/32 (22%) and the *TP53* tumor suppressor gene was lost in 7/32 (22%) of cases. Loss of *TP53* locus was the only single chromosomal imbalance associated with a poorer outcome ($p = 0.01$, log-rank test) (Figure S1C). On immunohistochemistry, p53 overexpression was seen in 15/27 (55%) cases. A comprehensive list of minimal common regions of imbalances with a frequency superior to 15% is provided in Table S1.

It was not possible to clearly delineate DIPG and supratentorial tumors on the basis of the copy number profiles, as exemplified by an unsupervised principal component analysis (PCA) generated using all 42332 quality control passing probes (Figure S1D). By contrast, a similar PCA analysis of gene expression profiling using all 15231 quality control passing gene probes demonstrated the clustering of the DIPG samples distinct from the majority of supratentorial high-grade gliomas, with the exception of some midline (thalamic) tumors (Figure 1A).

Supervised analysis using the 76 samples (23 DIPG and 53 HGG) was used to identify the genes most closely associated with pediatric high-grade gliomas arising in the brainstem versus supratentorially, and resulted in an expression signature comprising 712 genes ($p < 0.005$, Pearson correlation, Ward procedure) which could distinguish tumours based on location independent of WHO grade (Table S2). The corresponding heatmap showed that the GE profiles of midline tumors clustered in some cases with the ones of DIPG (Figure 1B). Figure 1C shows the distribution of the expression for transcription factors and neurogenesis regulators according to the three different locations. DIPG and supratentorial tumors could be distinguished by a different pattern of expression of specific homeobox and HLH genes. When analysing the expression levels of the major regulators of brainstem embryogenesis described in the literature, we observed a significant upregulation of *GAL3ST1*, *MAFB*, *OLIG2* and *HOXA2,3* and 4 in DIPG compared to supratentorial tumors (Figure S1E).

DIPG Comprise two Biological Subgroups with Distinct Survival and Pathological Characteristics

The unsupervised k-means algorithm was used to discover subgroups of DIPG based on their gene expression profiles. The most optimal Bayesian Information Criterion (BIC) value was obtained for the classification based on two clusters [22] (Figure S2A), as represented by the corresponding principal component analysis (Figure 2A). Supervised hierarchical clustering identified 643 genes differentially expressed between these two groups (False Discovery Rate (FDR) adjusted p -value < 0.01) (Table S3 and Figure 2B). The first group had a significantly worse survival, with 70% (9/13) of children succumbing to the disease before the median overall survival time of 10.6 months (range 2 to 25 months) of the entire cohort, whilst only 10% (1/10) of the patients in the second group did so (Figure 2C). Since the risk of death was not proportional over time in the two groups, we use a Cox model with an interaction between group and time. The hazard ratio for early deaths, ie before 10 months, was 0.122 for group 2 vs group 1 ($p = 0.041$). Significant association of the 2 GE groups was observed neither with age nor with the array CGH classification described above.

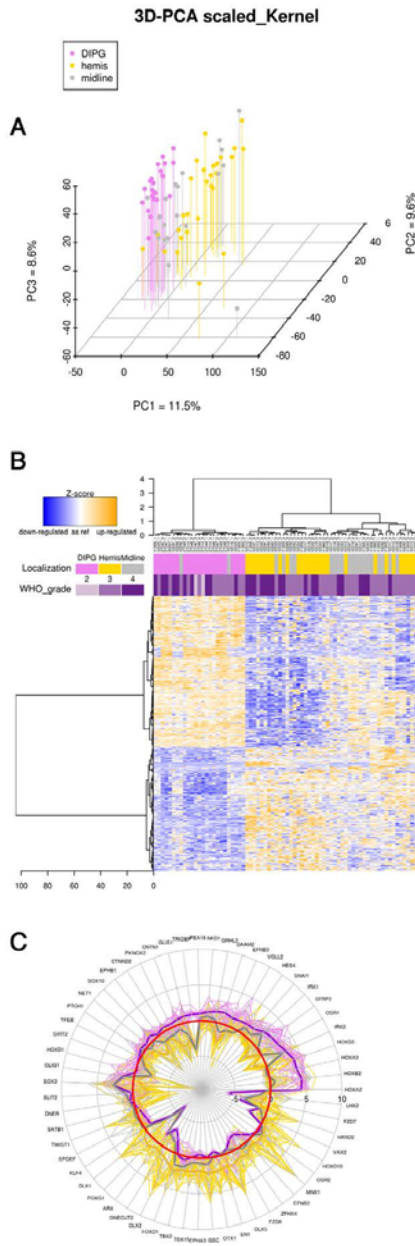


Figure 1. DIPG are different from supratentorial high-grade gliomas in children. Hemispheric, midline/thalamic tumors and DIPG are represented in gold, grey and violet respectively. Panel A: Gene expression of 23 DIPG and 49 supratentorial HGG were compared using a Principal Component Analysis on all 15231 quality control passing probes. Tumors are displayed according to their coordinates on the three first principal components, which describe 29.7% of the variance. Panel B: Heatmap of the 712 most differentially expressed genes between DIPG, midline and hemispheric tumors, selected using the moderated t-test of limma package of Bioconductor. Panel C: Radial plot of the expression of transcription factors and neurogenesis regulators according to the three tumor location, in log₂ ratios related to normal brain stem. The zero red line represent the expression level of normal adult brainstem. doi:10.1371/journal.pone.0030313.g001

Integrative analysis of the copy number and expression profiles using Spearman correlations demonstrated a significant influence of copy number on gene expression in group 1 tumours (306/15189 = 2% probes significantly correlated), however not for those in group 2 (3/15189 = 0.02% probes significantly correlated) (Figure 2D). These strong correlations were restricted to certain chromosomal abnormalities, in particular gain of 1q, loss of 19q, and amplification of 4q12. When considering both groups together the expression of 1460 genes (6% of the genome) was significantly correlated with their copy number; six of the twenty most correlated genes were located on chromosome 4q12 region: *CHIC2*, *SRP72*, *CLOCK*, *PPAT*, *SRD5A3* and *EXOC1* with Spearman correlation coefficient >0.9 and adjusted p<0.01 (Figure S2B).

Using gene set enrichment analysis [23], the expression profiles of the two groups were compared with the four subgroups of adult high grade gliomas recently described as proneural, neural, classical/proliferative and mesenchymal (http://tcga-data.nci.nih.gov/docs/publications/gbm_exp/) [24]. The proneural signature was highly enriched in the gene expression signature of group 1 (enrichment score = 0.66; nominal p = 0.004; FDR q = 0.089) (Figure 2E) while the mesenchymal signature was significantly associated with group 2 tumours (enrichment score = 0.8; nominal p = 0.004; FDR q = 0.007) (Figure 2F).

Mesenchymal Transition and a Pro-angiogenic Switch Define A Subset of DIPG

Since a mesenchymal gene expression signature was specifically represented in one of the two DIPG expression groups, we compared the expression of 53 transcription factors specific for this process as previously defined in adult high grade gliomas [25]. These genes were significantly upregulated in the group 2 DIPGs relative to the group 1 tumours (GSEA analysis: enrichment score 0.56, FDRq = 0.039, p nominal = 0.034), together with the master epithelial-mesenchymal transition regulators, *SNAIL1* and *SNAIL2/Slug* genes (Figure S3A). Expression of these genes alone was sufficient to distinguish group 1 and group 2 DIPG (Figure 3A). A subset of 7 transcription factors (*STAT3*, *BHLHE40*, *CEBPA* and *B*, *RUNX1*, *FOSL2* and *ZNF238*) controlled most genes of the mesenchymal signature of gliomas; all but *ZNF238* were significantly upregulated in the group 2 tumours compared to the other DIPG (Figure 3B). This transcriptional module was associated with a mesenchymal phenotype with upregulation of *TNC*, *OSMR*, *VIM* and *TKL40/CH3L1* and a more astrocytic histology (Table S3 & Figures 3A and 3C/D). Knowing that the *BRAF* V600E mutations could induce mesenchymal transition in some tumors [26] and that such mutations have been reported in a subset of pediatric glioma [27], we sequenced exon 15 of the *BRAF*

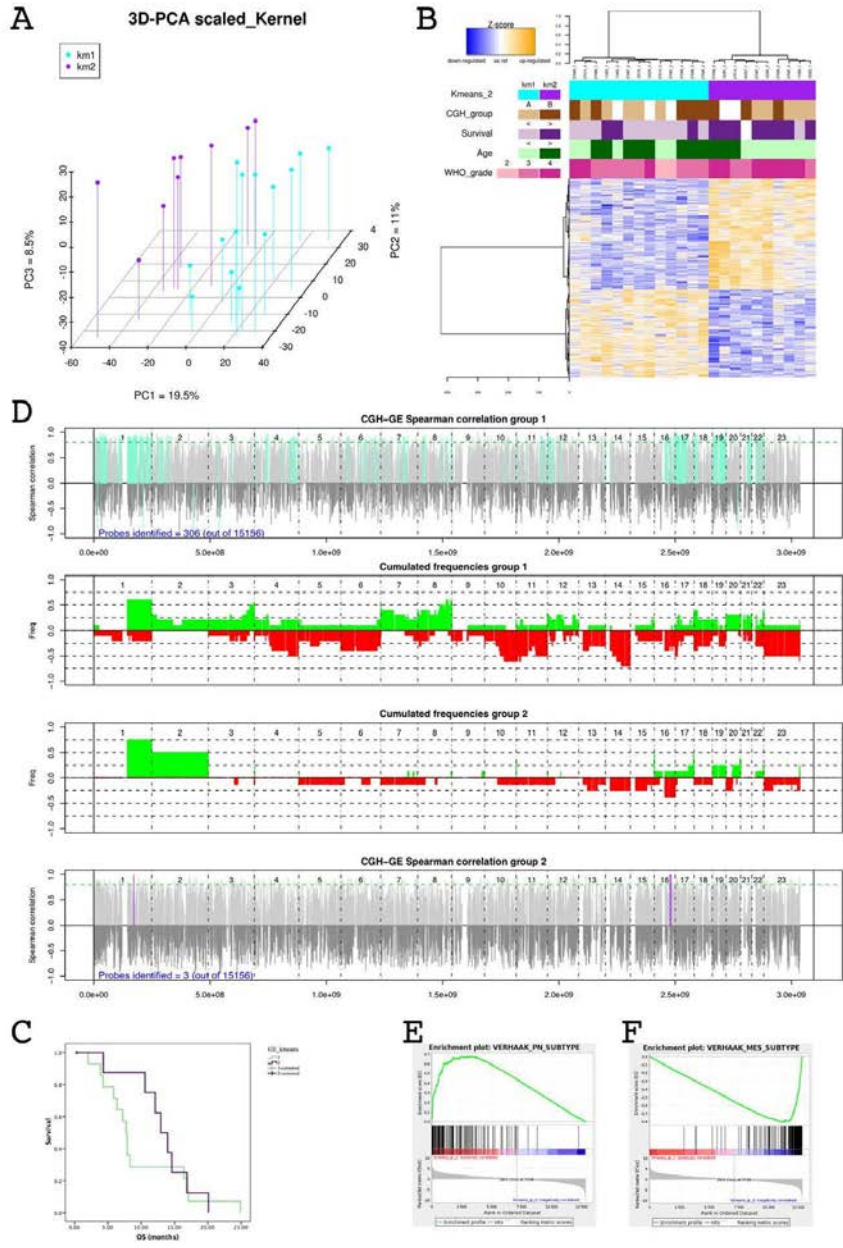


Figure 2. DIPG are divided into two groups with different gene expression signatures. Gene expression levels of 23 DIPG were analysed using an unsupervised procedure. Panel A: K-means algorithm followed by a model selection procedure using BIC defined two separated groups of DIPG that can be also clearly seen with a PCA on all probes that passed the quality control. Panel B: Heatmap of the 643 most differentially expressed genes between the two groups of DIPG, selected using the moderated t-test of limma package of Bioconductor. Panel C: Overall survival curves of the two groups of DIPG defining a group of patients who died early (70% of cases before the median survival time of 10.6 months, light green curve) and a group of patients who died later (90% of cases after the median survival time of 10.6 months, purple curve), ($p = 0.004$, chi-square test). Panel D: Integrated genomic analysis using DR-Integrator (R package) showing the correlation between probes of copy number and gene expression mapped on the same genomic coordinates (Refseq HG19) of a gene. In the upper panel (resp. the lower panel), colored vertical lines (cyan for group 1, purple for group 2) show probes for which copy number and expression were significantly correlated. The two panels in the middle show the CNA frequencies for the group 1 (resp. for the group 2). Most of the correlations between GE and CGH were found in group 1. Gene set enrichment analysis (GSEA) plot comparing group 1 GE profile to the signatures described for adult type gliomas. Group 1 gene expression profiles were enriched for proneural genes (Panel E) while group 2 gene expression profiles were enriched for mesenchymal genes (Panel F). doi:10.1371/journal.pone.0030313.g002

gene in 20 of the DIPG irrespective of their subgroup. No mutation was detected.

This mesenchymal phenotype was coupled with a hypoxia-induced angiogenic switch. Numerous proangiogenic genes were significantly overexpressed in this subgroup of DIPG compared to the other ones, including *VEGFA*, *VWF*, *PECAMI1*, *TREMI1*, *OSMR* and *PLAU* (Table S3 and Figure S3B). There was a strong correlation between *VEGFA* and *SNAI2/Slug* expression (Figure 3E), and between *VEGFA* and *YKL40* (Figure 3F) across the entire dataset, with a clear separation of the tumors in the two groups defined by the gene expression profiling. Endothelial proliferation was present in 8/9 mesenchymal group 2 tumours vs 8/14 in group 1 (89% vs 57%, $p = \text{NS}$, chi square test). On the extended cohort of 54 FFPE samples where endothelial proliferation could be evaluated, there was no correlation with survival, however an inverse correlation with Olig2 immunopositivity, a core biomarker of the proneural signature was noted ($p = 0.01$, chi square test). This angiogenic switch was associated with the activation of the *HIF1A* pathway as shown by the higher expression of *HIF1A* in group 2 ($p = 0.058$, Student t-test) and by the significant overexpression compared to group 1 of 5/8 of the hypoxia-related genes whose promoter is known to be highly responsive to *HIF1A*: *ENO2*, *HK1*, *HK2*, *LDHA*, *P4HA2* (Table S3).

This mesenchymal profile was further associated with a significant overexpression of numerous stem cell markers, including *BM11*, *CD34*, *CD44*, *CXCR4*, *LIF*, *DKK1*, *VIM* and *RUNX2*, in group 2 versus group 1 tumours (Figure S3C). Association of mesenchymal and stem cell markers was conserved in tumor cells with stem-like properties derived from three independent DIPG biopsies. These tumor stem-like cells yielded phenocopies of the original tumors in intracerebral xenografts (for complete description see [28]) and had a molecular profile as seen by qPCR similar to fetal neural stem cells with respect to stem cell markers (ie *SOX2*, *Musashi1*, *Nestin* and *FABP7/BLBP*) while overexpressing the mesenchymal markers *YKL40*, *SNAIL1* and *SNAIL2* compared to normal neural stem cells (Figure 3G). Of note, none of these tumor stem cells cultures, showed *PDGFRA* overexpression or amplification. The gene expression profile obtained from one of these DIPG models resembled mesenchymal subtype of DIPG as shown by unsupervised clustering using PCA (Figure S3D).

Oligodendroglial Differentiation and *PDGFRA* Amplification/mutation Define the Remaining Subset of DIPG

The group 1 of DIPGs as identified by gene expression profiling was characterized by the overexpression of oligodendroglial markers compared to group 2 (Figure 4A). Blinded morphological assessment revealed a significantly greater degree of oligodendroglial differentiation in these tumours compared with the mesenchymal group (Figure 4B and C). Strong expression of Olig2 by immunohistochemistry was seen in 13/13 tumors in this

group vs 3/8 in group 2 tumours (p value = 0.003, chi square test with McNemar correction) (Figure 4D and E). Of note, *SOX10*, a known transcription factor involved in oligodendroglial differentiation [29,30], was overexpressed in this subgroup compared to the other DIPG (\log_2 fold change 1.51 vs 0.21, adjusted p value = 0.0018). We used an extended cohort of 55 patients with histologically confirmed DIPG to study the impact of oligodendroglial differentiation on survival. Median overall survival of tumors with histological oligodendroglial features was 7.73 months versus 12.37 months for tumors that had predominantly astrocytic features ($p = 0.045$, log rank test) (Figure 4F).

The gene expression profile of group 1 DIPG was significantly enriched for the gene set describing the signature of *PDGFRA* amplified gliomas described in the TCGA [24] and in children [8] (GSEA analysis: enrichment score 0.59, FDRq = 0.038, p nominal = 0.052) (Figure 5A). Although *PDGFRA* was overexpressed in most of the tumors compared to normal brain, this overexpression was significantly stronger in the group 1 tumours ($p = 0.0055$) (Figure 4A). This overexpression was confirmed by immunohistochemistry on an independent cohort in 9/15 cases that were screened for the target-driven exploratory study of imatinib in children with solid malignancies [20] (Figure 5B & C). Eight of nine cases with gain/amplification of *PDGFRA* detected by arrayCGH were found in this subgroup; these imbalances were confirmed by FISH in six samples for which the analysis was possible (Figure 5D). Simultaneous amplification of *PDGFRA* and *MET* was observed in 4 samples (Figure 5E). A similar observation of co-amplification of two RTK was observed in one patient for *EGFR* and *PDGFRA* (Figure S4). The minimal common region of the *PDGFRA* amplicon also contained *LNX1*, *RPL21P44*, *CHIC2*, *GSK2*, *KIT* and *KDR*. Integration of copy number with gene expression data demonstrated a high degree of correlation only for *CHIC2*, *KIT*, *KDR* and *PDGFRA* only (Figure 5F).

Sequencing the *PDGFRA* gene in an extended series of DIPG samples revealed no mutations in the kinase domains, known hotspots in other tumors such as gastro-intestinal stroma tumors [31]. By contrast, novel missense mutations were observed in the extracellular domains in 3/34 (8.8%) cases, and in a further two high grade gliomas established as primary xenografts (Figure 5G). One of the mutations in the IGRG82 pediatric glioma xenograft has been previously described in an adult glioblastoma (C235Y) (http://tcga-data.nci.nih.gov/docs/publications/gbm_exp/). Both mutant-positive cases for which gene expression data was available were part of the group 1 DIPG, and harboured *PDGFRA* gene amplification, as did the additional case in the extended series.

DIPG subclasses signatures are enriched with genes of specific neural lineage

We conducted a GSEA to compare the GE profile of the two groups of DIPG to the gene list generated from 5 neural lineages

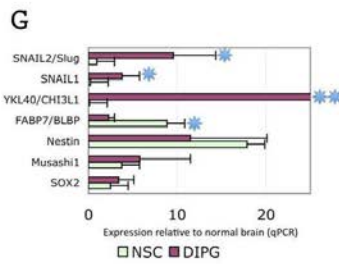
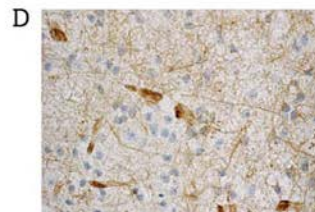
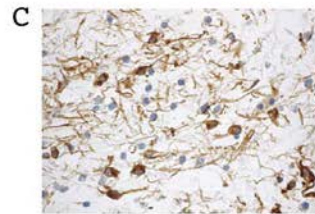
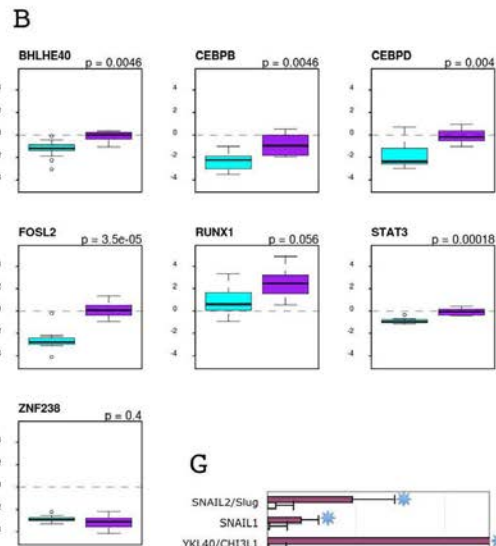
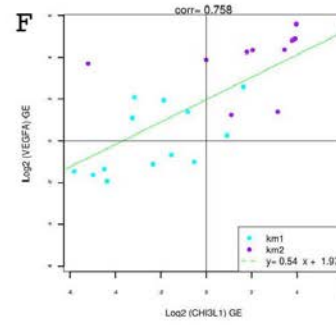
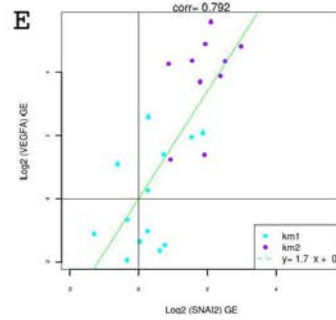
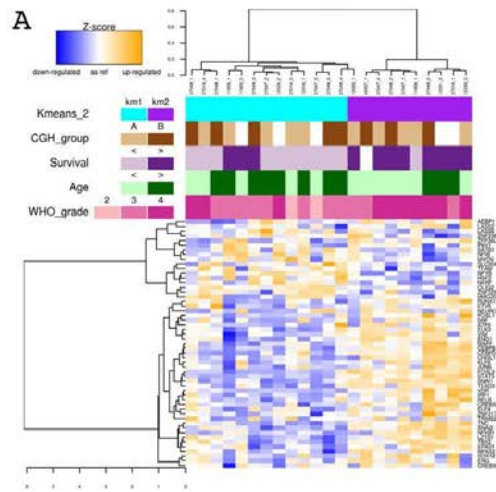


Figure 3. Description of the mesenchymal type of DIPG. DIPG from group 2 gene expression profile was enriched with genes involved in mesenchymal transition, angiogenesis and stem cell maintenance. Panel A: Heatmap of the transcription factors linked with mesenchymal gene expression signature (MGES) in adult glioblastomas. Biomarkers of mesenchymal phenotype (VIM, CH3L1 and TNC) and the two master regulators of epithelio-mesenchymal transition, SNAI1 and SNAI2/SLUG were added to the list provided by Carro et al [Carro et al., 2010]. Panel B: Boxplots comparing the 7 transcription factors driving the MGES in adult glioblastomas (Carro et al., 2010) in the two groups of DIPG (group 1 in cyan, group 2 in purple). Relative expression in log2 ratio compared to normal brainstem control is indicated. Vimentine immunohistochemistry in tumors of group 2 shows the positivity of tumors cells (Panel C) while in group 1 only vessels and reactive astrocytes were positive (Panel D). Panel E: Spearman correlation of the expression of SNAI2 and VEGFA. Group 1 tumors (cyan dots) segregate clearly from tumors of group 2 (purple dots). Panel F: Spearman correlation of the expression of CH3L1 and VEGFA. Group 1 tumors (cyan dots) segregate clearly from tumors of group 2 (purple dots). Panel G: Gene expression of stem cell and mesenchymal markers in DIPG tumorspheres derived from primary tumors of patients in stem cell medium as previously described (Thirant et al., 2011). Quantitative RT-PCR (qPCR) were performed using normal brain cortex as control. The spheroids cultured from three different DIPG were compared to normal neural stem cells (NSC) grown as neurospheres in the same medium. doi:10.1371/journal.pone.0030313.g003

isolated from mouse brain developed by Lei et al [32] from the transcriptome database of Cahoy et al [33]. Tumors of group 1 DIPG were enriched with the gene signature of mature oligodendrocytes and to a lower extent with the one of oligodendrocyte precursor cells (OPC), resembling in this respect to the proneural class of GBM glioblastoma (Figure 6). Conversely tumors of group 2 DIPG were enriched with the gene signatures of astrocytes and cultured astroglia (Figure 6). This later group of DIPG shared in this respect the GE signatures of the mesenchymal and classical classes of GBM that were enriched with the gene list of cultured astroglia and astrocytes, respectively.

Discussion

In this study, we report the first comprehensive genomic analysis of DIPG samples taken at diagnosis, and identify key biological features which distinguish them from other pediatric supratentorial HGG. The gene expression signatures associated with the location of a tumour was associated with differential reprogramming of embryonic signaling organizers, reflecting the discrete developmental origins of HGG presenting in different locations in the brain. Furthermore, our data indicate that DIPG arise from two distinct oncogenic pathways. The first group of DIPG exhibits an oligodendroglial phenotype associated with PDGFRA gain/amplification. Its gene expression profile is enriched for the proneural and PDGFRA-amplified glioma signatures. It comprises the most clinically aggressive tumours, independent of histological grade. The second group of DIPG exhibits a mesenchymal and pro-angiogenic phenotype orchestrated by a similar transcriptional module to that recently described in adult glioblastomas. These data greatly prolong our understanding of the molecular pathogenesis of pediatric DIPG and HGG, and have significant implications for the future clinical management of children with these tumours.

DIPG Represent a Biologically Distinct Group of HGG in Children

Pediatric DIPG and supratentorial high-grade gliomas, although harboring overlapping patterns of chromosomal imbalances, could be clearly differentiated through their gene expression signatures. Among the most differentially expressed genes with respect to tumour location, we identified numerous homeobox and HLH genes that were associated with brainstem tumours, and likely represent embryonic signaling organizers that have undergone transcriptional reprogramming during oncogenesis. The concept of location driving tumorigenesis in the brain [34] has been applied to other tumor types like ependymoma [35–38] and pilocytic astrocytomas [39], where developmentally-restricted gene expression signatures could be related to the site of tumor growth. Interestingly, genes found to be overexpressed in DIPG compared to supratentorial HGG, such as *LHX2* and *IRX2*, have been

previously described to be overexpressed in posterior fossa pilocytic astrocytomas and ependymomas compared to their supratentorial counterparts [36,38,39]. The converse may also be true, with *FOXF1* and *ZFX4* found to be upregulated in supratentorial HGG compared with DIPG, similar to data from ependymomas and pilocytic astrocytomas [36,38,39]. This suggests that there may be a common gene expression pattern related to the location and developmental origin of glial tumors irrespective of the histological diagnosis. Moreover, among the genes whose expression distinguished DIPG from the HGG in other location, we identified several genes involved in the SHh pathway such as *PTCH1*, *GLIS1*, *GJA1*, *SLC1A6*, *KCND2*, *PENK*, *GAD1* (see Table S2) already shown to be upregulated in mouse models [40]. This is in line with data from Monje et al. who have recently shown the possible role of the Sonic Hedgehog pathway in the oncogenesis of DIPG [41].

Of particular significance was the similarity of gene expression profiles of HGG arising in the midline/thalamus with DIPG, and their distinction from hemispheric tumours, likely indicating expansion from closely-related precursor populations, in these tumours for which the cell(s) of origin are yet not known. Although the adoption of different treatment strategies for DIPG and supratentorial HGG is well-established in clinical practice, the biological resemblance of midline/thalamic tumors and DIPG raises questions regarding the management of these specific neoplasms, currently focused on strategies designed for supratentorial HGG [42].

Mesenchymal transition with a stem cell-like phenotype is the hallmark of a subset of DIPG

While a mesenchymal phenotype appears only infrequently represented in pediatric supratentorial HGG [8], almost half of the pediatric DIPG were characterized by the overexpression of biomarkers of mesenchymal transition, stemness and a hypoxia-induced angiogenic switch. The transcriptional module driving the mesenchymal gene expression signature in adult glioblastoma [25] was also specifically overexpressed in this group compared to the proneural group. The acquisition of a mesenchymal phenotype [43], stemness [44], as well as the expression of hypoxia-related genes [45,46] have been associated with resistance to treatment including radiotherapy. The enhanced self-renewing capability of this subtype of DIPG further points to a distinct development lineage from the more differentiated PDGFRA-driven DIPG. In this respect, the higher expression of *STAT3* in the mesenchymal type of DIPG compared to the proneural one may play a key role in their opposite differentiation. Indeed, *STAT3* elimination promote neurogenesis and inhibit astrogenesis in neural stem cell, ie the phenotype of group 1 DIPG [47]. Glioma stem cells are associated with a perivascular niche, and appear to modulate vascular proliferation via VEGF, itself regulated via the HIF pathway. These three phenomena are closely interrelated in

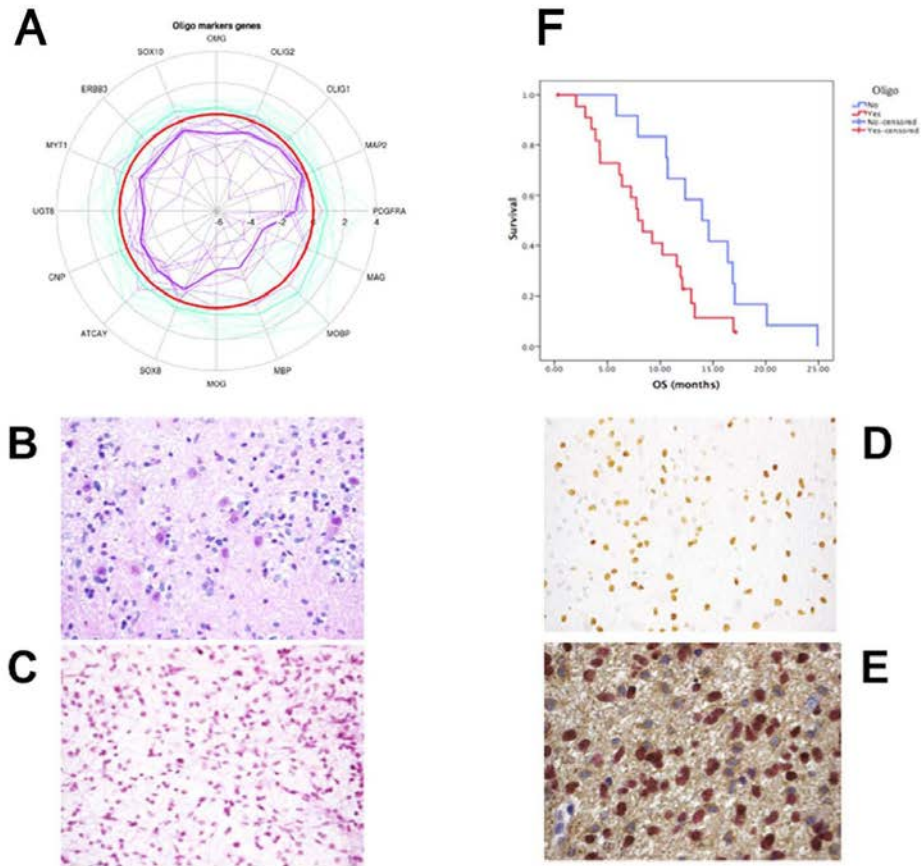


Figure 4. Description of the oligodendroglial/proneural type of DIPG. Panel A: Radial plot showing the expression of oligodendroglial markers in the two groups of DIPG, in log2 ratios related each other. The red circle represent the median expression level of the whole population of DIPG. Group 1 expresses higher levels of oligodendroglial markers than group 2 DIPG. Panel B: Morphological oligodendroglial differentiation in group 1 tumors (HES staining, ×40). Panel C: Morphological astrocytic differentiation in group 2 tumors (HES staining ×40). Panel D: Olig2 immunohistochemistry in a group 1 DIPG showing that probably not all cells in the biopsy are tumoral (×40). Panel E: Dual immunohistochemistry for Olig2 and GFAP showing that tumor cells in mitosis are GFAP negative but Olig2 positive (×100). Panel F: Overall survival of 55 DIPG according to the presence (red) or absence (blue) of oligodendroglial differentiation. Median OS was shorter in patients with oligodendroglial type of tumors (7.73 vs 12.37, $p=0.045$, log rank test). doi:10.1371/journal.pone.0030313.g004

several cancers including glioblastoma [48–51], and open the possibility that agents which target angiogenesis and/or drive differentiation of tumour stem cells may find application in this subset of DIPG to increase the antitumor effects of ionizing radiation.

Despite the involvement of Ras pathway in epithelio-mesenchymal transition via SNAIL2 [52] and its link with worse outcome of pediatric HGG [53], we did not find a correlation between *H-RAS* gain/amplification and its gene expression, nor activating

mutations in the *RAS* genes including *BRAF* V600E already described in some pediatric supratentorial gliomas [27], again highlighting differential oncogenic mechanisms in DIPG compared to other pediatric HGG.

Proneural and oligodendroglial differentiation associated with PDGFRA amplification

We have identified through unsupervised gene expression clustering a group of DIPG characterized by a 'proneural'

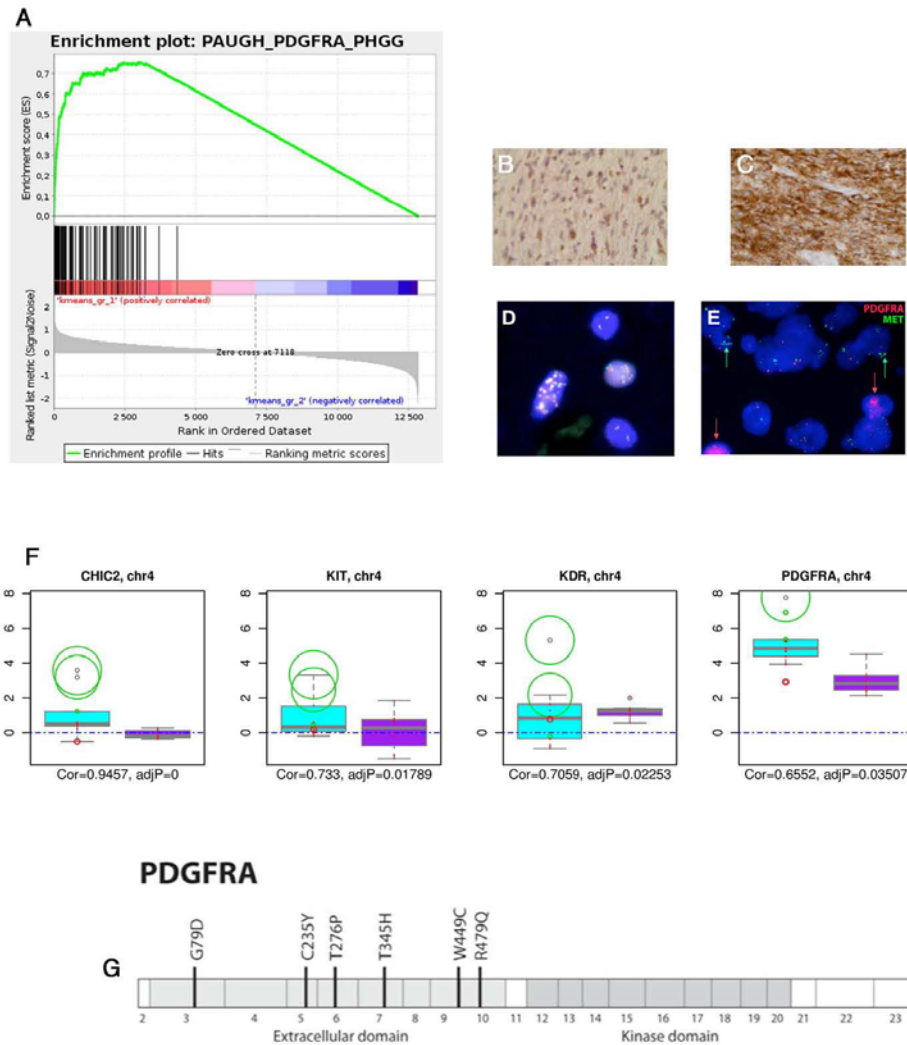


Figure 5. PDGFRA amplification/mutation is driving the oncogenesis of the oligodendroglial/proneural type of DIPG. Panel A: GSEA graph showing the enrichment of group 1 DIPG for the gene set describing the gene expression profile of PDGFRA amplified glioblastomas. Panel B: PDGFRA immunohistochemistry in the infiltrative part of a DIPG. Panel C: PDGFRA immunohistochemistry in the tumoral part of a DIPG. Panel D: FISH analysis of a DIPG using a FIP1L1/PDGFRA probe showing the amplification of the locus encompassing the two genes (most frequently seen). Panel E: Dual-FISH analysis of a DIPG with two probes one for PDGFRA and one for MET showing that the two oncogenes may be gained/amplified in different cells within the tumor. Panel F: Integrative genomic analysis using DR-Integrator (R package). Seven genes are present in the minimal common region (MCR) gained on chromosomal location 4q12 in DIPG. Boxplots represent the distribution of GE data and circles represent CNA data. The circles are centered on the corresponding GE measure on the distribution and their radii are proportional to the absolute value of CNA, red ones being losses and green ones gains. CNA and GE were highly correlated for four of these seven genes (*CHIC2*, *KIT*, *KDR*, *PDGFRA*). Panel G: Diagram of the PDGFRA gene showing the mutations discovered in DIPG samples and xenografts. doi:10.1371/journal.pone.0030313.g005

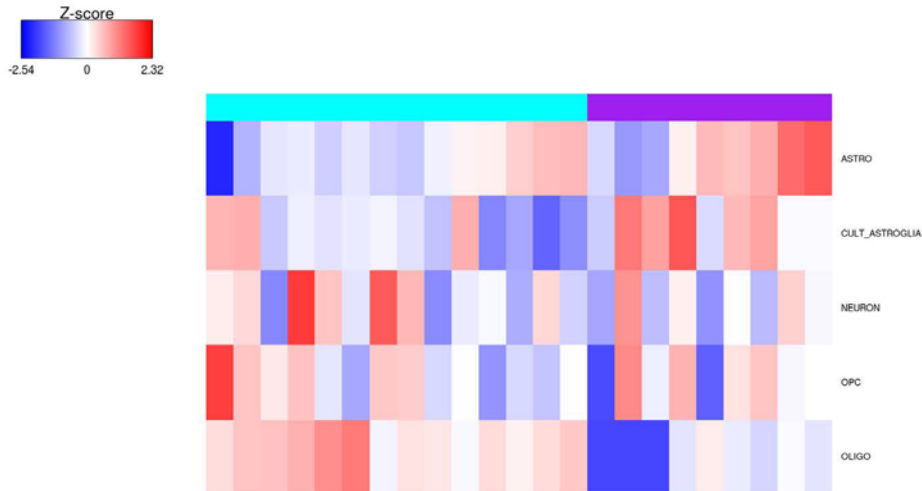


Figure 6. Comparison of gene expression signature of the two DIPG groups with specific neural lineages. A GSEA analysis was processed using the gene list previously described by Lei et al [32] and derived from the gene sets specifically enriched in astrocytes, oligodendrocytes, neurons, oligodendrocytes progenitors cells and cultured astroglial cells. Heatmap of the enrichment scores of each DIPG sample is represented with a red to blue color scale shows the range from the highest to lowest enrichment score.
doi:10.1371/journal.pone.0030313.g006

phenotype, an oligodendroglial differentiation, and *PDGFRA* amplifications/mutations. Moreover, the gene expression profile of group 1 DIPG was significantly enriched with genes describing the signature of *PDGFRA* amplified gliomas [8,24] supporting the hypothesis that *PDGFRA* amplification is associated with a robust gene expression profile across tumor location and patient's age. This association has been previously described in adult tumors [43,54–56], and include the expression of genes involved in neurogenesis and oligodendrocyte development, such as *Olig* transcription factors, *Mx2.2*, *PDGFRA* and *SOX10* [57]. DIPG with oligodendroglial phenotype and *Olig2* overexpression exhibited an even worse evolution and resistance to radiation than the other DIPG in our series. This could be explained by the recent findings that the central nervous system-restricted transcription factor *Olig2* opposes p53 response to genotoxic damage in neural progenitors and malignant glioma [58]. This is however in contrast with the adult gliomas where oligodendroglial differentiation and proneural phenotype are linked with a better prognosis [24]. Moreover, we did not observe *IDH1/2* mutation in 10 DIPG [59] while in adult proneural gliomas *IDH1* mutations are frequent [24]. In pediatric gliomas, *IDH1/2* mutations are almost exclusively seen in adolescents [59,60] who indeed do not represent the target population of DIPG. The presence of *IDH1* mutation in tumors from adolescents was not correlated with an oligodendroglial phenotype in our cohort of pHGG previously published [59]. Together with the fact that the group 1 DIPG is enriched preferentially with the signature of mature oligodendrocytes rather than oligodendrocyte progenitor cells, these data could suggest that this group of DIPG could be developed from a different oligodendroglial cell than their adult counterpart. This would be in line with the rarity of 1p19q co-deletion in pediatric gliomas with oligodendroglial features.

Integrative genomics showed that the gene expression of this group of DIPG was driven by copy number changes on the contrary to the other DIPG suggesting that chromosomal instability plays an important role in the phenotype of these tumors. Conversely, gene expression in the other group of DIPG may be more driven by epigenetic changes.

We found 28% (9/32) of *PDGFRA* gains or amplifications, all but one being included in the group 1 defined by unsupervised gene expression clustering. The PDGF autocrine/paracrine loop has been frequently implicated in oligodendroglomas [61] and has been used to create preclinical models of glioma [62,63], including brainstem tumors [64,65]. *PDGFRA* amplification has been shown to be more frequent in pediatric HGG than in adult ones [8] and a recent report found *PDGFRA* gain or amplification in four out of eleven post-mortem samples of DIPG [10]. In one of our previous study, *PDGFRA* protein was also more frequently detected by IHC in DIPG than in other pediatric HGG [20].

We identified 10% of pediatric DIPG to harbor *PDGFRA* missense mutations, considerably more frequently than the 2/206 (1%) reported in adult GBM (http://tcga-data.nci.nih.gov/docs/publications/gbm_exp/). These mutations were located in exons coding for the extracellular domains of the protein, potentially disrupting ligand interaction, but not in the tyrosine-kinase domain. Their oncogenic role can be suspected, especially as they are found exclusively in concert with gene amplification. Similarly, mutations have been found in the ectodomain but not in the tyrosine-kinase domain of *EGFR* gene in adult GBM [66]; these mutations were shown to be oncogenic. Moreover, similar to *EGFRvIII* mutants, deletions in the extracellular domain of *PDGFRA* have been already reported in as many as 40% of glioblastomas with *PDGFRA* amplification and were associated with increased tyrosine-kinase activity [67]. Unfortunately, the

assay used for PDGFRA sequencing did not allow us to exclude the possibility of in frame deletions and this would need further analysis on new samples.

Translational implications of targeting genomic alterations in DIPG

Lack of insight into disease mechanisms impeded the development of effective therapies in DIPG for years, with the selection of therapeutic agents to be used in conjunction with irradiation determined empirically or based on their efficacy in adult high-grade gliomas. Changing the paradigm of the treatment of this disease requires a better understanding of the key biological events driving this type of neoplasm. Our clinical and biological program allowed us to discover new potential therapeutic targets previously overlooked or ignored. For the first time, rationale design of trials with targeted therapies could be implemented in the armentarium against these aggressive neoplasms. PDGFRA indeed seems to be the most exciting target given also the existence of several inhibitors with a known toxicity profile in children, including patients with DIPG at relapse [20] or at diagnosis after irradiation [68]. Despite significant drug concentrations reached inside the glioblastoma [69], imatinib has shown limited efficacy in recurrent or newly diagnosed glioblastoma in adults [70] and response to the drug was not increased in patients with PDGFRA immunopositivity [71]. No information on the histology of the brainstem tumors was available in the Pediatric Brain Tumor Consortium (PBTC) phase II trial of imatinib [68] where most of the patients with brainstem gliomas received indeed the drug after the completion of their radiotherapy schedule. In a recent study of the 'Innovative Therapies in Children with Cancer' consortium, where imatinib was only given to patients with proven PDGFRA, PDGFRB or KIT over expression determined by immunohistochemistry [20], one child with recurrent DIPG harboring PDGFRA expression in 50% of the cells in the biopsy showed a sustained objective response (minus 31% for tumor size) for a period of ten months. Identifying the key predictive markers for efficacy of targeted agents will be a vital step in translating genomic data to the clinic, particularly where specific activating mutations are identified. The literature [70,71] indicates however that the effect of imatinib as single agent is limited and that combination with other agents such as irradiation should be considered [72,73]. In addition, insufficient drug penetration in the brain and in some part of the tumor may explain these disappointing results. Enhanced delivery would then need either blood to brain barrier opening [74] or P-gp and ABCG2 inhibition [75]. In this respect the DIPG orthotopic models newly described [63,64] will be valuable tools to study the appropriate way to deliver these drugs in addition to help our understanding of the disease. Combinatorial targeted approaches may also be valid given the observation of multiple oncogenic alterations activating the same downstream signaling cascades [76]. Our finding of simultaneous amplification of PDGFRA and MET in a subset of DIPG, for example, may justify the use of multikinase inhibitors or combinations of TKI, as has been demonstrated for pediatric glioblastoma cells *in vitro* [77].

Our integrated genetic profiling of diagnostic DIPG has identified two biologically and clinically distinct groups of DIPG, with clear differences from hemispheric HGG, and with likely differential treatment strategies warranted. These data highlight the importance of biologically driven guidance for novel therapeutic intervention in these currently untreatable tumors, and argue for the systematic biopsy of these lesions in order to facilitate this, in addition suggesting that some supratentorial deep-seated infiltrating HGG of the deep grey nuclei may deserve a similar approach.

Materials and Methods

Tumor and Nucleic Acids extraction

Tumor samples and clinical information were collected with written informed consent (see Supporting Information S1) of the parents/guardians before inclusion into protocols approved by the Internal Review Board of the Necker Sick Children's Hospital in Paris and the Gustave Roussy Cancer Institute in Villejuif [corresponding to two phase I/II trials, see references 20 and 21]. Only patient with classical diagnostic features of DIPG were included: 1) short clinical history of less than three months, 2) infiltrating neoplasm centered on the pons and involving at least 50% of the anatomical structure, 3) histology excluding a pilocytic astrocytoma or ganglioglioma.

Tumor biopsies were snap frozen in liquid nitrogen in the operating room to ensure preservation of high quality RNA, ground to powder and then RNA and DNA were extracted following two different protocols according to their respective efficiency: Rneasy Micro Kit (Qiagen) and/or TRIzol reagent (Invitrogen).

Microarray Analyses

DNA and RNA microarray hybridizations were carried out by the Functional Genomics Platform of the Integrated Research Cancer Institute in Villejuif (http://www.igr.fr/en/page/integrated-biology_1529) using the Agilent 44 K Whole Human Genome Array G4410B and G4112F, respectively (<http://www.agilent.com>). The microarray data related to this paper are MIAME compliant and the raw data have been submitted to the Array Express data repository at the European Bioinformatics Institute (<http://www.ebi.ac.uk/arrayexpress/>) under the accession number E-TABM-1107.

Bioinformatic Analyses

Raw copy number ratio data were transferred to the CGH Analytics v3.4.40 software for further analysis with the ADM-2 algorithm (<http://www.agilent.com>). A low-level copy number gain was defined as a $\log_2(\text{ratio}) > 0.3$ and a copy number loss was defined as a $\log_2(\text{ratio}) < -0.3$. A high-level gain or amplification was defined as a $\log_2(\text{ratio}) > 1.5$. Minimum common regions (MCR) were defined as chromosome regions that show maximal overlapping aberrations across multiple samples with the STAC v1.2 software [78]. Probe-level measurement MCRs do not include all genes that are altered within a given aberrant region in a particular tumor but define the recurrent abnormalities that span the region.

Raw gene expression data using normal brainstem as reference were transferred into R software for statistical analysis. In order to discover groups in GE data set, the k-means algorithm from R software has been run for two to five groups on the entire dataset. Then for each clustering the BIC value was calculated, according to Guillemot et al [22], in order to determine the best one, which was the one with two groups. GSEA analysis [23] was performed with the pre-ranked tool on gene list ranked by increasing FDR adjusted p-values, for each contrast of interest, with default parameter values. A nominal False Discovery Rate (FDR) of < 0.25 was considered statistically significant for GSEA. We ran GSEA analysis with t-test option as metric parameter.

For integrative genomics analysis, we used the DR-Integrator package for R [79].

Fluorescent In Situ Hybridization

FISH was performed from formalin-fixed-paraffin-embedded (FFPE) tumor samples or frozen tumor touch slides for the

xenografts. The FIP1L1/PDGFR α (Q-biogen/MP Medicals) and LSI EGFR (Vysis/Abbot) were used according to the manufacturer's instructions. PDGFRA and MET probes were labelled from BAC-clones RP11-58C6 and RP11-819D11 (PDGFRA) and RP11-165C4 and RP11-951I21 (MET) using the Bioprime kit (Invitrogen) and DIG-6-dUTP (Roche). Slides were pre-treated in 0.2 M HCl, 8% sodium thioyanate and 0.025% pepsin. Probes were hybridised overnight at 37°C. Slides were washed and incubated with conjugates streptavidin-Cy3 (Invitrogen) and anti-DIG-FITC (Roche).

Mutation screening of selected genes

For direct sequencing, the exon 15 of *BRAF* and all the individual exons of *PDGFRA* were PCR amplified using Taq DNA polymerase (Invitrogen) and primers that can be provided upon request. PCR products were sequenced with BigDye v3.1 and run on an AB3730 genetic analyser (Applied Biosystems). Traces were analysed using Mutation Surveyor software (Softgenetics). The effect of the mutations on the protein structure was predicted using PolyPhen (<http://genetics.bwh.harvard.edu/pph/>) and SIFT (<http://sift.jcvi.org/>) databases.

Histology and Immunohistochemistry on Primary Tumor Material

Tumor histology was reviewed by PV. Tumors were classified and graded according to the 2007 WHO classification. Representative formalin-zinc (formol 5%; Zinc 3 g/L; sodium chloride 8 g/L) fixed sections were deparaffinized and subjected to a Ventana autostainer (BenchMark XT, Ventana Medical system, Tucson, USA) with a standard pretreatment protocol included CC1 buffer for MIB (KI-67) and P53. A semi-automatised system using a microwave antigen retrieval (MicroMED T/T Mega; Hacker Instruments & Industries, Inc., Winnsboro, SC) for 30 minutes at 98°C (manufacturer recommendations) and the RTU Vectastain Universal detection system (Vector laboratories, Burlingame, CA, USA) for Olig2. Sections were then incubated with various commercial monoclonal primary antibodies against Olig2 (AF 2418, 1/150, R/D system, CA, USA), P53 (DO-1, 1/1, Ventana) and MIB-1 (1/100; Dako, Glostrup, Denmark). Diaminobenzidine was used as the chromogen. A minimal threshold at 10% of the total stained tumor cells served as a cut-off for defining the p53-positive status. A MIB-1 labeling index (MIB-1 LI) was obtained by counting the number of MIB-1-positive tumor cells in regions with the maximum number of labeled tumor cells. Ten microscopic high-power field sets were counted, and the MIB-1 LI was computed as a percentage of immunopositive cells from the total cells counted in selected fields. Light microscopic images were digitally captured using a Nikon eclipse E600 microscope (Nikon, Tokyo, Japan) equipped with Nikon DXM 1200 Digital camera. Photomicrographs were assembled for illustrations using the Adobe Photoshop version 7.0.1 software (Adobe, San Jose, California, USA).

Supporting Information

Figure S1 DIPG are different from supratentorial high-grade gliomas in children. Panel A: example of a biopsy sampling in a patient with DIPG. A maximum of 8 core biopsy samples can be obtained per patient. **Panel B: heatmap of the unsupervised hierarchical clustering of 29 DIPG.** From the 32 available samples, two had a completely flat profile and one was of insufficient quality. The analysis was then run on 29 samples. Gains are represented in green (the intensity being correlated to the $\log_2(\text{ratio})$) and amplifications as blue dots. Losses

are represented in red (the intensity being correlated to the $\log_2(\text{ratio})$). The lower panel indicate the general profile of genomic imbalances encountered in the 32 samples, y axis scale being the frequency of the aberrations. The colored right panel shows the profile of each individual sample and the black & white right panel shows the percentage of the genome with imbalances. **C: overall survival of the patients with CGHarray data according to the loss or the persistence of the *TP53* locus.** Overall survival was significantly lower in patients with *TP53* gene loss ($p = 0.01$, log-rank test). **D: principal component analysis (PCA) of pediatric high-grade gliomas (HGG) CGHarray data irrespective of their location.** Hemispheric HGG are indicated in yellow, midline HGG are indicated in grey and brainstem HGG or DIPG are indicated in pink. All the probes passing the quality control were used for the analysis. **E: box-plots comparing the expression of some of the key regulators of brainstem embryogenesis in DIPG (pink) and supratentorial HGG (yellow).** The adjusted p-value of the comparison is given in the upper left corner of each panel. All values are given relative to the expression found in normal adult brainstem. (TIFF)

Figure S2 DIPG comprises two biological subgroups with distinct survival and pathological characteristics. A: Identification of the most optimal Bayesian Information Criterion (BIC) value. The most optimal BIC value was obtained using the class prediction algorithm of Guillemot et al. (BIOTECHNO'08). The graphs show that the accuracy of class prediction did not improve with increasing number of groups. **B: Integrative analysis of genomic and gene expression data.** When considering all DIPG samples from whom both GE and CGHarray data were available, the expression of 1460 genes (i.e. 6% of the genome) was significantly correlated with copy numbers. The cheese-plots of the 20 genes with the highest correlation are provided. Complete data set is available upon request. (TIFF)

Figure S3 Mesenchymal transition and a pro-angiogenic switch define a subset of DIPG. A: The master epithelial to mesenchymal transition regulators, *SNAIL* and *SNAIL2/Slug* are upregulated in a subset of DIPG. The box-plots of the two DIPG subgroups identified are shown in purple and brown respectively. Gene expression are given compared to normal adult brainstem. The p-value is indicated for each gene in the upper right corner of the panel. **B: Angiogenic markers are overexpressed in a subgroup of DIPG.** The two different subgroups of DIPG are represented in purple and light green. The p-value is indicated for each gene in the upper right corner of the panel. Gene expression are given compared to normal adult brainstem. **C: Stem cell markers are overexpressed in a subgroup of DIPG.** The two different subgroups of DIPG are represented in purple and cyan. The p-value is indicated for each gene in the upper left corner of the panel. Gene expression are given compared to normal adult brainstem. **D: Gene expression profiling of one of the DIPG stem cell cultures.** Principal component analysis of one of the DIPG stem cell cultures together with all the primary DIPG samples. (TIFF)

Figure S4 Amplification of multiple RTK in the same tumor. Example of a DIPG sample for which simultaneous amplification of PDGFRA and EGFR could be observed by FISH. (TIFF)

Table S1
(XLSX)

Table S2
(XLS)

Table S3
(XLS)

Supporting Information S1
(PDF)

References

- Qadloui I, Sultan I, Gajjar A (2009) Outcome and prognostic features in pediatric gliomas: a review of 612 cases from the Surveillance, Epidemiology, and End Results database. *Cancer* 115: 5761–5770.
- Hargrave D, Bartels U, Bouffet E (2006) Diffuse brainstem glioma in children: critical review of clinical trials. *The Lancet Oncology* 7: 241–248.
- Donaldson SS, Laningham F, Fisher PG (2006) Advances toward an understanding of brainstem gliomas. *J Clin Oncol* 24: 1266–1272.
- Herrington B, Kieran MW (2009) Small molecule inhibitors in children with malignant gliomas. *Pediatr Blood Cancer* 53: 312–317.
- Jalali R, Raut N, Arora B, Gupta T, Dutta D, et al. (2010) Prospective evaluation of radiotherapy with concurrent and adjuvant temozolomide in children with newly diagnosed diffuse intrinsic pontine glioma. *Int J Radiat Oncol Biol Phys* 77: 113–118.
- Bax DA, Mackay A, Little SE, Carvalho D, Viana-Pereira M, et al. (2010) A distinct spectrum of copy number aberrations in pediatric high-grade gliomas. *Clin Cancer Res* 16: 3368–3377.
- Qi HQ, Jacob K, Fatet S, Ge B, Barnett D, et al. (2010) Genome-wide profiling using single-nucleotide polymorphism arrays identifies novel chromosomal imbalances in pediatric glioblastomas. *Neuro-oncology* 12: 153–163.
- Paugh BS, Qi C, Jones C, Liu Z, Adamowicz-Brice M, et al. (2010) Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease. *J Clin Oncol* 28: 3061–3068.
- Barrow J, Adamowicz-Brice M, Cartmill M, MacArthur D, Lowe J, et al. (2011) Homozygous loss of ADAM3A revealed by genome-wide analysis of pediatric high-grade glioma and diffuse intrinsic pontine gliomas. *Neuro Oncol* 13: 212–222.
- Zarghooni M, Bartels U, Lee E, Buczkowicz P, Morrison A, et al. (2010) Whole-genome profiling of pediatric diffuse intrinsic pontine gliomas highlights platelet-derived growth factor receptor alpha and poly (ADP-ribose) polymerase as potential therapeutic targets. *J Clin Oncol* 28: 1337–1344.
- Albright AL, Packer RJ, Zimmerman R, Roche LB, Boyett J, et al. (1993) Magnetic resonance scans should replace biopsies for the diagnosis of diffuse brain stem gliomas: a report from the Children's Cancer Group. *Neurosurgery* 33: 1026–1029. discussion 1029–1030.
- Hargrave D (2008) Pontine glioma. To biopsy or not to biopsy: that is the question. *Brit J Neurosurg* 22: 624.
- Schumacher M, Schulte-Monting J, Stoeter P, Warmuth-Metz M, Solymosi L (2007) Magnetic resonance imaging compared with biopsy in the diagnosis of brainstem diseases of childhood: a multicenter review. *J Neurosurg* 106: 111–119.
- Pincus DW, Richter EO, Yachnis AT, Bennett J, Bhatti MT, et al. (2006) Brainstem stereotactic biopsy sampling in children. *J Neurosurg* 104: 108–114.
- Louis DN, Rubin MP, Correa KM, Gusella JF, von Deimling A (1993) Molecular genetics of pediatric brain stem gliomas. Application of PCR techniques to small and archival brain tumor specimens. *J Neuropathol Exp Neurol* 52: 507–515.
- Gilbertson RJ, Hill DA, Hernan R, Kocak M, Geyer R, et al. (2003) ERBB1 is amplified and overexpressed in high-grade diffusely infiltrative pediatric brain stem glioma. *Clin Cancer Res* 9: 3620–3624.
- Okada H, Low KL, Kohanbash G, McDonald HA, Hamilton RL, et al. (2008) Expression of glioma-associated antigens in pediatric brain stem and non-brain stem gliomas. *J Neuro-oncol* 88: 245–250.
- Broniscer A, Baker JN, Baker SJ, Chi SN, Geyer JR, et al. (2010) Prospective collection of tissue samples at autopsy in children with diffuse intrinsic pontine glioma. *Cancer* 116: 4632–4637.
- Roujeau T, Machado G, Garnett MR, Miquel C, Puget S, et al. (2007) Stereotactic biopsy of diffuse pontine lesions in children. *J Neurosurg* 107: 1–4.
- Georger B, Morland B, Ndiaye A, Doe F, Kallia G, et al. (2009) Target-driven exploratory study of imatinib mesylate in children with solid malignancies by the Innovative Therapies for Children with Cancer (ITCC) European Consortium. *Eur J Cancer* 45: 2342–2351.
- Georger B, Hargrave D, Thomas F, Ndiaye A, Frappaz D, et al. (2010) Innovative Therapies for Children with Cancer pediatric phase I study of erlotinib in brainstem glioma and relapsing/refractory brain tumors. *Neuro-oncology* 13: 109–118.
- Guillemot V, Philippe C, Tenenhaus A, Rollin J, Gidrol X, et al. (2008) Grouping levels of Exposure with same Observable Effects before Class Prediction in Toxicogenomics. Paper presented at International Conference on

Acknowledgments

The authors also thank Sébastien Dubleumortier and Pr Nicole Brousse for the Centre de Ressources Biologiques from Necker Hospital (Paris).

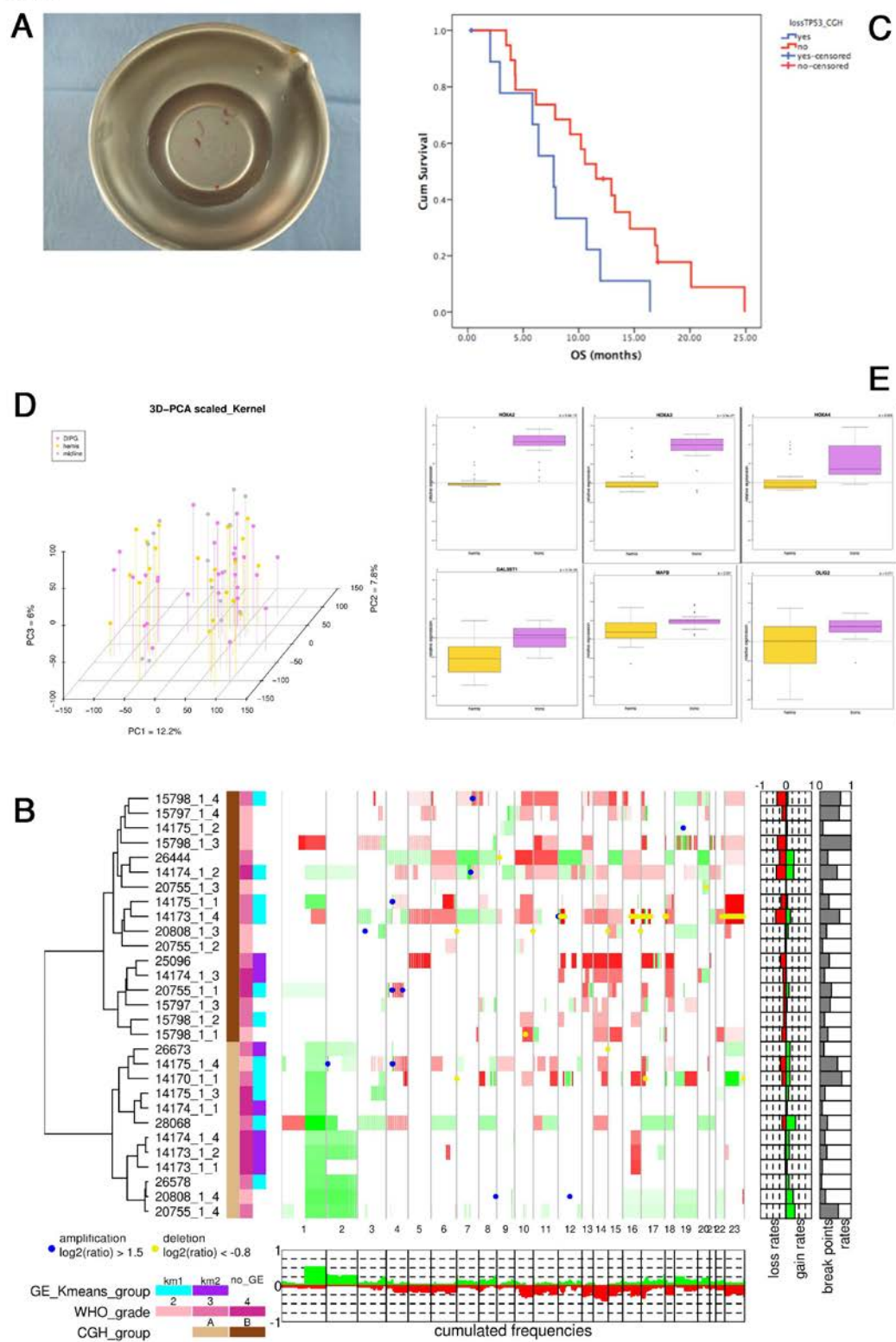
Author Contributions

Conceived and designed the experiments: JG BG CJ VL PD CP. Performed the experiments: SP CR LG-R FA PV DC RR DB M-PJ. Analyzed the data: GLT BJCP PV SP. Contributed reagents/materials/analysis tools: CS-R TR SP BG. Wrote the paper: JG SP CP CJ BG.

- Bioinformatics and Biomedical Technologies (BIOTECHNO'08) Digital Library, ISBN: 978-0-7695-3191-5, pp.164–169. Bucharest, Rumania, June 29th–July 5th.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102: 15545–15550.
- Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, et al. (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17: 98–110.
- Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, et al. (2010) The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 463: 318–325.
- Lin X, Baritaki S, Millicello L, Malaponte G, Bevelacqua Y, et al. (2010) The Role of BRAF Mutations in Melanoma and the Induction of EMT via Dysregulation of the NF-kappaB/SnaI/RKIP/PTEN Circuit. *Genes Cancer* 1: 409–420.
- Schiffman JD, Hodgson JG, VandenBerg SR, Flaherty P, Polley MY, et al. (2010) Oncogenic BRAF mutation with CDKN2A inactivation is characteristic of a subset of pediatric malignant astrocytomas. *Cancer Res* 70: 512–519.
- Thirant G, Besette B, Varlet P, Puget S, Gadusseau J, et al. (2011) Clinical relevance of tumor cells with stem-like properties in pediatric brain tumors. *PLoS ONE* 6: e16375.
- Takada N, Kucenas S, Appel B (2010) Sox10 is necessary for oligodendrocyte survival following axon wrapping. *Glia* 58: 996–1006.
- Pozniak CD, Langseth AJ, Dijkgraaf CJ, Choe Y, Werb Z, et al. (2010) Sox10 directs neural stem cells toward the oligodendrocyte lineage by decreasing Suppressor of Fused expression. *Proc Natl Acad Sci U S A* 107: 21795–21800.
- Negri T, Borzi F, Conca E, Brich S, Gronchi A, et al. (2009) Oncogenic and ligand-dependent activation of KIT/PDGFR α in surgical samples of imatinib-treated gastrointestinal stromal tumours (GISTs). *J Pathol* 217: 103–112.
- Lei L, Sonabend AM, Guarnieri P, Soderquist C, Ludwig T, et al. (2011) Glioblastoma models reveal the connection between adult glial progenitors and the proneural phenotype. *PLoS ONE* 6(5): e20041.
- Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, et al. (2008) A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci* 28(1): 264–278.
- Gilbertson RJ, Gutmann DH (2007) Tumorigenesis in the brain: location, location. *Cancer Res* 67: 5579–5582.
- Taylor MD, Poppleton H, Fuller C, Su X, Liu Y, et al. (2005) Radial glia cells are candidate stem cells of ependymoma. *Cancer Cell* 8: 323–335.
- Modena P, Lualdi E, Fachinetti F, Velman J, Reai JF, et al. (2006) Identification of tumor-specific molecular signatures in intracranial ependymoma and association with clinical characteristics. *J Clin Oncol* 24: 5223–5233.
- Palm T, Figarella-Branger D, Chapon F, Lacroix C, Gray F, et al. (2009) Expression profiling of ependymomas unravels localization and tumor grade-specific tumorigenesis. *Cancer* 115: 3955–3968.
- Andreiuolo F, Puget S, Peyre M, Dantas-Barbosa C, Boddaert N, et al. (2010) Neuronal differentiation distinguishes supratentorial and infratentorial childhood ependymomas. *Neuro-oncology* 12: 1126–1134.
- Sharma MK, Mansur DB, Reifenberger G, Perry A, Leonard JR, et al. (2007) Distinct genetic signatures among pilocytic astrocytomas relate to their brain region origin. *Cancer Res* 67: 890–900.
- Lee Y, Kawagoe R, Sasai K, Li Y, Russell HR, et al. (2007) Loss of suppressor-of-fused function promotes tumorigenesis. *Oncogene* 26: 6442–6447.
- Mojie M, Mitra SS, Frenet ME, Raveti TB, Kim J, et al. (2011) Hedgehog-responsive candidate cell of origin for diffuse intrinsic pontine glioma. *Proc Natl Acad Sci* 108: 4453–4458.
- Puget S, Crimmins DW, Garnett MR, Grill J, Oliveira R, et al. (2007) Thalamic tumors in children: a reappraisal. *J Neurosurg* 106: 354–362.
- Phillips HS, Kharabanda S, Chen R, Forrest WF, Soriano RH, et al. (2006) Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 9: 157–173.
- Bao S, Wu Q, McLendon RE, Hao Y, Shi Q, et al. (2006) Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* 444: 756–760.

45. Ducray F, de Reynies A, Chinot O, Idhah A, Figarella-Branger D, et al. (2010) An ANOGEF genomic and transcriptomic microarray study of the response to radiotherapy or to alkylating first-line chemotherapy in glioblastoma patients. *Molecular Can* 9: 234.
46. Kessler J, Hainel A, Wichmann H, Rot S, Kappler M, et al. (2010) HIF-1 α inhibition by siRNA or chetomin in human malignant glioma cells: effects on hypoxic radioresistance and monitoring via CA9 expression. *BMC Cancer* 10: 605.
47. Cao F, Hata R, Zhu P, Nakashiro K, Sakanaka M, et al. (2010) Conditional deletion of Stat3 promotes neurogenesis and inhibits astroglialogenesis in neural stem cells. *Biochem Biophys Res Commun* 394: 843–847.
48. Garcia JL, Perez-Caro M, Gomez-Moreta JA, Gonzalez F, Ortuz J, et al. (2010) Molecular analysis of ex-vivo CD133+ GBM cells revealed a common invasive and angiogenic profile but different proliferative signatures among high grade gliomas. *BMC Cancer* 10: 454.
49. Liu Q, Nguyen DH, Dong Q, Shitaku P, Chung K, et al. (2009) Molecular properties of CD133+ glioblastoma stem cells derived from treatment-refractory recurrent brain tumors. *J Neuro-oncol* 94: 1–19.
50. Kurrey NK, Jalgaonkar SP, Joglekar AV, Ghanate AD, Chaskar PD, et al. (2009) Snail and slug mediate radioresistance and chemoresistance by antagonizing p53-mediated apoptosis and acquiring a stem-like phenotype in ovarian cancer cells. *Stem cells (Dayton, Ohio)* 27: 2059–2068.
51. Lucas JT, Jr., Salimath BP, Siomiany MG, Rosenzweig SA (2010) Regulation of invasive behavior by vascular endothelial growth factor is HIF1-dependent. *Oncogene* 29: 4449–4459.
52. Wang Y, Ngo VN, Mariani M, Yang Y, Wright G, et al. (2010) Critical role for transcriptional repressor Snail2 in transformation by oncogenic RAS in colorectal carcinoma cells. *Oncogene* 29: 4658–4670.
53. Faury D, Nantal A, Dunn SE, Guiot MC, Haque T, et al. (2007) Molecular profiling identifies prognostic subgroups of pediatric glioblastoma and shows increased YB-1 expression in tumors. *J Clin Oncol* 25: 1196–1208.
54. Cooper LA, Gutman DA, Long Q, Johnson BA, Cholleti SR, et al. (2010) The proneural molecular signature is enriched in oligodendrogliomas and predicts improved survival among diffuse gliomas. *PLoS ONE* 5: e12548.
55. Ducray F, Idhah A, de Reynies A, Bieche I, Thillet J, et al. (2008) Anaplastic oligodendrogliomas with 1p19q codeletion have a proneural gene expression profile. *Molecular Can* 7: 41.
56. Freije WA, Castro-Vargas FE, Fang Z, Horvath S, Cloughesy T, et al. (2004) Gene expression profiling of gliomas strongly predicts survival. *Cancer Res* 64: 6503–6510.
57. Rowitch DH, Kriegstein AR (2010) Developmental genetics of vertebrate glial-cell specification. *Nature* 468: 214–222.
58. Mehta S, Hullard E, Kesari S, Mair CL, Golebiowski D, et al. (2011) The central nervous system-restricted transcription factor Olig2 opposes p53 responses to genotoxic damage in neural progenitors and malignant glioma. *Cancer Cell* 19: 359–371.
59. De Carli E, Wang X, Puget S (2009) IDH1 and IDH2 mutations in gliomas. *New Engl J Med* 360: 2248.
60. Pollack IF, Hamilton RL, Sobol RW, Nikiforova MN, Lyons-Weiler MA, et al. (2011) IDH1 mutations are common in malignant gliomas arising in adolescents: a report from the Children's Oncology Group. *Childs Nerv Syst* 27: 87–94.
61. Ferletta M, Uehrom L, Olofsson T, Ponten F, Westermark B (2007) Sox10 has a broad expression pattern in gliomas and enhances platelet-derived growth factor-B induced gliomagenesis. *Mol Cancer Res* 5: 891–897.
62. Lindberg N, Kastemar M, Olofsson T, Smits A, Uehrom L (2009) Oligodendrocyte progenitor cells can act as cell of origin for experimental glioma. *Oncogene* 28: 2266–2275.
63. Appoloni I, Calzolai F, Tuccel E, Cavaglia S, Terile M, et al. (2009) PDGF-B induces a homogeneous class of oligodendrogliomas from embryonic neural progenitors. *Int J Cancer* 124: 2251–2259.
64. Masui K, Suzuki SO, Torisu R, Goldman JE, Canoll P, et al. (2010) Glial progenitors in the brainstem give rise to malignant gliomas by platelet-derived growth factor stimulation. *Glia* 58: 1050–1065.
65. Becher OJ, Hambarzumyan D, Walker TR, Helny K, Nazarian J, et al. (2010) Preclinical evaluation of radiation and perifosine in a genetically and histologically accurate model of brainstem glioma. *Cancer Res* 70: 2548–2557.
66. Lee JC, Vivanco I, Beroukhi R, Huang JH, Feng WL, et al. (2006) Epidermal growth factor receptor activation in glioblastoma through novel missense mutations in the extracellular domain. *PLoS medicine* 3: e485.
67. Ozawa T, Brennan CW, Wang L, Squatrito M, Sasayama T, et al. (2010) PDGFRA gene rearrangements are frequent genetic events in PDGFRA-amplified glioblastomas. *Genes Dev* 24: 2205–2218.
68. Pollack IF, Jakacki RI, Blaney SM, Hancock ML, Kieran MW, et al. (2007) Phase I trial of imatinib in children with newly diagnosed brainstem and recurrent malignant gliomas: a Pediatric Brain Tumor Consortium report. *Neuro-oncology* 9: 145–160.
69. Holdhoff M, Supko JG, Gallia GL, Hann CL, Bonekamp D, et al. (2010) Intratumoral concentrations of imatinib after oral administration in patients with glioblastoma multiforme. *J Neuro-oncol* 97: 241–245.
70. Wen PY, Yung WKA, Lamborn KR, Dahia PL, Wang Y, et al. (2006) Phase I/II study of imatinib mesylate for recurrent malignant gliomas: North American Brain Tumor Consortium Study 99-08. *Clin Cancer Res* 12: 4899–4907.
71. Paulsson J, Lindh MB, Jarvis M, Puputti M, Nister M, et al. (2011) Prognostic but not predictive role of platelet-derived growth factor receptors in patients with recurrent glioblastoma. *Int J Cancer* 128: 1981–1988.
72. Holdhoff M, Kreuzer KA, Appelt C, Scholz R, Na IK, et al. (2005) Imatinib mesylate radiosensitizes human glioblastoma cells through inhibition of platelet-derived growth factor receptor. *Blood Cells Mol Dis* 34: 181–185.
73. Russell JS, Brady K, Burgan WE, Cerra MA, Oswald KA, et al. (2003) Gleevec-mediated inhibition of Rad51 expression and enhancement of tumor cell radiosensitivity. *Cancer Res* 63: 7377–7383.
74. Kast RE, Focosi D (2010) Three paths to better tyrosine kinase inhibition behind the blood-brain barrier in treating chronic myelogenous leukemia and glioblastoma with imatinib. *Translat Oncol* 3: 13–15.
75. Lagas JS, van Waterschoot RAB, van Tilburg VACJ, Hillbrand MJ, Lankheet N, et al. (2009) Brain accumulation of dasatinib is restricted by F-glycoprotein (ABC1) and breast cancer resistance protein (ABCG2) and can be enhanced by elacridar treatment. *Clin Cancer Res* 15: 2344–2351.
76. Stommel JM, Kimmelman AC, Ying H, Nabioullin R, Ponugoti AH, et al. (2007) Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science (New York, NY)* 318: 287–290.
77. Bax DA, Gaspar N, Little SE, Marshall L, Perryman L, et al. (2009) EGFRvIII deletion mutations in pediatric high-grade glioma and response to targeted therapy in pediatric glioma cell lines. *Clin Cancer Res* 15: 5753–5761.
78. Diskin SJ, Eck T, Greshock J, Mosse YP, Naylor T, et al. (2006) STAC: A method for testing the significance of DNA copy number aberrations across multiple array-CGH experiments. *Genome Res* 16: 1149–1158 (2006).
79. Salari K, Tshibani R, Pollack JR (2010) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. *Bioinformatics (Oxford, England)* 26: 414–416.

Figure S1



A

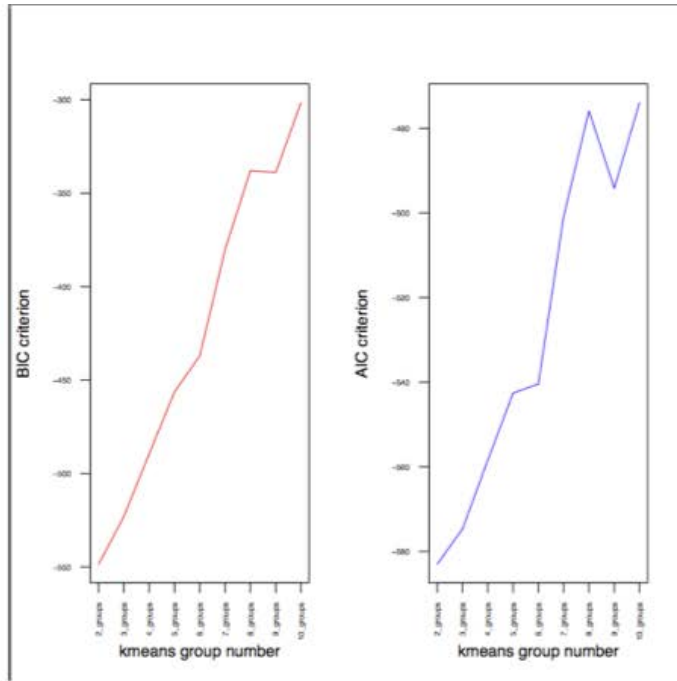


Figure S2

B

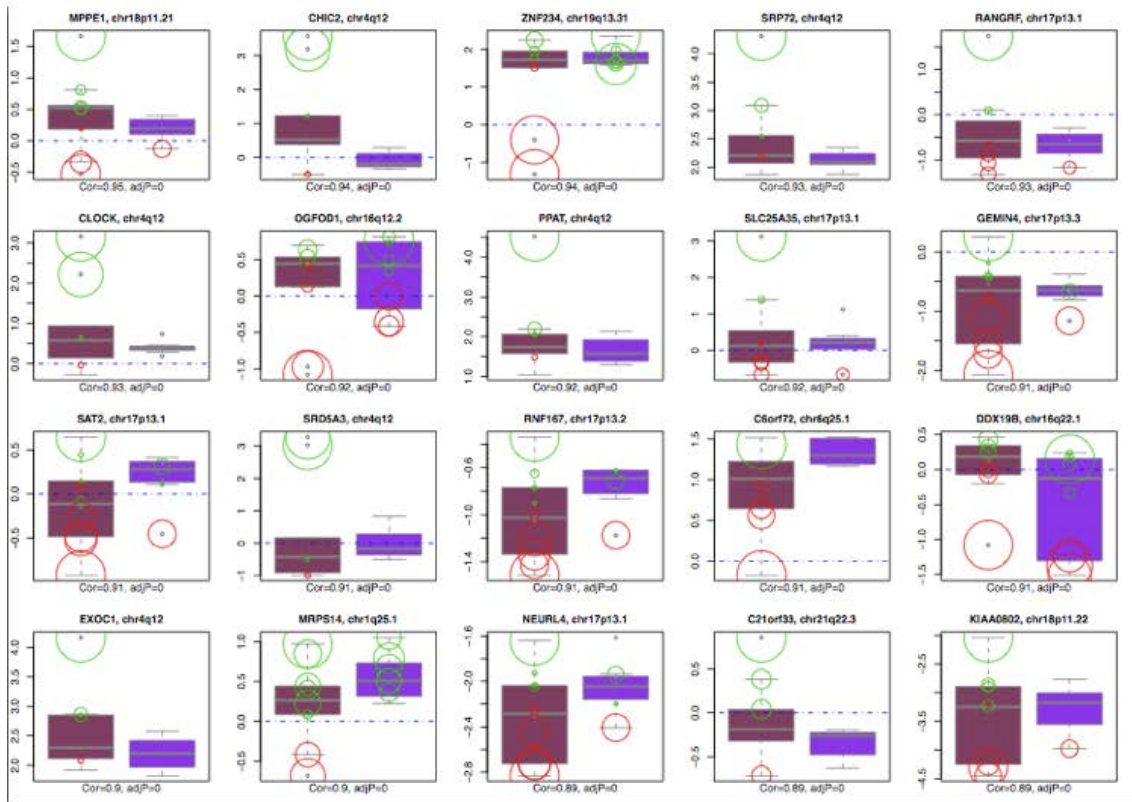


Figure S3

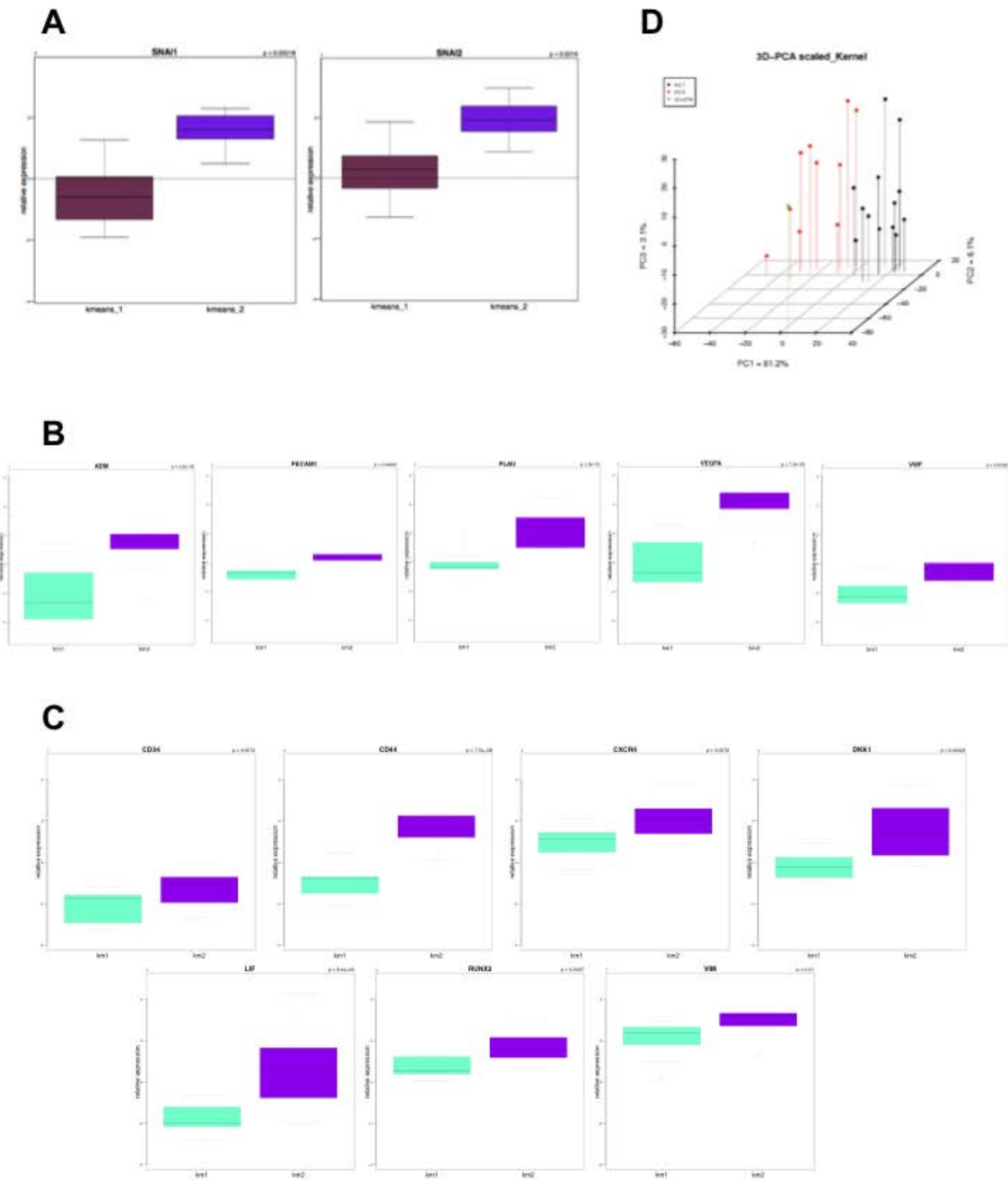
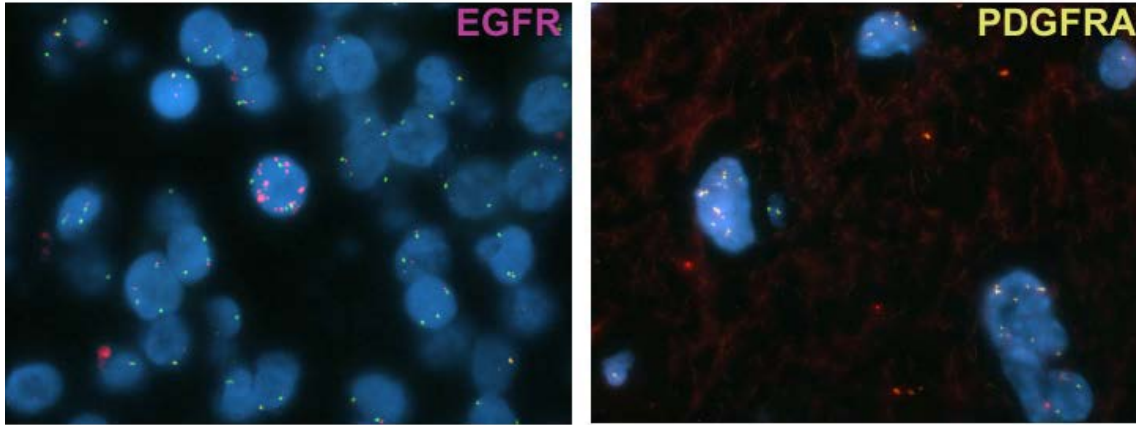


Figure S4



Novel Oncogenic *PDGFRA* Mutations in Pediatric High-Grade Gliomas

Barbara S. Paugh¹, Xiaoyan Zhu¹, Chunxu Qu², Raelene Endersby⁷, Alexander K. Diaz^{1,6}, Junyuan Zhang¹, Dorine A. Bax⁸, Diana Carvalho⁸, Rui M. Reis⁹, Arzu Onar-Thomas³, Alberto Broniscer⁴, Cynthia Wetmore⁴, Jinghui Zhang⁵, Chris Jones⁵, David W. Ellison⁵, and Suzanne J. Baker^{1,6}

Abstract

The outcome for children with high-grade gliomas (HGG) remains dismal, with a 2-year survival rate of only 10% to 30%. Diffuse intrinsic pontine glioma (DIPG) comprise a subset of HGG that arise in the brainstem almost exclusively in children. Genome-wide analyses of copy number imbalances previously showed that platelet-derived growth factor receptor α (*PDGFRA*) is the most frequent target of focal amplification in pediatric HGGs, including DIPGs. To determine whether *PDGFRA* is also targeted by more subtle mutations missed by copy number analysis, we sequenced all *PDGFRA* coding exons from a cohort of pediatric HGGs. Somatic-activating mutations were identified in 14.4% (13 of 90) of nonbrainstem pediatric HGGs and 4.7% (2 of 43) of DIPGs, including missense mutations and in-frame deletions and insertions not previously described. Forty percent of tumors with mutation showed concurrent amplification, whereas 60% carried heterozygous mutations. Six different mutations impacting different domains all resulted in ligand-independent receptor activation that was blocked by small molecule inhibitors of PDGFR. Expression of mutants in p53-null primary mouse astrocytes conferred a proliferative advantage *in vitro* and generated HGGs *in vivo* with complete penetrance when implanted into brain. The gene expression signatures of these murine HGGs reflected the spectrum of human diffuse HGGs. *PDGFRA* intragenic deletion of exons 8 and 9 were previously shown in adult HGG, but were not detected in 83 nonbrainstem pediatric HGG and 57 DIPGs. Thus, a distinct spectrum of mutations confers constitutive receptor activation and oncogenic activity to PDGFR α in childhood HGG. *Cancer Res*; 73(20): 6219–29. ©2013 AACR.

Introduction

Pediatric high-grade gliomas (HGG) comprise 15% to 20% of all childhood tumors of the central nervous system (1). Despite aggressive therapy, prognosis for pediatric HGG remains very poor, with a 2-year survival rate of less than 20% (2). A subset of HGGs arise in the brainstem as diffuse intrinsic pontine glioma (DIPG), a disease that occurs almost exclusively in children and has a 2-year survival rate of less than 10% (3). Although the histopathologies of pediatric and adult HGGs can be similar, genome-wide studies have shown significant differences in the

frequencies of specific copy number alterations as well as both similarities and differences in the gene expression signatures of HGGs in these two age groups (4–12). Somatic mutations in histone H3 occur in 78% of DIPGs and 36% of nonbrainstem pediatric glioblastomas, but occurred only rarely in young adults with glioblastoma and not in older patients with adult glioblastoma (11, 12). Thus, distinct molecular mechanisms drive gliomagenesis at different ages.

Platelet-derived growth factor receptor α (*PDGFRA*) is the most frequent target of focal amplification in pediatric HGGs arising within and outside the brainstem (4–7, 10), and somatic mutations of *PDGFRA* have been recently reported in pediatric HGGs (6, 12). In contrast, EGF receptor (*EGFR*) is the predominant receptor tyrosine kinase (RTK) targeted by both amplification and mutation in adult glioblastoma (13, 14). Pediatric HGGs with genomic amplification of *PDGFRA* showed concomitant increases in *PDGFRA* mRNA by gene expression profiling. Furthermore, PDGFR α overexpression without genomic amplification is commonly found in pediatric HGGs, and amplification of the genes encoding PDGF ligands or overexpression with and without *PDGFRA* aberrations were also reported, suggesting both autocrine and paracrine signaling.

PDGF and its receptors are involved in many cellular processes such as migration, survival, and proliferation and they are critical during developmental processes (15). Ligand binding induces receptor dimerization and results in

Authors' Affiliations: Departments of ¹Developmental Neurobiology, ²Computational Biology, ³Biostatistics, ⁴Oncology, and ⁵Pathology, St. Jude Children's Research Hospital; ⁶Interdisciplinary Biomedical Science Program, University of Tennessee Health Sciences Center, Memphis, Tennessee; ⁷Telethon Institute for Child Health Research, Centre for Child Health Research, The University of Western Australia, Perth, Australia; ⁸Divisions of Molecular Pathology and Cancer Therapeutics, The Institute of Cancer Research, London, United Kingdom; and ⁹Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, São Paulo, Brazil

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Corresponding Author: Suzanne J. Baker, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105. Phone: 901-595-2254; Fax: 901-595-2270; E-mail: suzanne.baker@stjude.org

doi: 10.1158/0008-5472.CAN-13-1491

©2013 American Association for Cancer Research.

phosphorylation of the receptor at multiple tyrosine residues. Activated PDGFRs transduce signals through multiple downstream pathways, including the PI3K/Akt, RAS/MAP kinase, Src kinase family, and PLC/PKC pathways, which have all been implicated in tumorigenesis (15, 16).

Abnormally activated PDGFR α signaling driven by viral expression of PDGFB ligand is sufficient to induce glioma formation *in vivo*, indicating that activation of PDGFR pathways is potentially an early event in tumorigenesis (17–19). Furthermore, simultaneous overexpression of PDGFB and loss of *TP53* induced murine HGG with increased incidence and shorter latency, indicating cooperativity between these pathways (20, 21). However, these studies focused on autocrine and paracrine activation of PDGFR signaling pathways by PDGFB ligand overexpression.

Here, we report that pediatric HGGs, including DIPGs, carry novel somatic-activating mutations of *PDGFRA* that are constitutively active, tumorigenic, and sensitive to small molecule inhibitors.

Materials and Methods

Clinical samples

Pediatric HGG samples were obtained from St. Jude Children's Research Hospital (Memphis, TN) and the Royal Marsden Hospital (London, United Kingdom; Supplementary Table S2). Ethical Review Committee approval was obtained from each institution/consortium. Genomic DNA was extracted as previously described from snap-frozen (22) or formalin-fixed paraffin-embedded material (10).

Mutation analysis of *PDGFRA*

All coding exons of *PDGFRA* were sequenced by direct sequencing of PCR-amplified products from genomic DNA in the tumors listed in the Supplementary Table S2, including 39 cases of nonbrainstem pediatric HGGs and 43 cases of DIPGs, using primers listed in the Supplementary Table S4, or by exome sequencing for three DIPG samples. For an additional 51 cases of nonbrainstem pediatric HGG, DNA was extracted from formalin-fixed paraffin-embedded tissue and amplified and sequenced using primers published previously (9). Identified *PDGFRA* mutations were validated by independent PCR and matched normal samples were sequenced when available. Expression of mutated receptor was confirmed by reverse transcription (RT)-PCR and sequencing using primers listed in the Supplementary Table S4 for available cDNA samples. Eighty-three nonbrainstem pediatric HGGs samples were screened by RT-PCR for *KDR-PDGFR* gene fusion (23) and the single case identified was validated by independent PCR and sequencing. cDNA from 83 nonbrainstem pediatric HGG and 57 DIPG cases were screened for *PDGFRA*^{S8,9}, the previously reported 243 base-pair deletion in exons 8 and 9 as described (23).

In vitro analyses of overexpression of wild-type and mutant *PDGFRA*

Wild-type and mutated *PDGFRA* open reading frames were cloned into the MSCV-IRES-GFP (MIG) retroviral vector and used to generate retrovirus (24). Cortical astrocyte cultures

were established from 2-day-old mice (*GFAP-cre;Tpp53^{loxP/loxP}*) as described previously (25). At passage one, p53-null astrocytes were transduced with retroviruses expressing wild-type PDGFR α , PDGFR α mutants or empty vector, and *in vitro* and tumorigenesis experiments were carried out before passage six. For proliferation assays, 5.5×10^3 cells per well were plated on 96-well plates in triplicate. Cells were grown in Dulbecco's Modified Eagle Medium (DMEM)/F-12 supplemented with 10% FBS and 20 ng/mL mouse EGF (Millipore), but without exogenous addition of the PDGF ligand. Proliferation was measured using XTT (2,3-bis[2-methoxy-4-nitro-5-sulphophenyl]-2H-tetrazolium-5-carboxanilide inner salt) assay (Roche) at 24-hour intervals over a 4-day period, without replacing the growth medium. For inhibitor studies, cells were allowed to attach for 4 hours after seeding, then 225 nmol/L (100 ng/mL) crenolanib (AROG Pharmaceuticals), 50 nmol/L dasatinib (LC Laboratories), or vehicle (0.1% dimethyl sulfoxide, DMSO) were added to the cells in a single dose and growth was assayed by XTT as above. Data were normalized to the cell number measured at time zero of the experiment, which was acquired within the first 8 hours from cell seeding (4 hours for cell attachment and 4 hours for development of XTT). For cell-cycle analyses, 2×10^6 cells were seeded per 10 cm dish and the next day cells were treated with 225 nmol/L (100 ng/mL) crenolanib (AROG Pharmaceuticals), 50 nmol/L dasatinib (LC Laboratories), or vehicle (0.1% DMSO) for 24 hours. Following inhibitor treatment, cells in the supernatant from each dish were collected, pooled with trypsinized cells, and washed with PBS. Cells were fixed in 70% ethanol overnight at 4°C and then stained using Guava Cell Cycle Reagent (Millipore). Data were acquired on the Guava EasyCyte using CytoSoft software (Millipore).

Tumorigenesis studies

Mouse experiments were approved by the Institutional Animal Care and Use Committee and are in compliance with national and institutional guidelines. A total of 2×10^6 transduced astrocytes were implanted intracranially into athymic nude mice for tumorigenesis studies as previously described (25). On the manifestation of brain tumor symptoms, mice were anesthetized and perfused with PBS. GFP-labeled tumors were dissected using a fluorescent dissecting microscope. For each tumor, a portion was snap-frozen for protein analyses and RNA extraction, and the remainder was fixed in 4% paraformaldehyde in PBS at 4°C overnight, then processed, embedded in paraffin, and cut into 5 μ m sections. Hematoxylin and eosin (H&E)-stained sections from all collected tumors were evaluated by a clinical neuropathologist (D.W. Ellison) and graded according to World Health Organization (WHO) criteria (26). Immunohistochemistry was conducted with microwave antigen retrieval in a citrate solution using the following primary antibodies from Cell Signaling Technology: PDGFR α (#5241), phospho-Akt S473 (#9271), and phospho-4E-BP1 Thr37/46 (#2855). Anti-rabbit biotinylated secondary antibodies were used in conjunction with horseradish peroxidase-conjugated streptavidin (Elite ABC; Vector Labs). Staining was developed with NovaRED substrate (Vector Labs) and the sections were counterstained with hematoxylin (Vector Labs).

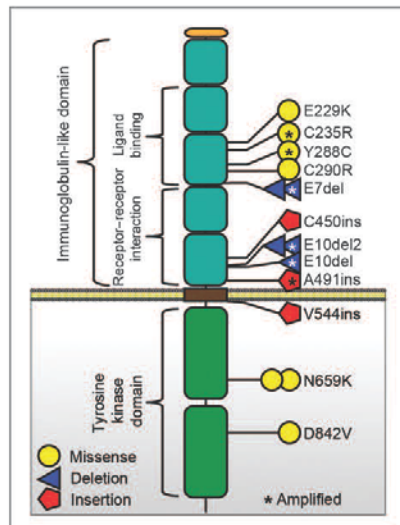


Figure 1. *PDGFRA* somatic mutations identified in pediatric HGGs and DIPGs. Sanger sequencing of *PDGFRA* from genomic DNA revealed multiple mutations, including missense mutations, in-frame insertions, and in-frame deletions. The schematic shows the location of the mutations and affected domains of PDGFR α . A star (*) indicates samples with concomitant mutation and amplification of the *PDGFRA* locus. The signal peptide is designated by an orange oval and the transmembrane domain by a brown rectangle.

mutations in the kinase domain, N659K and D842V were previously reported in gastrointestinal stromal tumors but not in gliomas (32, 33). The remaining mutations are all novel, including four different missense mutations (E229K, C235R, Y288C, and C290R), three in-frame insertions (C450ins, A491ins, and V544ins), and three in-frame deletions (E7del, E10del, and E10del2; Supplementary Table S1) not found in previous studies of a large collection of pediatric HGGs or adult HGGs (6, 12, 23, 34). Three of the identified mutations (E7del, E10del2, and N659K) were recurrent, each of them found in two different HGG cases.

We also evaluated whether activating *PDGFRA* mutations previously shown in adult HGG occur frequently in pediatric HGG. Forty percent of adult glioblastomas with *PDGFRA* amplification harbor an in-frame deletion of 243 base pairs in exons 8 and 9 of the extracellular portion of the receptor (*PDGFRA* ^{Δ 8,9}) that renders the receptor constitutively active (23, 35). A gene fusion between *PDGFRA* and *KDR* (*VEGFR2*), which rendered the receptor constitutively active and tumorigenic *in vivo*, has been reported in one out of 215 cases of adult glioblastoma. We used RT-PCR to screen for these alterations that would have been missed by exonic sequencing. We found

one case of *KDR-PDGFR* in 83 nonbrainstem pediatric HGGs analyzed (Supplementary Fig. S1B). There were no examples of the *PDGFRA* ^{Δ 8,9} deletion in 83 nonbrainstem pediatric HGG and 57 DIPG screened by RT-PCR (Supplementary Fig. S1C). Thus, a different spectrum of intragenic mutations target *PDGFRA* in HGGs from different age groups.

PDGFR α mutants are constitutively active

To better understand the consequence of PDGFR α mutation in pediatric gliomagenesis, retroviral constructs expressing wild-type PDGFR α or six selected PDGFR α mutants that affect different regions of the receptor were generated for functional studies. p53-null PMA cultures were chosen as a relevant cellular background to assess PDGFR α function, because 70% of pediatric HGGs outside of the brainstem and 46% of DIPGs with amplified and/or mutated *PDGFRA* also have inactivating mutations of *TP53* (Supplementary Table S2). All six mutants were constitutively active, inducing strong phosphorylation of PDGFR α at Tyr-572/574, Tyr-742, Tyr-720, and Tyr-754 when expressed in p53-null PMAs in the absence of serum. These phosphotyrosine sites have been shown to associate with activation of different downstream signaling pathways. Src family members bind to phosphorylated PDGFR α at Tyr-572/574, phospho-Tyr-742 promotes interaction with p85, the regulatory subunit of PI3K, and leads to PI3K pathway activation, whereas phosphorylated Tyr-720 and Tyr-754 recruit and activate SHP-2 (Src homology-2 domain-containing phosphatase), which regulates activation of Src family kinases and the MAPK pathway (16, 36–38). Different levels of receptor phosphorylation were observed depending on the specific mutation, with the E10del2 exhibiting the lowest receptor activation. All mutants were expressed to similar levels as shown by total PDGFR α (Fig. 2A, left). In contrast, overexpression of wild-type PDGFR α did not lead to receptor activation in the ligand-free condition. However, treatment of these cells with PDGF-AA, triggered wild-type PDGFR α phosphorylation to levels similar to or greater than the mutant receptors (Fig. 2A, right). Furthermore, phosphorylation of known downstream signaling targets of PDGFR α confirmed the constitutive activity of PDGFR α mutants. In the absence of ligand, elevated levels of phospho-Akt (Ser473 and Thr308) and higher phosphorylation levels of S6 ribosomal protein and PRAS40 were observed in comparison with wild-type PDGFR α -expressing cells, indicating ligand-independent activation of the PI3K pathway by the mutated receptor. Activation of MAPK signaling, detected by phosphorylation of p44/42 MAPK was not increased by mutant PDGFR α in the absence of ligand. Ligand stimulation further increased phosphorylation levels of PI3K pathway components but had only modest effect on the MAPK pathway activation (Fig. 2A and Supplementary Fig. S2). Expression of PDGFR α mutants as well as wild-type receptor added a significant proliferative advantage in comparison with empty vector control cells as determined by XTT assays (Fig. 2B).

PDGFR α signaling is abrogated by inhibitors

To analyze whether wild-type and PDGFR α mutants have differential responses to small molecule inhibitors, two distinct

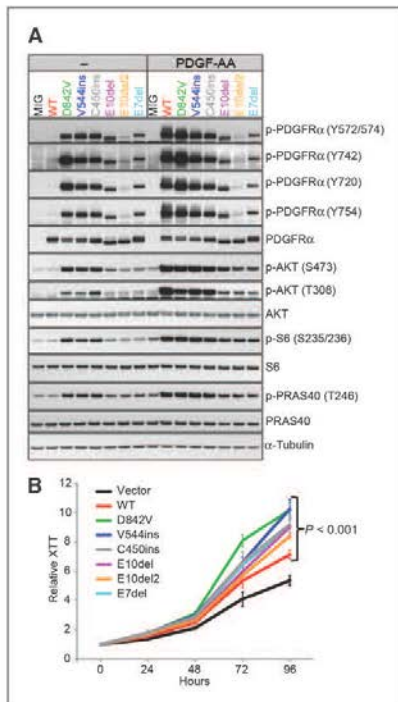


Figure 2. PDGFR α mutants are constitutively active and confer proliferative advantage. A, Western blot analysis of whole cell lysates from p53-null PMAs transduced with retroviruses expressing wild-type PDGFR α (WT), PDGFR α mutants, or empty vector. Cells were grown in serum-free conditions overnight and then with or without PDGF-AA for 30 minutes. PDGFR α signaling activation of downstream targets was monitored using indicated antibodies in the presence and absence of PDGF-AA. B, proliferation of p53-null PMAs transduced with retroviruses expressing wild-type, PDGFR α mutants, or empty vector was measured by an XTT assay in triplicate. Error bars show SD. A representative figure of three independent experiments is shown.

compounds were selected. Dasatinib (BMS-354825) is a potent, multityrosine kinase inhibitor that exerts broad antiproliferative activity by targeting PDGFR α , PDGFR β , ABL, SRC family kinases, KIT, and several other tyrosine kinases (39, 40). In contrast, crenolanib (CP-868,596) is a relatively specific inhibitor of PDGFR α and PDGFR β , that is more than 100-fold more selective for PDGFRs versus a variety of other kinases (e.g., KIT, VEGFR-2, TIE-2, FGFR-2, EGFR, ERBB2, and SRC; refs. 41, 42). Both drugs have been shown to inhibit PDGFR kinase activity by competing with ATP. To compare wild-type and mutant PDGFR α signaling alterations in response to these

inhibitors, cells were serum-starved overnight, then treated with crenolanib or dasatinib for 3 hours before adding PDGF-AA ligand. As previously reported, the D842V mutant was resistant to dasatinib treatment, however it was effectively inhibited by crenolanib (42). Both inhibitors significantly blocked activation of wild-type PDGFR α and all other mutants analyzed as shown by decreased phosphorylation of the receptor at Tyr-572/574, Tyr-742, Tyr-720, and Tyr-754 (Fig. 3A). Furthermore, downstream signaling activation was significantly diminished, including decreased phosphorylation of PI3K and MAPK pathway effectors (Fig. 3A and Supplementary Fig. S3A). D842V and V544ins were the only two mutants that induced a substantial increase in SHP2 phosphorylation (Supplementary Fig. S2) and the only two mutants for which crenolanib induced a substantial decrease in phospho-SHP2 (Supplementary Fig. S3A).

The effect of each inhibitor on cell proliferation was examined by XTT assay (Fig. 3B and Supplementary Fig. S3B), using a concentration greater than the IC₅₀ determined to block receptor phosphorylation (Supplementary Fig. S4). Crenolanib selectively inhibited proliferation of PDGFR α -expressing cells, both wild-type and all analyzed mutants, whereas it did not affect proliferation of the empty vector control cells (Fig. 3B and Supplementary Fig. S3B). In contrast, dasatinib exerted broad antiproliferative effects and significantly inhibited empty vector control cells in addition to PDGFR α -expressing cells. Furthermore, concordant with the inability of dasatinib to block phosphorylation of the D842V mutant, cells expressing this mutant showed a diminished growth inhibitory response compared with other mutants (Fig. 3B). Crenolanib and dasatinib both exert cytostatic effects on PDGFR α -expressing p53-null PMAs, arresting cells in G₀-G₁ (Supplementary Fig. S5). However, neither inhibitor induced significant cell death, as assessed by annexin V staining and TUNEL (terminal deoxynucleotidyl transferase-mediated dUTP nick end labeling) assays (data not shown).

PDGFR α mutations are oncogenic *in vivo* and drive development of HGGs

To determine whether expression of wild-type PDGFR α or PDGFR α mutants renders p53-null PMAs tumorigenic, 2 million cells were implanted intracranially into the parietal lobe of 2-month-old athymic nude mice and monitored daily for signs of morbidity. Mice were euthanized and brain tumor tissue was collected when mice became symptomatic. Brain tumors formed in 100% of mice implanted with cells expressing any of six different mutant forms of PDGFR α . Only one of 19 mice implanted with cells overexpressing wild-type PDGFR α developed a brain tumor and none of the empty vector control cells developed into brain tumors *in vivo*. The tumor-induced morbidity was detected between 23 and 72 days for all mutants except E7del-expressing tumors, which occurred at 103 to 119 days and the single tumor from wild-type PDGFR α , which occurred at 120 days after implantation (Fig. 4A). Standard histopathologic preparations from all tumors were evaluated (by D.W. Ellison) and classified according to WHO criteria (Fig. 4B and Supplementary Table S3). Overall, 63% (44 of 70) of tumors analyzed were grade 3 anaplastic astrocytoma, 26%

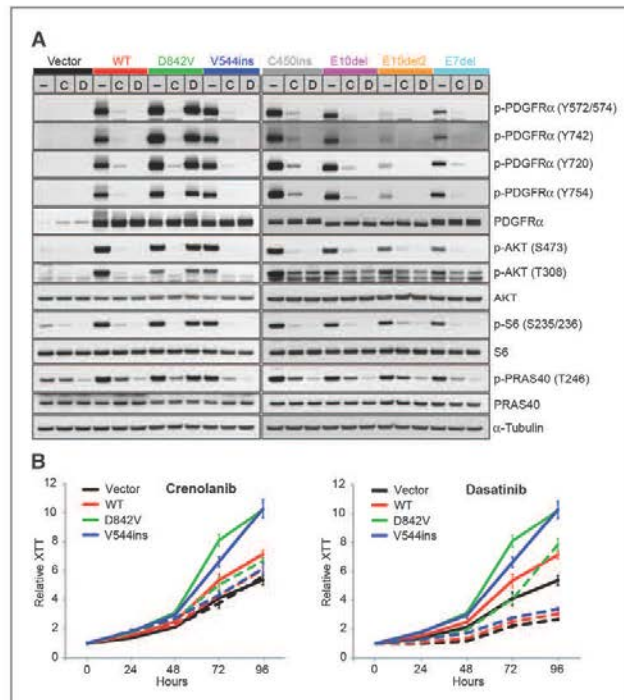


Figure 3. Wild-type and mutant PDGFR α are sensitive to small-molecule inhibitors. **A**, Western blot analysis of whole cell lysates from p53-null PMAs transfected with retroviruses expressing wild-type PDGFR α (WT), PDGFR α mutants, or empty vector. Following overnight serum starvation, cells were pretreated with crenolanib (C), dasatinib (D), or vehicle (-) for 3 hours and then stimulated with PDGF-AA for 30 minutes. Inhibition of PDGFR α signaling was analyzed using indicated antibodies. **B**, growth of p53-null PMAs transfected with retroviruses expressing wild-type, PDGFR α mutants, or empty vector was measured with XTT assay in triplicate in the presence of crenolanib (left graph), dasatinib (right graph), or vehicle. Single dose of inhibitors (dashed lines) or vehicle (solid lines) was administered and growth was measured. Error bars show SD. Representative figures of three independent experiments are shown.

(18 of 70) were grade 3 anaplastic oligoastrocytoma, 10% (7 of 70) were grade 4 glioblastoma, and the single tumor that developed from cells expressing wild-type PDGFR α was the only low-grade tumor, a grade 2 oligoastrocytoma. Tumors expressing the D842V or V544ins mutation were significantly associated with anaplastic astrocytoma histology ($P < 0.0001$ and $P = 0.0012$, respectively) and tumors with E10del were significantly associated with anaplastic oligoastrocytoma histology ($P = 0.002$). The rest of the evaluated tumors expressing C450ins, E10del2, and E7del mutations were not significantly associated with a specific morphology ($P > 0.2$).

Immunohistochemical analysis for PDGFR α showed strong overexpression of the receptor throughout all tumors. PDGFR α -driven murine gliomas showed growth patterns similar to human HGGs including an easily visualized focal mass as well as diffuse infiltration into the normal brain parenchyma (Fig. 4C and Supplementary Fig. S6A). These tumors were immunopositive for p-4E-BP1 (Thr37/46) and p-Akt (Ser473), suggesting activated PDGFR α signaling (Fig. 4C). Furthermore, activation of PDGFR α in tumor tissues

was confirmed by Western blot analysis, which showed high levels of phosphorylated receptor on Tyr-572/574, Tyr-742, Tyr-720, and Tyr-754 in all tumors examined but not in normal brain (Fig. 4D). Moreover, phosphorylation of PI3K pathway components, SHP2 and Src family kinases were strongly elevated compared with normal cortex. In contrast, levels of total STAT3 and phosphorylated STAT3 were similarly increased in all analyzed tumors compared with normal brain tissue, which may indicate cell type-specific differential expression of STAT3, not necessarily selective activation of STAT3 signaling in tumor. There was no significant difference in levels of MAPK pathway activation observed between tumor tissues and normal brain (Fig. 4D and Supplementary Fig. S6B).

Gene expression signatures of PDGFR α -driven murine HGGs resemble human HGGs

To evaluate the similarity of the molecular signatures of PDGFR-driven murine HGGs to human disease, gene expression profiles were analyzed for representative mutant and

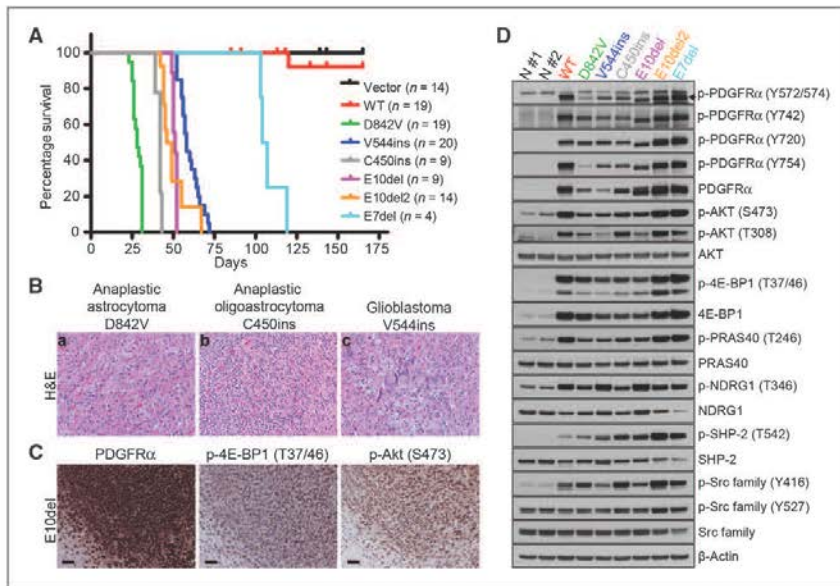


Figure 4. PDGFRα mutants are tumorigenic *in vivo*. A, Kaplan-Meier curves showing time to morbidity of mice following intracranial implantation of p53-null PMAs expressing wild-type PDGFRα (WT), PDGFRα mutants, or empty vector. B, diverse histologic phenotypes were generated in PDGFRα-driven HGGs. Most tumors diffusely infiltrated brain parenchyma while also producing circumscribed masses. a, representative anaplastic astrocytoma showing moderately pleomorphic tumor cells, some with astrocytic differentiation, diffusely infiltrating cerebral white matter (D842V; H&E $\times 200$); b, representative anaplastic oligoastrocytoma with admixed oligodendroglial and astrocytic phenotypes (C450ins; H&E $\times 200$); c, representative glioblastoma with focal giant cell phenotype (V544ins; H&E $\times 200$). C, immunohistochemical analysis of a representative brain tumor stained for PDGFRα, p-4E-BP1 (T37/46), p-Akt (S473), and counterstained with hematoxylin. Scale bar, 50 μ m. D, Western blot analysis of whole cell lysates from tissues of wild-type and mutant PDGFRα-driven brain tumors. Lysates from normal adult cortex (lanes N #1 and #2) were included as controls. Signaling pathway activation in PDGFRα-driven murine HGGs was assayed using the indicated antibodies.

wild-type PDGFRα generated brain tumors. In addition, gene expression profiles of EGFRvIII-driven murine HGGs were assayed for comparison (25). UHC segregated the tumors into two distinct subgroups (Supplementary Fig. S7). Interestingly, PDGFRα and EGFRvIII-expressing tumors were distributed between both subgroups and there was no significant association of specific PDGFRα mutations or tumor histopathology with either of the identified subgroups. Furthermore, using single sample GSEA the gene expression profile of each murine tumor was compared with published signature gene sets for human HGG subgroups (proneural, proliferative, and mesenchymal) and murine cell type-specific signatures (oligodendrocyte progenitor cells, oligodendrocytes, neurons, astrocytes, and cultured astroglia; refs. 30, 31). We also compared the gene expression signatures of seven independent untransduced p53-null PMA cultures, which showed a significant similarity to the published expression signatures of astrocytes and cultured astroglia (31). Interestingly, transformation of p53-null PMAs by PDGFRα

mutants resulted in gliomas with a range of expression signatures representing the three major expression subgroups observed in human HGGs (Fig. 5; ref. 30). Of note, 37.8% (14 of 37) of PDGFRα tumors showed proneural, 32.4% (12 of 37) proliferative, and 16.2% (6 of 37) mesenchymal subgroup expression signatures. There was no significant association of specific PDGFRα mutations with a particular expression subgroup described in pediatric and adult HGGs (4, 5, 30). Interestingly, tumors driven by E10del2 mutation were significantly associated with the oligodendrocyte progenitor cell-gene expression signature ($P = 0.03$); however, there was no significant association between specific mutation and expression signature of cell types.

Discussion

Aberrations of PDGFRα signaling via amplification and/or mutation of PDGFRA are frequent in pediatric HGGs. Here, targeted sequencing identified novel activating somatic mutations of PDGFRA in pediatric HGGs. Two recent studies

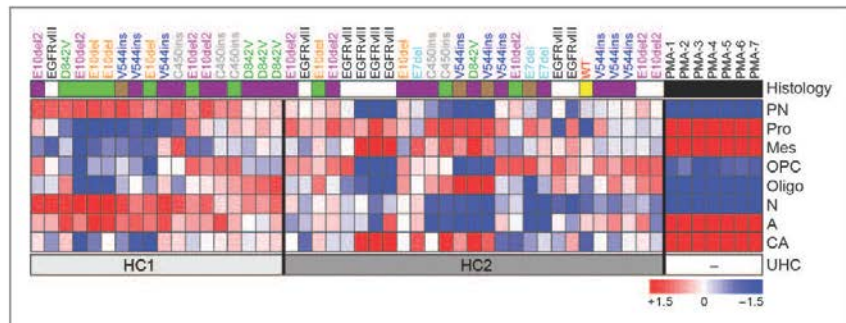


Figure 5. Gene expression profiles of murine PDGFR α -driven HGGs resemble human HGGs. Heat map of single sample GSEA of representative murine brain tumors and untransduced p53-null PMA cultures using gene sets defining human HGG expression subgroups (PN, proneural; Pro, proliferative; Mes, mesenchymal) and murine cell-type-specific signatures (OPC, oligodendrocyte progenitor cells; Oligo, oligodendrocyte; N, neurons; A, astrocytes; and CA, cultured astroglia). The order of the tumor samples is arranged on the basis of the UHC analysis using 1,000 most variable probe sets, which revealed two major expression subgroups HC1 and HC2 (Supplementary Fig. S7). The histology of each tumor is indicated above the heatmap: anaplastic astrocytoma (purple), anaplastic oligoastrocytoma (green), glioblastoma (brown), oligoastrocytoma (yellow), and white marks tumors with no diagnosis available. Untransduced p53-null PMA cultures are marked with black.

reported *PDGFR α* mutations in pediatric HGGs, but the functional consequence on PDGFR α activation was not explored and none of the described mutations overlapped with our findings (6, 12). Interestingly, mutations were not represented by a single hotspot as observed in gastrointestinal stromal tumors. One of the identified mutations in this study, V544ins, a small duplication resulting in a 14 amino acid in-frame insertion, was found in the transmembrane domain, potentially facilitating oligomerization of the receptor via the transmembrane domain or disrupting the inhibitory conformation of the juxtamembrane region and consequently leading to ligand-independent receptor dimerization and activation as observed for PDGFR β , c-Kit, and other RTKs (43–46).

In adult glioblastomas, the frequency of *PDGFR α* amplification and mutation is lower than in pediatric tumors and *PDGFR α* mutations comprise a distinct set of alterations compared with childhood disease (8, 9, 23, 34). Approximately 40% of amplified *PDGFR α* in adult glioblastomas is not wild-type, but contains recurrent genomic deletions of exons 8 and 9, resulting in an in-frame deletion in the extracellular domain and constitutive receptor activation. Several in-frame deletions occur in pediatric HGG; however, the recurrent alterations in adult glioblastoma were not found in pediatric tumors. Instead, mutations were not restricted to a single functional domain, but were distributed among different regions of the PDGFR α protein, including the kinase domain and extracellular regions involved in ligand binding and receptor-receptor interaction. The six PDGFR α mutants analyzed represented alterations in different functional domains, and all were constitutively active and tumorigenic. Thus, alternative genetic mechanisms are used to generate oncogenic mutations of *PDGFR α* in childhood and adult HGGs.

Numerous mouse glioma models have been generated by expression of exogenous PDGF, driving paracrine stimulation of cells expressing endogenous PDGFR (17–19, 46–48). In this study, gliomas were driven by the *PDGFR α* mutations found in pediatric HGGs. Both wild-type and mutant PDGFR α conferred a proliferative advantage to p53-null PMAs *in vitro*; however, the wild-type showed minimal tumorigenic activity, resulting in only 1 of 19 mice developing a low grade glioma, whereas all six of the mutants induced HGG formation with 100% penetrance. In contrast, a previous study showed that wild-type PDGFR α -transformed *Ink4a/Arf*-null PMAs to generate gliomas (49). The difference in the tumorigenic activity of wild-type PDGFR α may be explained in part by a higher propensity for transformation in the recipient cells, as the *Ink4a/Arf*-null PMAs formed tumors at lower frequency in the absence of PDGFR α , whereas the early passage p53-null PMAs in the present study did not form any tumors. In addition, *Ink4a/Arf* and p53 loss may differentially cooperate with PDGFR α overexpression to drive glioma (49). It is also possible that the levels of wild-type receptor expression in our model system were not sufficient to model the gene amplification that drives tumorigenesis in human tumors. However, the wild-type receptor was expressed at levels similar to the E7del, E10del, and E10del2 mutations (Fig. 2), which were also amplified in human tumors.

The fact that all PDGFR α mutations showed ligand-independent phosphorylation, whereas wild-type receptor activation required ligand stimulation, suggests lack of sufficient ligand in the brain of implanted adult animals, or different selective advantage conferred by wild-type *PDGFR α* amplification versus mutation with or without amplification. Amplification of wild-type *PDGFR α* occurred more frequently in tumors within the brainstem (26%, 11 of 43 DIPG vs. 11%, 9 of 84 nonbrainstem HGG, $P = 0.04$), whereas *PDGFR α* sequence

References

- Qaddoumi I, Sultan I, Gajjar A. Outcome and prognostic features in pediatric gliomas: a review of 6212 cases from the surveillance, epidemiology, and end results database. *Cancer* 2009;115:5761-70.
- Gottardo NG, Gajjar A. Chemotherapy for malignant brain tumors of childhood. *J Child Neurol* 2008;23:1149-59.
- Warren KE. Diffuse intrinsic pontine glioma: poised for progress. *Front Oncol* 2012;2:205.
- Paugh BS, Broniscer A, Qu C, Miller CP, Zhang J, Tatevossian RG, et al. Genome-wide analyses identify recurrent amplifications of receptor tyrosine kinases and cell-cycle regulatory genes in diffuse intrinsic pontine glioma. *J Clin Oncol* 2011;29:3999-4008.
- Paugh BS, Qu C, Jones C, Liu Z, Adamowicz-Brice M, Zhang J, et al. Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease. *J Clin Oncol* 2010;28:3061-8.
- Puget S, Philippe C, Bax DA, Job B, Varlet P, Junier MP, et al. Mesenchymal transition and PDGFRA amplification/mutation are key distinct oncogenic events in pediatric diffuse intrinsic pontine gliomas. *PLoS ONE* 2012;7:e30313.
- Zarghooni M, Bartels U, Lee E, Buczkowicz P, Morrison A, Huang A, et al. Whole-genome profiling of pediatric diffuse intrinsic pontine gliomas highlights platelet-derived growth factor receptor alpha and poly (ADP-ribose) polymerase as potential therapeutic targets. *J Clin Oncol* 2010;28:1337-44.
- McLendon R, Friedman A, Bigner D, Van Meir EG, Brat DJ, Mastrogiannis GM, et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 2008;455:1061-8.
- Parsons DW, Jones S, Zhang X, Lin JC, Lsary RJ, Angenendt P, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science* 2008;321:1807-12.
- Bax DA, Mackay A, Little SE, Carvalho D, Viana-Pereira M, Tambar N, et al. A distinct spectrum of copy number aberrations in pediatric high-grade gliomas. *Clin Cancer Res* 2010;16:3368-77.
- Wu G, Broniscer A, McEachron TA, Lu C, Paugh BS, Beckwith J, et al. Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat Genet* 2012;44:251-3.
- Schwartzentruber J, Korshunov A, Liu XY, Jones DT, Pfaff E, Jacob K, et al. Driver mutations in histone H3.3 and chromatin remodeling genes in paediatric glioblastoma. *Nature* 2012;482:226-31.
- Pollack IF, Hamilton RL, James CD, Finkelstein SD, Burnham J, Yates AJ, et al. Rarity of PTEN deletions and EGFR amplification in malignant gliomas of childhood: results from the Children's Cancer Group 945 cohort. *J Neurosurg* 2006;105:418-24.
- Funari FB, Fenton T, Bachoo RM, Mukasa A, Stommel JM, Stegh A, et al. Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes Dev* 2007;21:2683-710.
- Andrae J, Gallini R, Betscholtz C. Role of platelet-derived growth factors in physiology and medicine. *Genes Dev* 2008;22:1276-312.
- Heidin CH, Ostman A, Ronnstrand L. Signal transduction via platelet-derived growth factor receptors. *Biochim Biophys Acta* 1998;1378: F79-113.
- Uhrbom L, Hesselager G, Nister M, Westermarck B. Induction of brain tumors in mice using a recombinant platelet-derived growth factor B-chain retrovirus. *Cancer Res* 1998;58:5275-9.
- Dai C, Celestino JC, Okada Y, Louis DN, Fuller GN, Holland EC. PDGF autocrine stimulation dedifferentiates cultured astrocytes and induces oligodendrogliomas and oligoastrocytomas from neural progenitors and astrocytes *in vivo*. *Genes Dev* 2001;15:1913-25.
- Lindberg N, Kastemar M, Olofsson T, Smits A, Uhrbom L. Oligodendrocyte progenitor cells can act as cell of origin for experimental glioma. *Oncogene* 2005;28:2266-75.
- Hesselager G, Uhrbom L, Westermarck B, Nister M. Complementary effects of platelet-derived growth factor autocrine stimulation and p53 or ink4a-Arf deletion in a mouse glioma model. *Cancer Res* 2003;63: 4305-9.
- Hambardzumyan D, Parada LF, Holland EC, Charest A. Genetic modeling of gliomas in mice: new tools to tackle old problems. *Glia* 2011;59:1155-68.
- Torchia EC, Boyd K, Rehg JE, Qu C, Baker SJ. BWS/FLI-1 induces rapid onset of myeloid/erythroid leukemia in mice. *Mol Cell Biol* 2007;27:7918-34.
- Ozawa T, Brennan CW, Wang L, Squatrito M, Sasayama T, Nakada M, et al. PDGFRA gene rearrangements are frequent genetic events in PDGFRA-amplified glioblastomas. *Genes Dev* 2010;24:2205-18.
- Persons DA, Allay JA, Allay ER, Ashmun RA, Orlie D, Jane SM, et al. Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis. *Blood* 1999;93:488-99.
- Enderby R, Zhu X, Hay N, Ellison DW, Baker SJ. Nonredundant functions for Akt isoforms in astrocyte growth and gliomagenesis in an orthotopic transplantation model. *Cancer Res* 2011;71:4106-16.
- Louis DN, Ohgaki H, Westler OD, Cavenee WK, Burger PC, Jouvet A, et al. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* 2007;114:97-109.
- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;8: 118-27.
- Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 2009;462:108-12.
- Chow LM, Enderby R, Zhu X, Rankin S, Qu C, Zhang J, et al. Cooperativity within and among Pten, p53, and Rb pathways induces high-grade astrocytoma in adult brain. *Cancer Cell* 2011;19:305-16.
- Phillips HS, Kharbada S, Chen R, Forrest WF, Soriano RH, Wu TD, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 2006;9:157-73.
- Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, Christopherson KS, et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci* 2008;28:264-78.
- Corless CL, Schroeder A, Griffith D, Town A, McGreevey L, Harrell P, et al. PDGFRA mutations in gastrointestinal stromal tumors: frequency, spectrum and *in vitro* sensitivity to imatinib. *J Clin Oncol* 2005;23: 5357-64.
- Medeiros F, Corless CL, Duensing A, Hornick JL, Oliveira AM, Heinrich MC, et al. KIT-negative gastrointestinal stromal tumors: proof-of-concept and therapeutic implications. *Am J Surg Pathol* 2004;28: 889-94.
- Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 2010;17:98-110.
- Kumabe T, Sohma Y, Kayama T, Yoshimoto T, Yamamoto T. Overexpression and amplification of alpha-PDGFR gene lacking exons coding for a portion of the extracellular region in a malignant glioma. *Tohoku J Exp Med* 1992;168:265-9.
- Zhang SQ, Yang W, Kontaridis MI, Bivona TG, Wen G, Araki T, et al. Shp2 regulates SRC family kinase activity and Ras/Erk activation by controlling Csk recruitment. *Mol Cell* 2004;13:341-55.
- Bazenet CE, Gelderbloos JA, Kazlauskas A. Phosphorylation of tyrosine 720 in the platelet-derived growth factor alpha receptor is required for binding of Grb2 and SHP-2 but not for activation of Ras or cell proliferation. *Mol Cell Biol* 1996;16:8926-36.
- Yu JC, Heidaran MA, Pierce JH, Gutkind JS, Lombardi D, Ruggiero M, et al. Tyrosine mutations within the alpha platelet-derived growth factor receptor kinase insert domain abrogate receptor-associated phosphatidylinositol-3 kinase activity without affecting mitogenic or chemotactic signal transduction. *Mol Cell Biol* 1991;11:3780-5.
- Lombardo LJ, Lee FY, Chen P, Norris D, Barrish JC, Behnia K, et al. Discovery of N-(2-chloro-6-methyl-phenyl)-2-(6-(4-(2-hydroxyethyl)-piperazin-1-yl)-2-methylpyrimidin-4-ylamino)thiazole-5-carboxamide (BMS-354825), a dual Src/Abl kinase inhibitor with potent antitumor activity in preclinical assays. *J Med Chem* 2004;47:8658-61.
- Carter TA, Wodicka LM, Shah NP, Velasco AM, Fabian MA, Treiber DK, et al. Inhibition of drug-resistant mutants of ABL, KIT, and EGF receptor kinases. *Proc Natl Acad Sci U S A* 2005;102:11011-6.

41. Lewis NL, Lewis LD, Eder JP, Reddy NJ, Guo F, Pierce KJ, et al. Phase I study of the safety, tolerability, and pharmacokinetics of oral CP-868,596, a highly specific platelet-derived growth factor receptor tyrosine kinase inhibitor in patients with advanced cancers. *J Clin Oncol* 2009;27:5262-9.
42. Heinrich MC, Griffith D, McKinley A, Patterson J, Presnell A, Ramachandran A, et al. Crenolanib inhibits the drug-resistant PDGFRA D842V mutation associated with imatinib-resistant gastrointestinal stromal tumors. *Clin Cancer Res* 2012;18:4375-84.
43. Oates J, King G, Dixon AM. Strong oligomerization behavior of PDGFbeta receptor transmembrane domain and its regulation by the juxtamembrane regions. *Biochim Biophys Acta* 2010;1798:605-15.
44. Iruela PM, Luo Y, Bakht O, Lal CC, Smith SO, DiMaio D. Definition of an inhibitory juxtamembrane WW-like domain in the platelet-derived growth factor beta receptor. *J Biol Chem* 2002;277:38627-34.
45. Hirota S, Isozaki K, Moriyama Y, Hashimoto K, Nishida T, Ishiguro S, et al. Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science* 1998;279:577-80.
46. Becher OJ, Hambarzumyan D, Walker TR, Helmy K, Nazarian J, Albrecht S, et al. Preclinical evaluation of radiation and perifosine in a genetically and histologically accurate model of brainstem glioma. *Cancer Res* 2010;70:2548-57.
47. Lai L, Sonabend AM, Guarnieri P, Soderquist C, Ludwig T, Rosenfeld S, et al. Glioblastoma models reveal the connection between adult glial progenitors and the proneural phenotype. *PLoS ONE* 2011;6:e20041.
48. Jackson EL, Garcia-Verdugo JM, Gil-Perotin S, Roy M, Quinones-Hinojosa A, VandenBerg S, et al. PDGFR alpha-positive B cells are neural stem cells in the adult SVZ that form glioma-like growths in response to increased PDGF signaling. *Neuron* 2006;51:187-99.
49. Liu KW, Feng H, Bachoo R, Kazlauskas A, Smith EM, Symes K, et al. SHP-2/PTPN11 mediates gliomagenesis driven by PDGFRA and INK4A/ARF aberrations in mice and humans. *J Clin Invest* 2011;121:905-17.
50. Bachoo RM, Maher EA, Ligon KL, Sharpless NE, Chan SS, You MJ, et al. Epidermal growth factor receptor and Ink4a/Arf: convergent mechanisms governing terminal differentiation and transformation along the neural stem cell to astrocyte axis. *Cancer Cell* 2002;1:269-77.
51. Kaul A, Chen YH, Emmett RJ, Dahiya S, Gutmann DH. Pediatric glioma-associated KIAA1549: BRAF expression regulates neuroglial cell growth in a cell type-specific and mTOR-dependent manner. *Genes Dev* 2012;26:2561-6.
52. Tchougounova E, Kastemar M, Brasater D, Holland EC, Westermark B, Ulröbom L. Loss of Arf causes tumor progression of PDGFB-induced oligodendroglioma. *Oncogene* 2007;26:6289-96.
53. Peulikas P, Rosen N. Mutant BRAF^{V600E} melanomas-dependence and resistance. *Cancer Cell* 2011;19:11-5.
54. Engelman JA, Janne PA. Mechanisms of acquired resistance to epidermal growth factor receptor tyrosine kinase inhibitors in non-small cell lung cancer. *Clin Cancer Res* 2008;14:2895-9.
55. Vivanco I, Robins HI, Rohle D, Campos C, Grommes C, Nghiemphu PL, et al. Differential sensitivity of glioma- versus lung cancer-specific EGFR mutations to EGFR kinase inhibitors. *Cancer Discov* 2012;2:458-71.

Supplementary Figure Legends

Figure-S1: Missense mutation resulting in alternative splicing and RT-PCR screens for *KDR-PDGFR*A fusion and *PDGFR*A^{Δ8, 9}

- A) Sanger sequencing of *PDGFR*A from tumor and normal reference sample identified a heterozygous somatic missense mutation in genomic DNA from HGG028 and 2HGG171T (N468S). Subsequent validation of expression of the mutated allele by sequencing cDNA amplified by RT-PCR revealed deletion of 13 amino acids, indicating that this point mutation lead to generation of an alternative splicing site and subsequent deletion of amino acids 456-468 within exon 10 (E10del2) rather than encoding an amino acid substitution (Table S1). Example sequencing chromatograms are shown for HGG028 for genomic DNA and cDNA.
- B) RT-PCR screen for gene fusion between *PDGFR*A and *KDR* previously described in a single case of adult HGG identified a single case of pediatric HGG with this rearrangement. The sequence surrounding the fusion site of *KDR-PDGFR*A is shown. The *KDR-PDGFR*A fusion transcript is an in-frame fusion of *KDR* exon 13 to *PDGFR*A exon 10, with an intervening cryptic exon from *KDR* intron 13 as previously reported.
- C) In-frame deletion of 243 base pairs (bp) in exons 8 and 9 of the extracellular portion of the receptor (*PDGFR*A^{Δ8, 9}) previously described in adult HGGs was not found in RT-PCR screen of 83 non-brainstem pediatric HGG, and 57 DIPG. An example DNA gel is shown. For all analyzed cases 885 bp wild-type RT-PCR product was detected and no Δ8,9 was found (642 bp). For HGG159 a smaller mutant band was amplified

(807bp), which corresponds to E10del mutation found by Sanger sequencing (Table S1).

Figure-S2: Analysis of PDGFR α downstream effectors

Western blot analysis of whole cell lysates from p53 null PMAs transduced with retroviruses expressing wild type PDGFR α (WT), PDGFR α mutants or empty vector. Cells were serum starved overnight and then treated with PDGF-AA or vehicle for 30 minutes. Activation of PDGFR α downstream signaling targets was analyzed using specified antibodies in the presence and absence of PDGF-AA. The top immunoblot showing p-PDGFR α (Y720) and α -tubulin loading control are the same as on Figure 2.

Figure-S3: Mutant PDGFR α is sensitive to small molecule inhibitors

- A) Whole cell lysates from p53 null PMAs transduced with retroviruses expressing wild type PDGFR α (WT), PDGFR α mutants or empty vector were subjected to western blot analysis. Serum starved cells were pre-treated with vehicle (-), Crenolanib (C), or Dasatinib (D) for 3 hours and then stimulated with PDGF-AA for 30 minutes. Inhibition of PDGFR α signaling was analyzed using indicated antibodies. The top immunoblots showing p-PDGFR α (Y720) and α -tubulin are the same as on Figure 3.
- B) Growth of p53 null PMAs transduced with retroviruses expressing PDGFR α mutants (C450ins, E10del, E10del2 and E7del) was measured with XTT assay in triplicate in the presence of crenolanib (left graph), dasatinib (right graph) or vehicle. Single

dose of inhibitors (dashed lines) or vehicle (solid lines) was administered and growth was measured. Error bars show standard deviation.

Figure-S4: Dose-dependent sensitivity to crenolanib

Wild type PDGFR α , D842V or V544ins expressing p53 null PMAs were serum starved overnight and then pre-treated with indicated concentrations of crenolanib or vehicle for 3 hours and stimulated with PDGF-AA for 30 minutes. The dose-dependent inhibitory effect of crenolanib on phosphorylation of PDGFR α was analyzed by western blot using indicated antibodies. The IC₅₀ and IC₉₀ concentrations were calculated using densitometry. α -Tubulin is shown as a loading control.

Figure-S5: Crenolanib and dasatinib induce G0/G1 arrest

Cell cycle analysis of fixed PDGFR α or empty vector expressing cells following 24 hour vehicle (V), crenolanib (C), or dasatinib (D) treatment. Cell cycle analysis was performed by flow cytometry. Each data point is an average from triplicate measurement. Both inhibitors caused cell cycle arrest in G0/G1. Error bars show standard deviation. Student's two-tailed t-test p-values are shown for comparison to vehicle treated controls.

Figure-S6: Signaling in PDGFR α -driven HGGs

A) Immunohistochemistry of whole brain sections of two representative PDGFR α -driven gliomas stained for PDGFR α , which clearly marks focal tumor mass. Higher

magnification images illustrate invasiveness of both tumors, a common feature of human HGGs. Brain sections were counterstained with hematoxylin.

- B) Western blot analysis of whole cell lysates from tissues of wild-type and mutant PDGFR α -driven brain tumors. Lysates from normal adult cortex (lanes N #1 and #2) were included as controls. Signaling pathway activation in PDGFR α -driven murine HGGs was assayed using the indicated antibodies. The top immunoblots showing p-PDGFR α (Y720) and PDGFR α and β -actin are the same as shown on Figure 4.

Figure-S7: PDGFR α -driven murine HGGs show a range of gene expression signatures

Gene expression profiles generated from murine PDGFR α -driven tumors and EGFRvIII-expressing tumors were analyzed by unsupervised hierarchical clustering using the top 1000 most variable probe sets selected using the median absolute deviation (MAD) scores. The dendrogram is shown on top with two main clusters identified, HC1 and HC2. The heat map is featuring all 1000 selected probes. The histology of each tumor is indicated above the heat map: anaplastic astrocytoma (purple), anaplastic oligoastrocytomas (green), glioblastoma (brown), one oligoastrocytoma (yellow) and white bars mark tumors with no diagnosis available. Tumors driven by EGFRvIII transformation of p53-null PMAs were previously shown to be astrocytomas of grade III or grade IV (glioblastoma).

Supplementary Tables

Table S1: *PDGFRA* mutations in pediatric HGG

Table S2: Clinical sample information, PDGFR α signaling alterations and *TP53* mutations in pediatric HGG

Table S3: Histopathology of PDGFR α -driven murine tumors

Table S4: *PDGFRA* sequencing primers

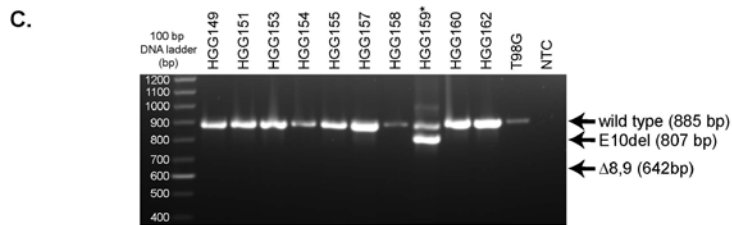
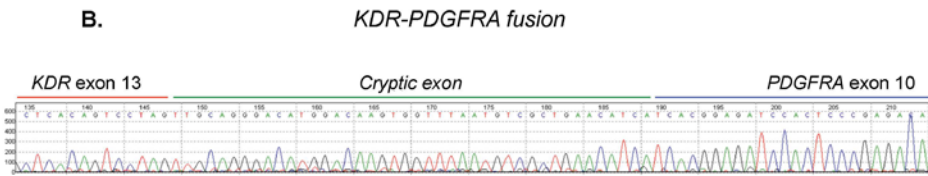
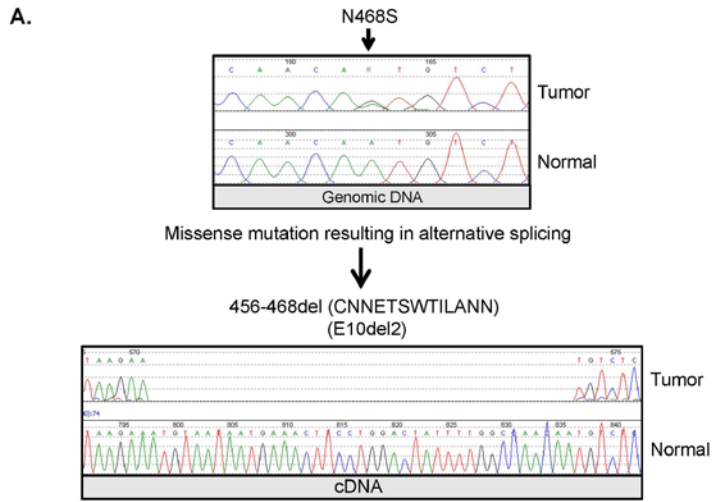


Figure S1

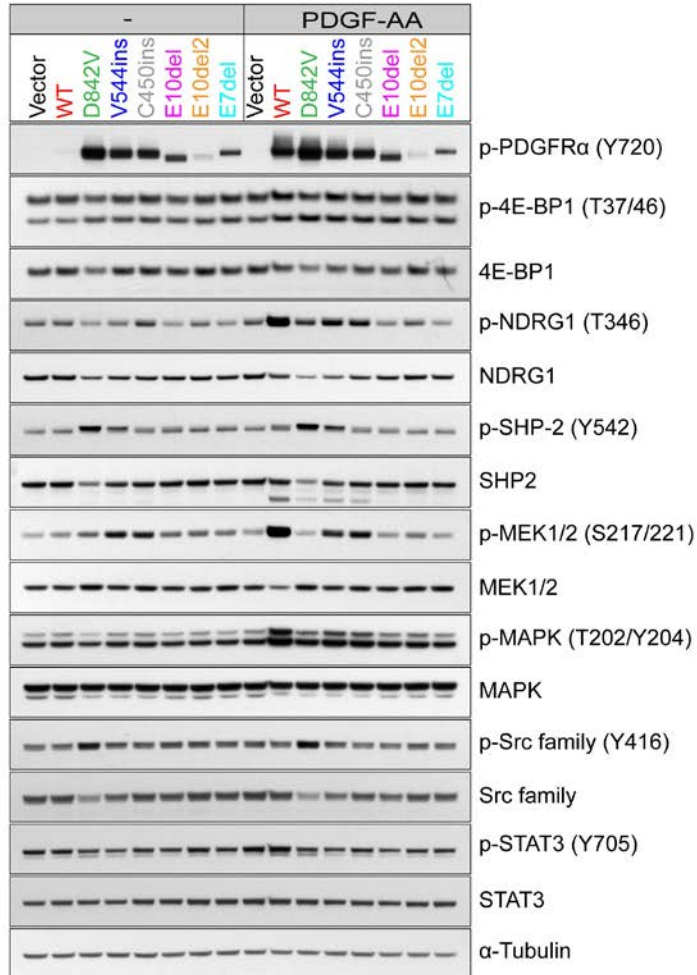


Figure S2

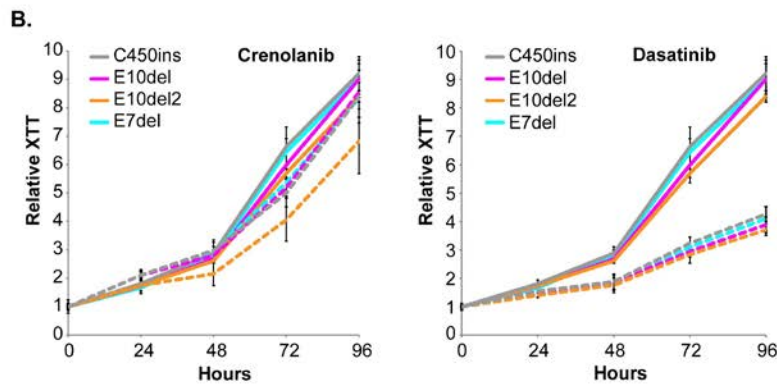
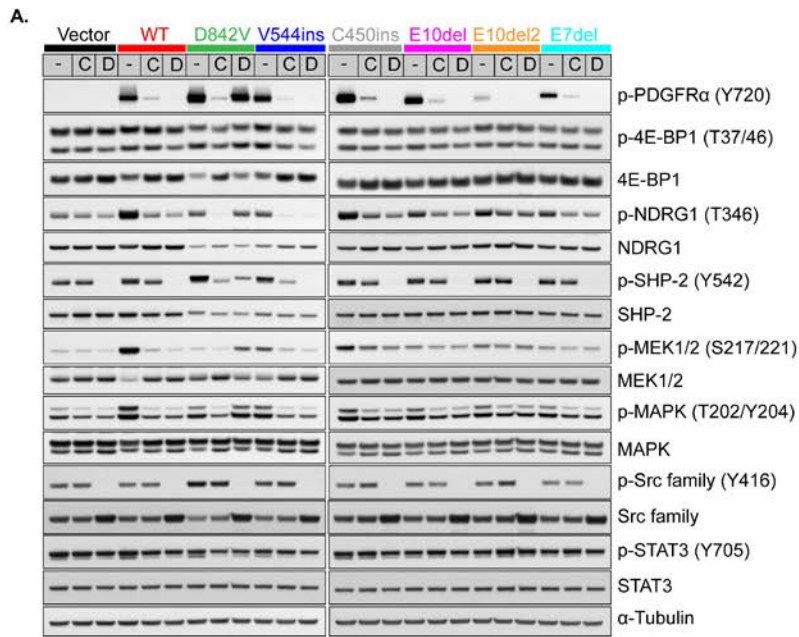


Figure S3

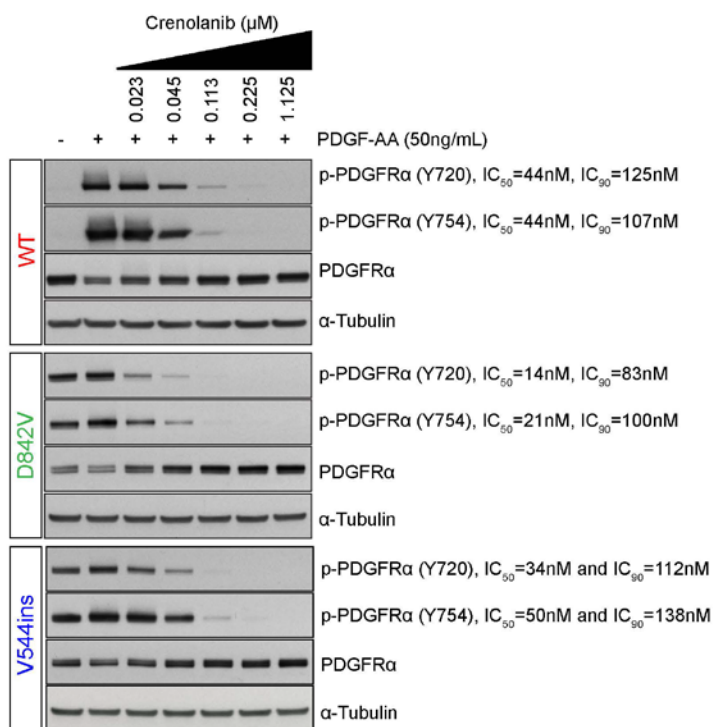


Figure S4

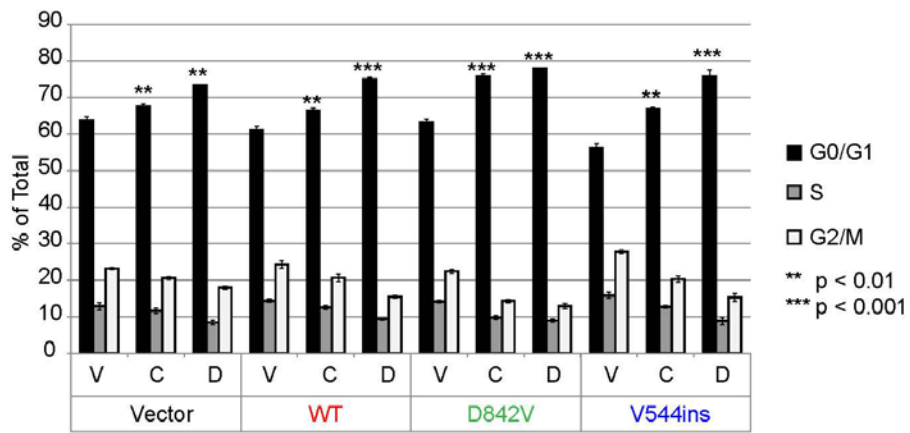


Figure S5

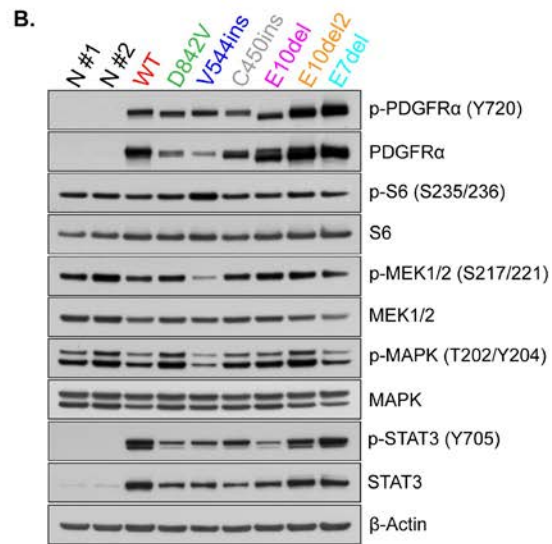
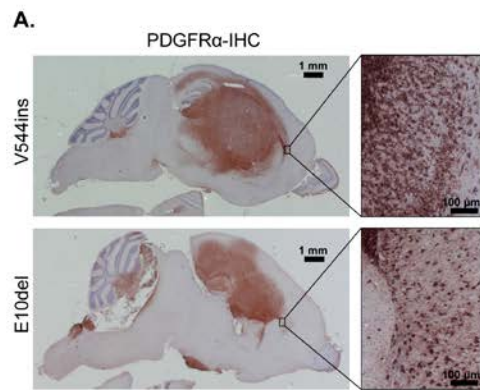


Figure S6

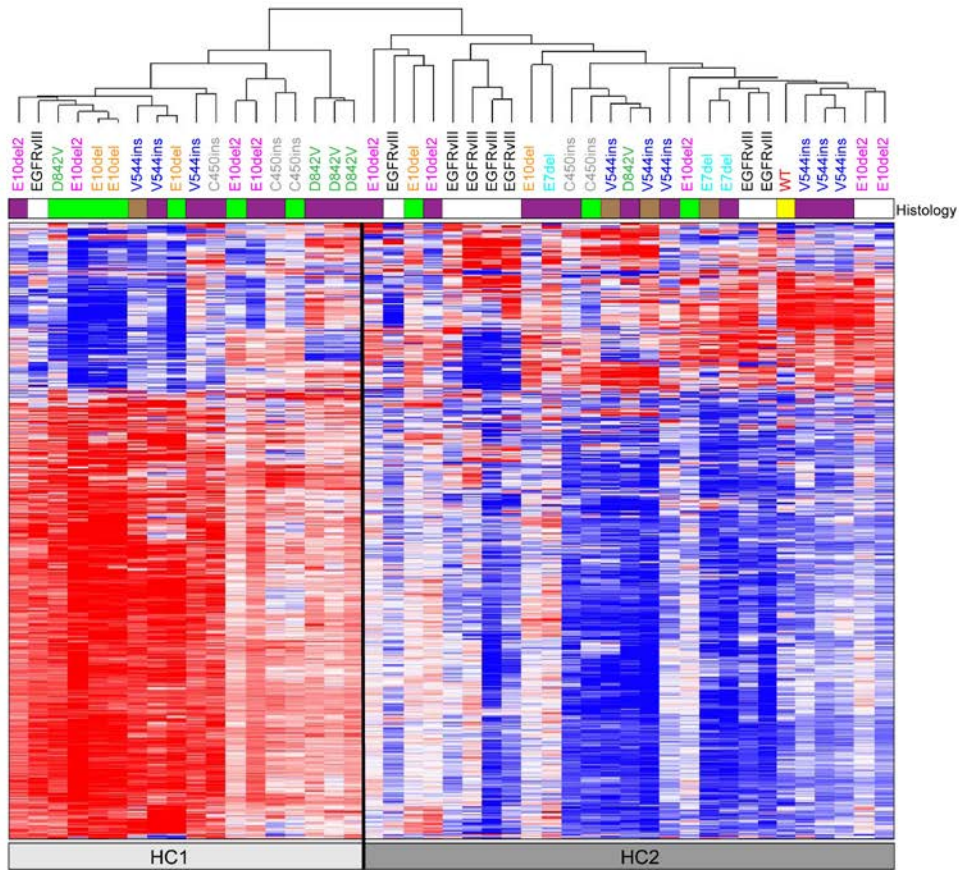


Figure S7

RESEARCH

Open Access

The prognostic role of intragenic copy number breakpoints and identification of novel fusion genes in paediatric high grade glioma

Diana Carvalho^{1,2,3}, Alan Mackay¹, Lynn Bjerke¹, Richard G Grundy⁴, Celeste Lopes², Rui M Reis^{3,5} and Chris Jones^{1*}

Abstract

Background: Paediatric high grade glioma (pHGG) is a distinct biological entity to histologically similar tumours arising in older adults, and has differing copy number profiles and driver genetic alterations. As functionally important intragenic copy number aberrations (iCNA) and fusion genes begin to be identified in adult HGG, the same has not yet been done in the childhood setting. We applied an iCNA algorithm to our previously published dataset of DNA copy number profiling in pHGG with a view to identify novel intragenic breakpoints.

Results: We report a series of 288 iCNA events in pHGG, with the presence of intragenic breakpoints itself a negative prognostic factor. We identified an increased number of iCNA in older children compared to infants, and increased iCNA in *H3F3A* K27M mutant tumours compared to G34R/V and wild-type. We observed numerous gene disruptions by iCNA due to both deletions and amplifications, targeting known HGG-associated genes such as *RB1* and *NF1*, putative tumour suppressors such as *FAF1* and *KIDINS220*, and novel candidates such as *PTPRE* and *KCND2*. We further identified two novel fusion genes in pHGG – *CSGALNACT2:RET* and the complex fusion *DHX57:TMEM178:MAP4K3*. The latter was sequence-validated and appears to be an activating event in pHGG.

Conclusions: These data expand upon our understanding of the genomic events driving these tumours and represent novel targets for therapeutic intervention in these poor prognosis cancers of childhood.

Keywords: Fusion, Paediatric, Glioblastoma, Copy number, Intragenic

Background

DNA copy number and gene expression studies have highlighted key distinctions between high grade gliomas (HGG) arising in childhood and far more commonly, much later in adult life [1-4]. Indeed, recent exome-level sequencing initiatives have conclusively shown the existence of subgroups of HGG marked by distinct driver mutations [5], which are significantly enriched in young children (*H3F3A* K27M), teenagers and young adults (*H3F3A* G34R/V), and middle-aged adults (*IDH1/2*) [6]. Specific driving events for infants and elderly patients remain to be elucidated, however they too represent biological sub-entities, with infants having few genomic alterations [4], and elderly patients harbouring frequent amplification of *EGFR* and other genomic events [2,3].

The identification of driving genetic alterations at the DNA copy level are necessarily most commonly focussed on assessing the amplification/deletion of genes in their entirety, and approaches to ascribe significance to genomic events make use of overlapping regions across multiple samples to find genes consistently within regions of gain/loss [7]. This approach has the result of ignoring genes for whom the breakpoint, *i.e.* the specific location of copy number change, is found within the coding regions. Such events may be more than mere bystanders of the “driving” aberration, and may themselves play significant roles in tumour initiation and maintenance.

One key implication of copy number breakpoints occurring within genes is the possibility of generating novel fusions. Gene fusions can occur through both intra- and inter-chromosomal translocations, bringing together coding regions from two or more genes within a single reading frame allowing expression of a novel protein. Such gene fusions are common in cancer, but

* Correspondence: chris.jones@icr.ac.uk

¹Divisions of Molecular Pathology and Cancer Therapeutics, Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK

Full list of author information is available at the end of the article



have historically been thought to be largely restricted to haematological malignancies and selected solid tumours such as sarcomas. Recent evidence has overturned this, with numerous novel gene fusions being discovered in a wide range of cancer types, exemplified by the identification of common *TMPRSS2:ERG* fusions in prostate cancer [8] and the *EML4:ALK* fusion in non-small cell lung cancer [9].

The first fusion gene found in glioblastoma was the rearrangement located at an amplified region at chromosome 4q12, resulting in the fusing of the kinase domain of *PDGFRA* with the regulatory domains of *KDR* (*VEGFR2*) [10]. This *KDR:PDGFRA* was found to be activating and tumorigenic, however to date only a single additional case has been found, in a paediatric high grade glioma (pHGG) [11], and thus these fusions do not represent a common event. Another low frequency fusion has more recently been identified in approximately 3% of adult HGG, involving *FGFR1* or *FGFR3* partnering with *TACC1* or *TACC3* [12]. These *FGFR:TACC* fusions have been shown to localize to mitotic spindle poles, have constitutive kinase activity and induce mitotic and chromosomal segregation defects and aneuploidy [12]. The types of integrated analysis that identified these mutations have also begun to identify more common rearrangements, such as numerous fusions involving *EGFR*, the most frequently seen partner producing the *EGFR-SEPT14* fusion demonstrated to activate STAT3 signaling and confer mitogen independence and sensitivity to EGFR inhibition [13].

Such analyses are clearly proving extremely valuable in furthering our understanding of HGG biology and generating novel targets for therapeutic intervention. As similar approaches are yet to be undertaken in the paediatric setting, we have applied an algorithm designed to identify intragenic copy number breakpoints in our previously published study of DNA copy number [4]. We identify numerous potentially functional gene disruptions and a novel validated complex fusion, *DHX57:TMEM178:MAP4K3*.

Methods

Published DNA copy number data

We previously carried out a DNA copy number profiling study of 100 pHGG cases on Affymetrix 500 K SNP arrays [4]. The data have been deposited at the Gene Expression Omnibus (GEO, www.ncbi.nlm.nih.gov/geo/) with accession number GSE19578. Copy number assignment was carried out as per the original publication, using Affymetrix Genotyping Analysis Software (GTTYPE version 4.s) improved using Bayesian Linear Model with Mahalanobis distance classifier algorithm (BRLMM) and standard dChipSNP normalization and model-based expression algorithms [4]. Log₂-transformed data was used for all subsequent analysis in the present study.

iCNA algorithm

We implemented the iCNA algorithm developed as part of the GTS package under R2.11.0 (cbio.mskcc.org/~brennan) [14]. Breakpoints are calculated according to user-defined 'delta' values representing shifts in log₂ ratios between two contiguous genomic regions after segmenting the copy number data using circular binary segmentation (cbs) [15]. Using a delta of 0.4, breakpoint boundaries are identified and errors estimated by permutation-based calculations of neighbouring probe data. Confidence intervals are assigned and those falling within the 95% window considered 'high confidence'. An estimate is calculated for the expected rate of breaks for each gene based upon gene size and rate of breaks per sample, with a p value obtained based upon (observed-expected)/standard error. A corrected p value of <0.05 is considered significant. Manual inspection of copy number plots was undertaken to ensure sufficient probe coverage was present at identified loci in order to prioritise the most convincing breakpoints. Those with substantial gaps at either side of the break were excluded.

Custom oligonucleotide array CGH

We designed two fine-tiling oligonucleotide microarrays to cover the specific amplicons observed at chromosome 2p22.1 and 10q11.21. This was undertaken using the Agilent custom array design tool e-Array (Agilent, Santa Clara, CA, USA; <https://earray.chem.agilent.com/>), and comprised 700 probes covering 43.56–43.70 Mb on chromosome 10 and 5000 probes covering 39–40 Mb on chromosome 10 with a median probe interval of 200 bp on 2x105K microarray. Due to limited amount of material, DNA was whole genome amplified (WGA) using the GenomePlex® Complete Whole Genome Amplification Kit (Sigma, Gillingham, UK) starting with 10 ng of sample and control DNA, and following the manufacturer's protocol. WGA DNA was labelled using the Agilent Genomic DNA ULS labelling kit, hybridised as per manufacturer's instructions, and scanned on the Agilent 2505B Microarray Scanner System. Data has been submitted to ArrayExpress with accession number E-MTAB-2340.

siRNA knockdown

siRNA was carried out using a Dharmacon SMARTpool™ (Dharmacon, Lafayette, CO, USA) against *MAP4K3* (#003588) with paediatric glioma cells SF188, KNS42, UW479, Res259 and Res186 [16] and a panel of breast carcinoma lines. Cells were plated and transfected 24 hours later with siRNA using Lipofectamine RNAiMax™ (Invitrogen, Paisley, UK) as per manufacturer's instructions, alongside transfections of siControl. Twenty four hours following transfection, cells were trypsinised and media replenished after 48 hours and 96 hours, with cell viability assessed after seven days using CellTiter-Glo™ Luminescent Cell Viability

Assay (Promega, Madison, WI, USA) as per manufacturer's instructions.

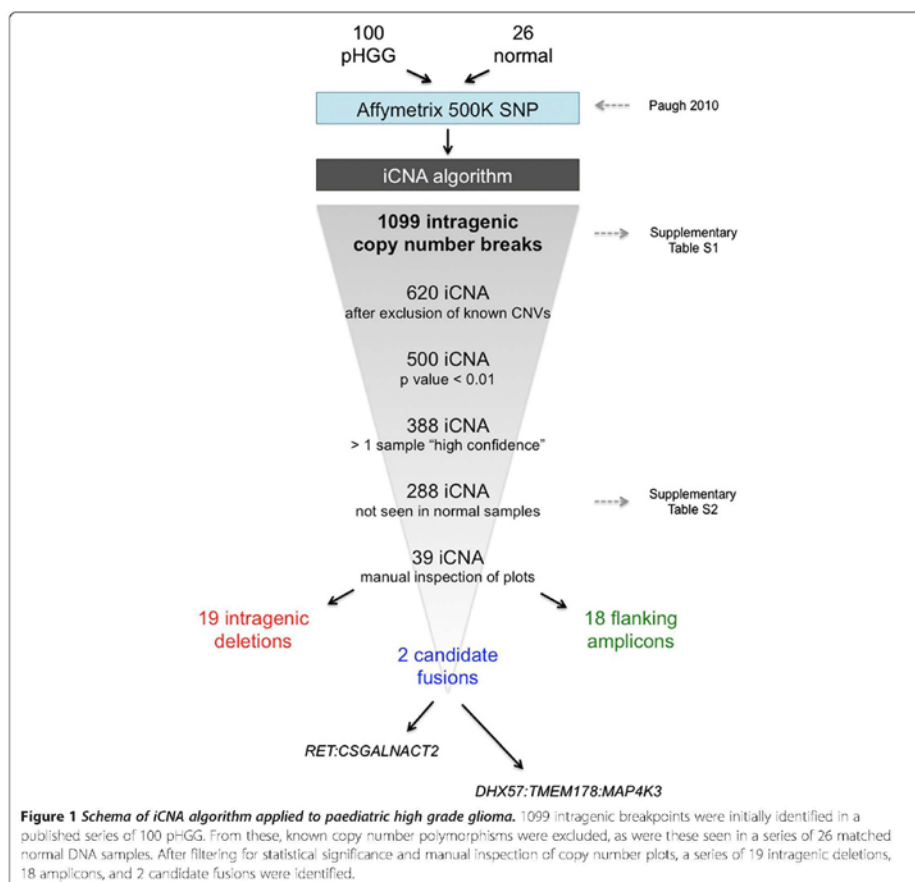
Results

Intragenic copy number breakpoints in paediatric HGG

We have previously carried out DNA copy number profiling of a large series of pHGG samples using Affymetrix 500 K SNP arrays, and reported numerous genes encompassed within areas of focal amplification and deletion [4]. We now applied an algorithm (iCNA [14]) designed to identify copy number breakpoints contained within the sequence of known genes. A full schema of the analytical process is given in Figure 1.

This algorithm was applied to 100 pHGG and 26 matched normal DNA samples, resulting in the identification of 1099 unique DNA copy number breaks contained within gene sequences across all tumour samples (Additional file 1: Table S1). Of these, 479 were found to map to known regions of copy number variations found commonly in the germlines of the general population by cross-referencing the breakpoints with The Centre for Applied Genomics Database of Genomic Variants [17], leaving a total of 620 events.

These were filtered to 500 after excluding those with p values > 0.05, and further reduced to 388 with at least one sample harbouring a given aberration at 'high



confidence'. A further 100 of these were excluded as they were also found in at least one of the normal samples profiled, representing either technical artifacts associated with the array platform used, or low frequency normal copy number polymorphisms. A final list of 288 iCNA is provided in Additional file 2: Table S2.

Most pHGG samples harboured at least one iCNA (median = 3), although seven cases were found to contain none. Several cases were found to contain many more aberrations (maximum = 19), though these were in the minority (Figure 2a). The number of iCNA events per sample was found to be prognostic in this multi-institutional series of cases, with pHGG containing more than 10 iCNA (n = 9) found to have a significantly poorer survival (median = 7.8 months), and those with no iCNA a better survival (median = 24 months) than the rest of the tumours (median = 13.2 months) (p = 0.026, log-rank test) (Figure 2b).

There were no differences in the number of iCNA between grade III (n = 20, median = 3, range 0–7) and grade IV tumours (n = 58, median = 3, range 0–19) (p = 0.456, t-test), though the cases with the highest number of iCNA were all grade IV glioblastoma (Figure 2c). Similarly, there were no differences between primary tumours (n = 68, median = 3, range 0–19) and those which arose as secondary

malignancies after cranio-spinal irradiation (n = 10, median = 2.5, range 0–11) (p = 0.698, t-test) (Figure 2d). Infants (less than 3 years at diagnosis) had significantly fewer iCNA (n = 10, median = 2.5, range 0–6) than older children (n = 68, median = 4, range 0–19) (p = 0.050, t-test) (Figure 2e). Tumours with the K27M mutation in the gene encoding the histone variant H3.3, *H3F3A*, harboured significantly more iCNA (n = 5, median = 6, range 1–16) than either G34R/V mutant tumours (n = 4, median = 4.5, range 3–7) or wild-type (n = 14, median = 3, range 0–11) (p = 0.043, ANOVA) (Figure 2f). This was independent of location of tumour, with no differences in number of iCNA between supratentorial GBM (n = 51) and DIPG (n = 7, p = 0.684, t-test) (data not shown).

Intragenic amplifications and deletions

The 288 iCNA were further subjected to individual manual inspection of the data plots in order to identify the most robust copy number shifts associated with intragenic breaks. This resulted in a list of 39 unique events in 51 samples (Table 1). The recurrent changes included copy number loss, resulting primarily in either the absence of the 3' end of a gene or small deletions wholly within the coding sequence. These intragenic deletions included those targeting known tumour suppressors in glioblastoma

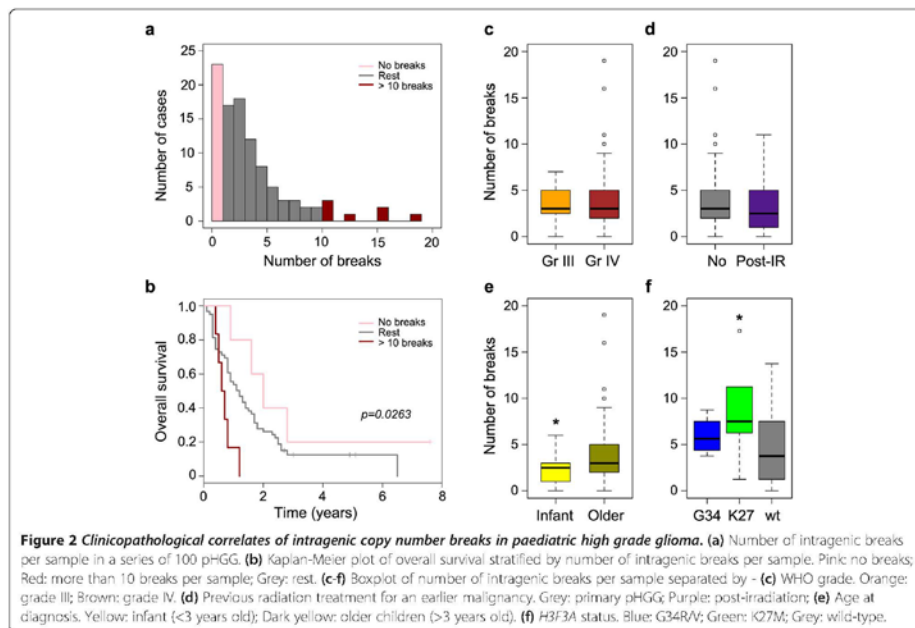


Table 1 Nominated intragenic copy number aberration candidates

Candidate.ID	Gene	Chromosome	Sample	Copy number change	Comments
1	FAF1	1	HGG091	Deletion	
	FAF1	1	HGG140	Deletion	
2	CD84	1	HGG088	Amplification	
3	LGALS8	1	HGG070	Deletion	
4	KIDINS220	2	HGG077	Amplification	
5	DHXS7	2	HGG063	Amplification	Fusion candidate
6	TMEM178	2	HGG063	Amplification	Fusion candidate
7	WDR49	3	HGG010	Amplification	
8	PDCD10	3	HGG157	Deletion	
9	CHIC2	4	HGG077	Amplification	
10	ITGA1	5	HGG029	Deletion	
11	EPHA7	6	HGG139	Deletion	
12	LANCL2	7	HGG060	Amplification	
13	ECOP	7	HGG060	Amplification	
14	KCND2	7	HGG152	Amplification	
	KCND2	7	HGG162	Amplification	
14	SND1	7	HGG090	Deletion	
15	CSMD3	8	HGG054	Deletion	
	CSMD3	8	HGG140	Deletion	
	CSMD3	8	HGG153	Deletion	
16	SLC24A2	9	HGG151	Deletion	
	SLC24A2	9	HGG011	Deletion	
17	MTAP	9	HGG022	Deletion	
	MTAP	9	HGG007	Deletion	
18	ANKRD26	10	HGG068	Deletion	
19	RET	10	HGG139	Amplification	Fusion candidate
20	CSGALNACT2	10	HGG139	Amplification	Fusion candidate
21	PTPRE	10	HGG086	Deletion	
	PTPRE	10	HGG145	Deletion	
22	RAB6IP1	11	HGG092	Amplification	
23	PSMA1	11	HGG092	Amplification	
24	TMTC1	12	HGG010	Amplification	
	TMTC1	12	HGG068	Amplification	
25	LRRK2	12	HGG068	Amplification	
26	MYO1A	12	HGG029	Amplification	
27	XRCC6BP1	12	HGG029	Amplification	
28	OSBPL8	12	HGG065	Amplification	
	OSBPL8	12	HGG162	Amplification	
29	RB1	13	HGG154	Deletion	
30	PCDH17	13	HGG059	Deletion	
31	CD276	15	HGG006	Deletion	
32	MEF2A	15	HGG011	Deletion	
33	DNAH2	17	HGG143	Amplification	

Table 1 Nominated intragenic copy number aberration candidates (Continued)

34	NF1	17	HGG154	Deletion
	NF1	17	HGG140	Deletion
35	BRIP1	17	HGG077	Deletion
36	KCNB1	20	HGG139	Amplification
37	SYN3	22	HGG017	Deletion
	SYN3	22	HGG146	Deletion
38	TIMP3	22	HGG146	Deletion
39	PHF21B	22	HGG072	Deletion

39 unique intragenic breakpoints found within 51 cases of paediatric high grade glioma. Direction of copy number shift (gain/loss) is reported, as well as candidate fusion events.

such as *NF1* (17q11.2, n = 2) (Figure 3a) and *RBI* (13q14.2, n = 1) (Additional file 3: Figure S3), as well as putative novel GBM-associated genes including *EAF1* (1p33, n = 2) and *MTAP* (9p21.3, n = 2) (Additional file 3: Figure S3). In addition, there were novel deletions in the protein phosphatase *PTPRE* (10q26.2, n = 2) (Figure 3b) and recurrent internal microdeletions in the gene *CSMD3* (CUB and Sushi multiple domains 3) (8q23.3, n = 3), all of which overlapped to result in the loss of exon 4 (Additional file 3: Figure S3).

Copy number gains within gene coding regions tended to be associated with regions flanking known oncogenic amplicons. These included amplification of the *MYCN* locus at chromosome 2p24.3, which in case HGG077 breaks within the coding region of the kinase D-interacting substrate *KIDINS220* (Figure 4a); amplification of *PDGFRA* at 4q12, harbouring an iCNA in *CHIC2* in the same case (though only covered by two probes); and recurrent breakpoints in the gene encoding the potassium voltage-gated channel *KCND2* at 7q31.31 in association with amplification of *MET*, though curiously this targeted either 5' or 3' ends in two different cases (Figure 4b). Similarly, common amplification events encompassing *EGFR* (7p12) and *CDK4* (12q14) had intragenic breakpoints at both ends in cases HGG060 (*LANCL2* and *ECOP*) and HGG029 (*MYO1A* and *XRCC6BPT*), respectively (Additional file 4: Figure S4).

Identification of novel fusion genes

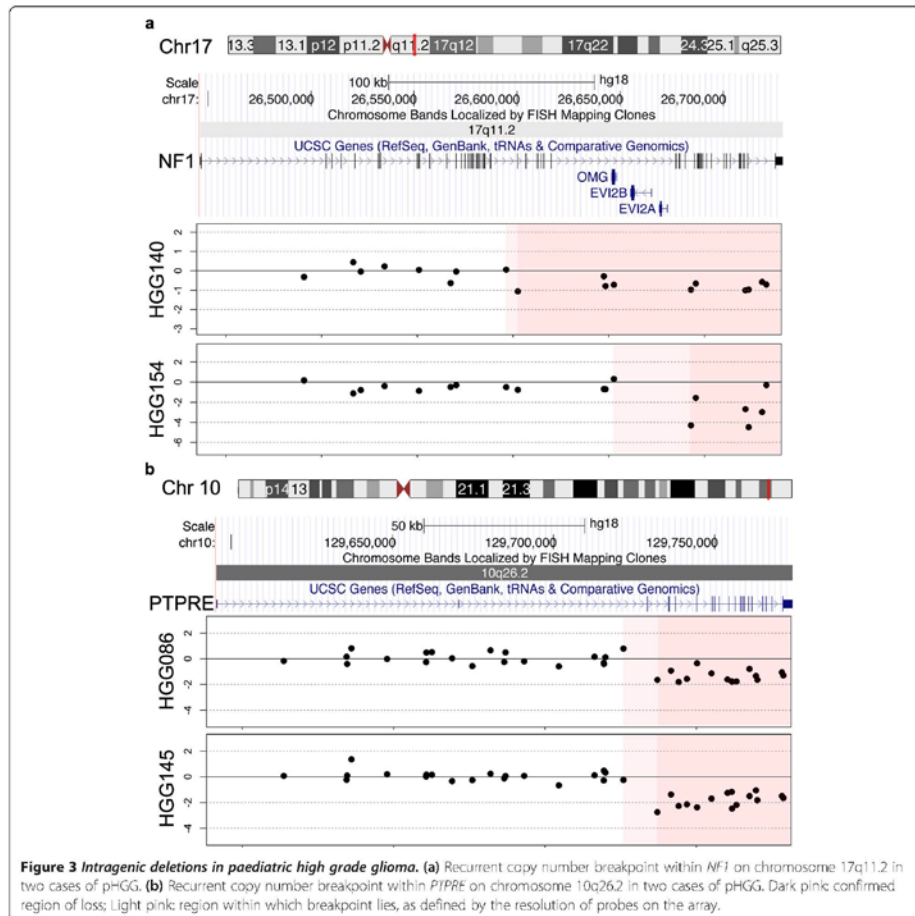
For the most part, iCNA events resulted in an imbalance of certain regions of coding genes in isolation, with the predicted consequence a disruption of full-length gene expression. For certain events however, a 5' end of one gene was found amplified at a similar copy number to a 3' end of a second gene within the same case. We reasoned that such instances may represent candidate fusion genes, and we identified two such examples in our cohort.

The first was at chromosome 10q11.21 and reflected a single amplicon, breaking within the genes *RET* and *CSGALNACT2* such that we propose a hypothetical fusion gene encompassing the 5' regulatory regions of

CSGALNACT2 and the 3' kinase domain of *RET*. In order to determine the precise breakpoints to allow validation of this novel fusion, we designed custom oligonucleotide arrays spanning the amplicon in order to carry out high-resolution array CGH on the reference case HGG139, a relapse sample of glioblastoma in which this genomic event was not present in the primary tumour. Although the breakpoint for *CSGALNACT2* was identified within intron 2, leaving the catalytic domains intact, the breakpoint within *RET* could not be accurately determined to closer than 10 kb between introns 1 and 2 (Additional file 5: Figure S5). As material was limited for this case, we were unable to confirm the precise nature of the putative *CSGALNACT2:RET* fusion by PCR-based techniques.

The second fusion candidate was located at an amplified region of chromosome 2p22.1 in case HGG063, an anaplastic astrocytoma. At Affymetrix 500 K SNP resolution, this appeared to be a single amplicon with breaks within the coding regions of the RNA helicase *DHX57* and the transmembrane protein *TMEM178* (Figure 5a). Applying the same approach as above, using custom-designed oligonucleotide arrays for high-resolution array CGH revealed two amplicons within this structure, with further intragenic breakpoints within the mitogen-activated protein kinase *MAP4K3* (Figure 5b). Designing PCR primers to amplify across the highly specific breakpoints confirmed the presence of the fusion, which was further validated by direct sequencing (Figure 6).

The resultant fusion gene, *DHX57:TMEM178:MAP4K3*, is a complex three gene fusion formed from a series of intragenic breaks, amplifications and inversions to produce a sequence comprising exons 1–12 of *DHX57*, exons 2–4 of *TMEM178* and exons 13–34 of *MAP4K3*, associated with regions of microhomology (Figure 6). This would produce a protein with the zinger finger and DEAD-like helicase domains of *DHX57*, the claudin family transmembrane domains of *TMEM178* and the citron domain of *MAP4K3*. Selective knockdown of *MAP4K3* by siRNA leads to a significant reduction in cell viability in five paediatric glioma cell lines as assayed by CellTiter Glo, an



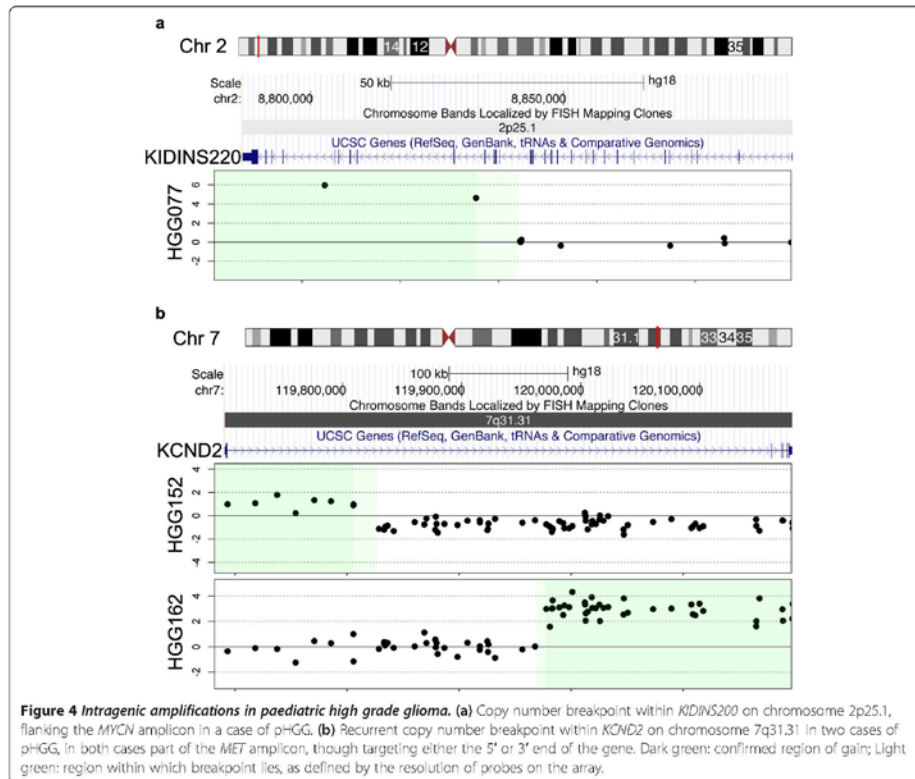
effect not seen in 18/20 breast cancer cells ($p = 0.0017$, pHG vs breast cancer, t -test) (Additional file 6: Figure S6).

Discussion

Comprehensive copy number profiling of adult and paediatric high grade gliomas was among the first data to demonstrate the biological differences between these similar-looking histological malignancies [18]. In this context, the focus has been on large-scale genomic copy number changes. A more refined analysis of copy number and exon-level expression data has identified new

insights into genomic architecture and novel fusion proteins in adult glioblastoma [12,13]. Here we leverage a large dataset we have previously generated [4] in the paediatric disease to carry out a scan of intragenic breakpoints, leading to the identification of novel gene disruptions and candidate gene fusions.

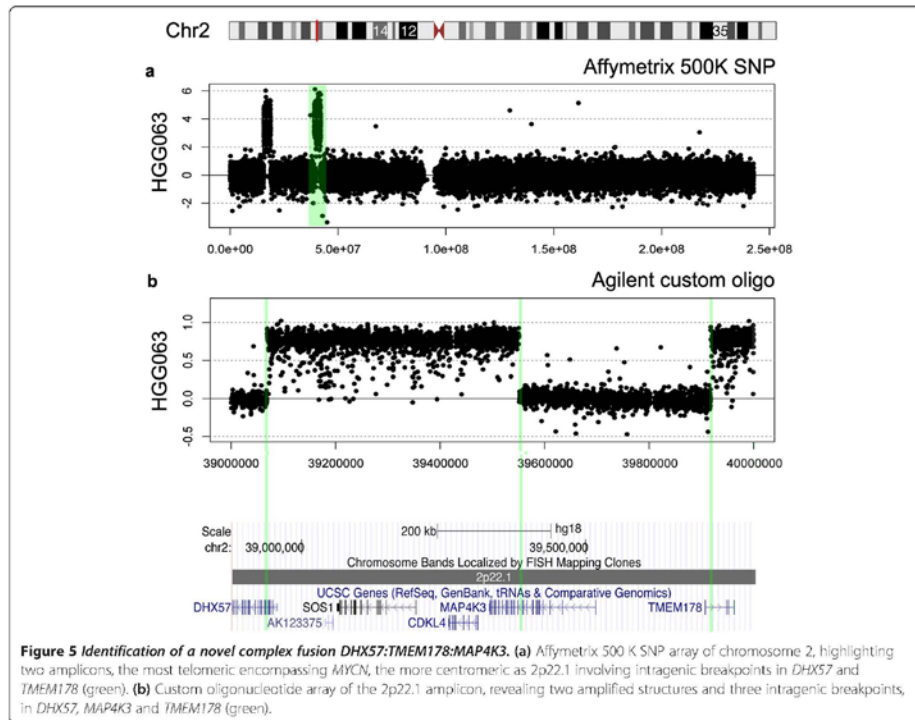
The presence of intragenic copy number aberrations was confirmed in the vast majority of pHG cases, and was itself prognostic, with an absence of iCNAs conferring a longer overall survival in paediatric patients. This was associated with the infant age group, known to have



a better clinical outcome than older children [19], and further highlights the biological distinctiveness of this age group. By contrast, the presence of large numbers of intragenic breaks conferred a shorter survival time, but was not a result of the grade of the tumour, nor associated with a second malignancy due to radiation treatment for an earlier cancer. We had previously reported an association of post-irradiated HGG with *PDGFRA* amplification and chromosome 1q gain [4], so it appears these are relatively selective radiation-induced changes, rather than reflecting a generalised genomic instability in secondary tumours from these patients. Importantly, we identified an increased number of iCNA in tumours harbouring an *H3F3A* K27M mutation, regardless of anatomical location. This is a group of thalamic and pontine HGG associated with a particularly dismal prognosis [18], for whom understanding the mechanisms of

genomic instability and the identification of novel gene disruptions is of considerable interest.

The majority of intragenic breakpoints we identified were associated with gene disruption. This includes deletions of known tumour suppressors such as *RBI* and *NFI*, but also more novel glioblastoma associated genes. *FAF1* and *MTAP* were both recurrently targeted by intragenic deletion events in pHGG. These genes are localised close to known cyclin-dependent kinase inhibitors and tumour suppressors *CDKN1C* and *CDKN2A/B*, respectively, but both *FAF1* and *MTAP* have recently been proposed to harbour tumour suppressor activity in their own right. *FAF1* is associated with a FAS-mediated apoptosis response and restoration of the FAF1 protein in adult glioma cell lines significantly increases cell death [20], whilst in *MTAP*-deficient cells, methylthioadenosine, generated during polyamine biosynthesis, is not

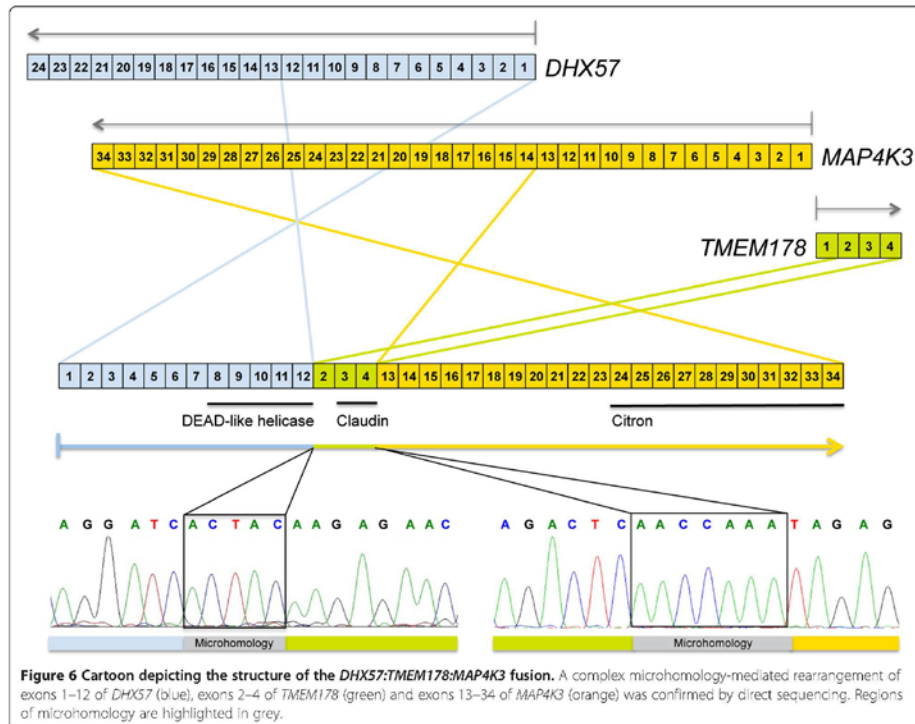


cleaved and the salvage pathway for adenine and methionine is absent [21]. It seems that such mechanisms are also likely in a subset of paediatric tumours.

Of note we identified novel deletions in the protein phosphatase epsilon, *PTPRE*. This has not been reported previously, although there are several reports of the tumour suppressive capacity of the related *PTPRD* [22,23]. This gene also appears targeted by intragenic deletions, and human astrocytes lacking *PTPRD* exhibited increased growth, as it is thought the protein usually functions to dephosphorylate the oncoprotein *STAT3* [23]. The wholly intragenic microdeletions observed in *CSMD3* in four cases may represent another novel mechanism of gene disruption. *CSMD3* encodes a gene with multiple CUB and Sushi domains whose function is poorly understood. Recently, *CSMD3* was identified as the second most frequently mutated gene (next to *TP53*) in lung cancer, where it was demonstrated that loss of *CSMD3* results in increased proliferation of airway epithelial cells [24].

Gene disruption may also play a significant functional role when known gain-of-function oncogenes are amplified. We report numerous intragenic breakpoints which may have been overlooked in the context of identifying the 'driver' event within a common amplicon, but which may themselves be tumorigenic. These include disruptions of *KIDINS220*, a functional mediator of multiple receptor signalling pathways and essential for cortical development [25,26]; *CHIC2*, frequently deleted/rearranged in myeloid malignancies [27]; and *KCND2*, encoding a potassium voltage-gated channel, which is expressed in both neuronal and glial cells and has been shown to regulate ERK signalling in ganglioglioma [28]. All of these gene disruptions represent novel avenues for understanding the underlying biology of pHGG.

Of most interest was the use of the iCNA algorithm to identify potential novel fusion genes, as was demonstrated in adult glioblastoma with the identification of the *KDR:PDGFRA* fusion [10], which we also found in a case of pHGG [11]. Our analysis nominated two potential



candidates – the first we were unable to conclusively validate, *CSGALNACT2:RET*. Such a putative fusion would retain the kinase domain of the *RET* oncoprotein, but would lose the autoregulatory portion of the protein, instead fusing it to the N terminal of chondroitin sulfate N-acetyl-galactosaminyltransferase 2. Although a precise cancer-related function has not been ascribed to the latter enzyme, it is thought to play an important role in morphogenesis in zebrafish models [29,30]. Whilst not validated, oncogenic *RET* rearrangements and fusions are common in thyroid and lung cancer [31,32], and the presence of infrequent activating fusions in HGG do not seem unlikely.

We were able to validate a novel complex fusion involving three genes with intragenic breakpoints and amplification/rearrangement on chromosome 2p22.1. The resulting fusion gene, *DHX57:TMEM178:MAP4K3* encompasses key regulatory domains from all three proteins, though a specific function is hard to predict. The helicase properties of the *DHX57* component may be a candidate for oncogenicity, with numerous other DEAD-box helicases appearing to play a role in regulation of DNA repair,

apoptosis and drug sensitivity [33]. *MAP4K3* has been associated with several malignancies in both an oncogenic and tumour suppressor capacity [34,35]. In particular, one function that has been ascribed includes activation of mTOR signalling via the TORC1 complex [36], a pathway commonly activated by diverse mechanisms in pHGG [18].

In the context of pHGG, although the kinase domain is not retained in the fusion, *MAP4K3* plays some functional role as selective knockdown by siRNA leads to a significant and selective reduction in cell viability in paediatric glioma cell lines. Thus we hypothesise that the *DHX57:TMEM178:MAP4K3* is activating as disruption of the protein would otherwise seem incompatible with tumour cell growth and proliferation.

Conclusion

In summary these data represent a key addition to our understanding of the genomic alterations driving pHGG and provide novel avenues for developing sorely-needed novel therapeutic strategies for children with these otherwise incurable tumours.

Additional files

Additional file 1: Table S1. Initial output from ICNA algorithm detailing 1099 intragenic copy number breakpoints in 100 cases of paediatric high grade glioma.

Additional file 2: Table S2. Final output from ICNA algorithm detailing 288 filtered intragenic copy number breakpoints in 100 cases of paediatric high grade glioma.

Additional file 3: Figure S3. *Intragenic deletions in paediatric high grade glioma.* (a) Recurrent copy number breakpoint within *FAF1* on chromosome 1p33 in two cases of pHGG. (b) Recurrent copy number breakpoint within *MTAP* on chromosome 9p21.3 in two cases of pHGG. (c) Copy number breakpoint within *RBT* on chromosome 13q14.2 in a case of pHGG. (b) Recurrent copy number breakpoint within *CMD3* on chromosome 8q23.3 in three cases of pHGG. Dark pink: confirmed region of loss; Light pink: region within which breakpoint lies, as defined by the resolution of probes on the array.

Additional file 4: Figure S4. *Intragenic amplifications in paediatric high grade glioma.* (a) Copy number breakpoints within *LANCL2* and *ECOP* on chromosome 7p11.2, flanking the *EGFR* amplicon in a case of pHGG. (b) Recurrent copy number breakpoints within *MYO1A1* and *XRCC6BP1* on chromosome 12q13.3 and 12q14.1, flanking the *CDK4* amplicon in a case of pHGG. Dark green: confirmed region of gain; Light green: region within which breakpoint lies, as defined by the resolution of probes on the array.

Additional file 5: Figure S5. *Identification of a novel candidate fusion CSGALNACT2-RET.* (a) Affymetrix 500 K SNP array of chromosome 10, highlighting an amplicon at 10q11.21 (green). (b) Custom oligonucleotide array of the 10q11.21 amplicon, revealing a clear breakpoint within *CSGALNACT2* (green), but a less clear boundary within *RET* (grey).

Additional file 6: Figure S6. *siRNA knockdown of MAP4K3 in paediatric glioma and breast carcinoma cells.* Paediatric glioma cells (green) were highly sensitive to knockdown of MAP4K3, with 5/5 cells showing significant effects on cell viability. By contrast, only 2/20 breast cancer cells (blue) showed a similar dependency on MAP4K3 expression for cell viability. The screen was carried out in three independent experiments and was highly reproducible for all cell lines, with R^2 values ranging from 0.68-0.94 (breast) and 0.78-0.92 (glioma). The different sensitivity of glioma cells to MAP4K3 knockdown as compared to breast carcinoma cells was statistically significant ($p = 0.0017$, pHGG vs breast cancer, t -test).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DC designed and performed experiments and analysed and interpreted data. LB designed and interpreted experiments. AM analysed and interpreted data. RG provided clinical samples. RMR, CL and CJ designed the study. DC and CJ wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We are grateful for support from the Rosetrees Trust, the Brain Tumour Charity and Fundação para a Ciência e Tecnologia, Portugal (PhD Studentship SFRH/BD/33473/2008). DC, AM, LB and CJ acknowledge NHS funding to the Biomedical Research Centre.

Author details

¹Divisions of Molecular Pathology and Cancer Therapeutics, Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK. ²University of Coimbra, Palácio dos Góios, R. da Ilha, Coimbra 3000-214, Portugal. ³Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, and ICVS/3B's-PT Government Associate Laboratory, Braga/Guimarães, Portugal. ⁴Childhood Brain Tumour Research Centre, University of Nottingham, Kings Meadow Campus, Lenton Lane, Nottingham NG7 2NR, UK. ⁵Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, SP, Brazil.

Received: 16 December 2013 Accepted: 11 February 2014
Published: 18 February 2014

References

1. Bax DA, MacIay A, Little SE, Carvalho D, Vana-Pereira M, Tambar N, Grigorádis A, Ashworth A, Reis RM, Ellison DW, Al-Saraj S, Hargrave D, Jones C: **A distinct spectrum of copy number aberrations in paediatric high grade gliomas.** *Clin Cancer Res* 2010, **16**(13):3368-3377.
2. McLendon R, Friedman A, Bigner D, Van Meir EG, Brat DJ, Mastrogiannis M, Olson JJ, Mikkelsen T, Lehman N, Aldape K, Alfred Yung WK, Bogler O, Vandenberg S, Berger M, Prados M, Muzny D, Morgan M, Scherer S, Sabo A, Nazareth L, Lewis L, Hall O, Zhu Y, Ren Y, Alvi Q, Yao J, Hawes A, Jhangiani S, Fowler G, San Lucas A, et al: **Comprehensive genomic characterization defines human glioblastoma genes and core pathways.** *Nature* 2008, **455**:1061-1068.
3. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivari A, McLendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz LA Jr, Hartigan J, Smith DR, Strausberg RL, Marle SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, et al: **An integrated genomic analysis of human glioblastoma multiforme.** *Science* 2008, **321**(5897):1807-1812.
4. Paugh BS, Qu C, Jones C, Liu Z, Adamowicz-Brice M, Zhang J, Bax DA, Coyle B, Barrow J, Hargrave D, Lowe J, Gajjar A, Zhao W, Broniscer A, Ellison DW, Grundy RG, Baker SJ: **Integrated molecular genetic profiling of pediatric high grade gliomas reveals key differences with the adult disease.** *J Clin Oncol* 2010, **28**(18):3061-3068.
5. Schwartzentruber J, Korshunov A, Liu XY, Jones DT, Pfaff E, Jacob K, Sturm D, Fontebasso AM, Quang DA, Tonjes M, Hovestadt V, Albrecht S, Kool M, Nantel A, Konermann C, Lindroth A, Jager N, Rausch T, Ryzhova M, Korbel JO, Hielscher T, Hauser P, Garami M, Kleimer A, Bogner L, Ebinger M, Schuhmann MU, Scheurle W, Peikun A, Fruhwald MC, et al: **Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma.** *Nature* 2012, **482**(7384):226-231.
6. Sturm D, Witt H, Hovestadt V, Khuong-Quang DA, Jones DT, Konermann C, Pfaff E, Tonjes M, Sill M, Bender S, Kool M, Zaparka M, Becker N, Zucknick M, Hielscher T, Liu XY, Fontebasso AM, Ryzhova M, Albrecht S, Jacob K, Wolter M, Ebinger M, Schuhmann MU, van Meter T, Fruhwald MC, Hauch H, Peikun A, Radlwimmer B, Niehues T, von Kromowski G, et al: **Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma.** *Cancer Cell* 2012, **22**(4):425-437.
7. Beroukhim R, Getz G, Nghiemphu L, Baretina J, Hsueh T, Linhart D, Vivanco L, Lee JC, Huang JH, Alexander S, Du J, Kau T, Thomas RK, Shah K, Soto H, Perner S, Prensner J, DeBiasi RM, Demichellis F, Hatton C, Rubin MA, Garraway LA, Nelson SF, Liaw L, Mischel PS, Cloughesy TF, Meyerson M, Golub TA, Lander ES, Mellinghoff IK, Sellers WR: **Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma.** *Proc Natl Acad Sci USA* 2007, **104**(50):20007-20012.
8. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM: **Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer.** *Science* 2005, **310**(5748):644-648.
9. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H: **Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer.** *Nature* 2007, **448**(7153):561-566.
10. Ozawa T, Brennan CW, Wang L, Squatrito M, Sasayama T, Nakada M, Huse JT, Pedraza A, Utsuki S, Yasui Y, Tandon A, Forchlenko EI, Oka H, Levine RL, Fujii K, Ladanyi M, Holland EC: **PDGFRA gene rearrangements are frequent genetic events in PDGFRA-amplified glioblastomas.** *Genes Dev* 2010, **24**(19):2205-2218.
11. Paugh BS, Zhu X, Qu C, Endersby R, Diaz AK, Zhang J, Bax DA, Carvalho D, Reis RM, Onar-Thomas A, Broniscer A, Wetmore C, Zhang J, Jones C, Ellison DW, Baker SJ: **Novel Oncogenic PDGFRA Mutations in Pediatric High-Grade Gliomas.** *Cancer Res* 2013, **73**(20):6219-6229.
12. Singh D, Chan JM, Zoppoli P, Niola F, Sullivan R, Castano A, Liu EM, Reichel J, Porrati P, Pellegatta S, Qiu K, Gao Z, Ceccarelli M, Riccardi R, Brat DJ, Guha A, Aldape K, Goffinos JG, Zagzag D, Mikkelsen T, Finocchiaro G, Lasorella A,

- Rabadan R, Iavarone A: **Transforming fusions of FGFR and TACC genes in human glioblastoma.** *Science* 2012, **337**(6099):1231–1235.
13. Frattini V, Trifonov V, Chan JM, Castano A, Lia M, Abate F, Keir ST, Ji AX, Zoppoli P, Niola F, Danussi C, Dolgalev I, Poratti P, Pellegatta S, Heguy A, Gupta G, Pisapia DJ, Canoll P, Bruce JN, McLendon RE, Yan H, Aldape K, Finocchiaro G, Mikkelsen T, Prive GG, Bigner DD, Lasorella A, Rabadan R, Iavarone A: **The integrated landscape of driver genomic alterations in glioblastoma.** *Nat Genet* 2013, **45**(10):1141–1149.
 14. Wiedenmeyer R, Brennan C, Heffernan TP, Xiao Y, Mahoney J, Protopopov A, Zheng H, Bignelli G, Fumari F, Cavenee WK, Hahn WC, Ichimura K, Collins VP, Chu GC, Stratton MR, Ligon KL, Futreal PA, Chin L: **Feedback circuit among INK4 tumor suppressors constrains human glioblastoma development.** *Cancer Cell* 2008, **13**(4):355–364.
 15. Olishen AB, Venkatraman ES, Lucito R, Wigler M: **Circular binary segmentation for the analysis of array-based DNA copy number data.** *Biostatistics* 2004, **5**(4):557–572.
 16. Bax DA, Little SE, Gaspar N, Penryman L, Marshall L, Viana-Pereira M, Jones T, Williams RD, Vassal GM, Workman P, Sheer D, Reis RM, Pearson ADJ, Hargrave D, Jones C: **Molecular and phenotypic characterisation of paediatric glioma cell lines as models for preclinical drug development.** *PLoS ONE* 2009, **4**(4):e5209.
 17. Iafate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C: **Detection of large-scale variation in the human genome.** *Nat Genet* 2004, **36**(9):949–951.
 18. Jones C, Penryman L, Hargrave D: **Paediatric and adult malignant glioma: close relatives or distant cousins?** *Nat Rev Clin Oncol* 2012, **9**(7):400–413.
 19. Wu W, Lamborn KR, Buckner JC, Novotny PJ, Chang SM, O'Fallon JR, Jaeckle KA, Prados MD: **Joint NCTG and NABTC prognostic factors analysis for high-grade recurrent glioma.** *Neuro Oncol* 2010, **12**(2):164–172.
 20. Zheng S, Fu J, Vegesna R, Mao Y, Heathcock LE, Torres-Garcia W, Ethilbarasan R, Wang S, McKenna A, Chin L, Brennan CW, Yung WK, Weinstein JN, Aldape KD, Sulman EP, Chen K, Koul D, Verhaak RG: **A survey of intragenic breakpoints in glioblastoma identifies a distinct subset associated with poor survival.** *Genes Dev* 2013, **27**(13):1462–1472.
 21. Bertino JR, Waud WR, Parker WB, Lubin M: **Targeting tumors that lack methylthioadenosine phosphorylase (MTAP) activity: current strategies.** *Cancer Biol Ther* 2011, **11**(7):627–632.
 22. Solomon DA, Kim JS, Cronin JC, Sibenaller Z, Ryken T, Rosenberg SA, Rensom H, Jean W, Bigner D, Yan H, Samuels Y, Waldman T: **Mutational inactivation of PTPRD in glioblastoma multiforme and malignant melanoma.** *Cancer Res* 2008, **68**(24):10300–10306.
 23. Veeriah S, Brennan C, Meng S, Singh B, Fagin JA, Solit DB, Paty PB, Rohle D, Vivanco L, Chmielecki J, Pao W, Ladanyi M, Gerald WL, Liu L, Cloughesy TC, Mischel PS, Sander C, Taylor B, Schultz N, Major J, Heguy A, Fang F, Mellingerhoff IK, Chan TA: **The tyrosine phosphatase PTPRD is a tumor suppressor that is frequently inactivated and mutated in glioblastoma and other human cancers.** *Proc Natl Acad Sci USA* 2009, **106**(23):9435–9440.
 24. Liu P, Morrison C, Wang L, Xiong D, Vedell P, Cui P, Hua X, Ding F, Lu Y, James M, Ebben JD, Xu H, Adjei AA, Head K, Andrae JW, Tschannen MR, Jacob H, Pan J, Zhang Q, Van den Bergh F, Xiao H, Lo KC, Patel J, Richmond T, Watt MA, Albert T, Selzer R, Anderson M, Wang J, Wang Y, Stames S, Yang P, You M: **Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing.** *Carcinogenesis* 2012, **33**(7):1270–1276.
 25. Cesca F, Yabe A, Spencer-Dene B, Arifgoni A, Al-Qatari M, Henderson D, Phillips H, Koltzenburg M, Benfenati F, Schiavo G: **Kidins220/ARMS is an essential modulator of cardiovascular and nervous system development.** *Cell Death Dis* 2011, **2**:e226.
 26. Neubrand VE, Cesca F, Benfenati F, Schiavo G: **Kidins220/ARMS as a functional mediator of multiple receptor signalling pathways.** *J Cell Sci* 2012, **125**(Pt 8):1845–1854.
 27. Cools J, Bilhou-Nabera C, Wlodarska I, Cabrol C, Talmant P, Bernard P, Hagemeijer A, Marynen P: **Fusion of a novel gene, BTL, to ETV6 in acute myeloid leukemias with a t(4;12)(q11-q12p13).** *Blood* 1999, **94**(5):1820–1824.
 28. Adams JP, Anderson AE, Varga AW, Dineley KT, Cook RG, Pfaffinger PJ, Sweatt JD: **The A-type potassium channel Kv4.2 is a substrate for the mitogen-activated protein kinase ERK.** *J Neurochem* 2000, **75**(6):2277–2287.
 29. Filippek-Gomiok B, Holmborn K, Haltina T, Habicher J, Oliveira MB, Hellgren C, Eriksson I, Kjellen L, Kreuger J, Ledlin J: **Expression of chondroitin/dermatan sulfate glycosyltransferases during early zebrafish development.** *Dev Dyn* 2013, **242**(8):964–975.
 30. Holmborn K, Habicher J, Kasza Z, Eriksson AS, Filippek-Gomiok B, Gopal S, Couchman JR, Ahlberg PE, Wieweger M, Spillmann D, Kreuger J, Ledlin J: **On the roles and regulation of chondroitin sulfate and heparan sulfate in zebrafish pharyngeal cartilage morphogenesis.** *J Biol Chem* 2012, **287**(40):33905–33916.
 31. Fusco A, Santoro M: **20 years of RET/PTC in thyroid cancer: clinico-pathological correlations.** *Arq Bras Endocrinol Metabol* 2007, **51**(5):731–735.
 32. Gainor JF, Shaw AT: **Novel targets in non-small cell lung cancer: ROS1 and RET fusions.** *Oncologist* 2013, **18**(7):865–875.
 33. Fuller-Pace FV: **DEAD box RNA helicase functions in cancer.** *RNA Biol* 2013, **10**(1):121–132.
 34. Lam D, Dickens D, Reid EB, Loh SH, Moiso N, Martins LM: **MAP4K3 modulates cell death via the post-transcriptional regulation of BHLH-only proteins.** *Proc Natl Acad Sci USA* 2009, **106**(29):11978–11983.
 35. Wright JH, Wang X, Manning G, LaMere BJ, Le P, Zhu S, Khatry D, Flanagan PM, Buckley SD, Whyte DB, Howlett AR, Bischoff JR, Lipson KE, Jallal B: **The STE20 kinase HGK is broadly expressed in human tumor cells and can modulate cellular transformation, invasion, and adhesion.** *Mol Cell Biol* 2003, **23**(6):2068–2082.
 36. Findlay GM, Yan L, Procter J, Mieulet V, Lamb RF: **A MAP4 kinase related to Ste20 is a nutrient-sensitive regulator of mTOR signalling.** *Biochem J* 2007, **403**(1):13–20.

doi:10.1186/2051-5960-2-23
Cite this article as: Carvalho et al.: **The prognostic role of intragenic copy number breakpoints and identification of novel fusion genes in paediatric high grade glioma.** *Acta Neuropathologica Communications* 2014 **2**:23.

Submit your next manuscript to BioMed Central and take full advantage of:

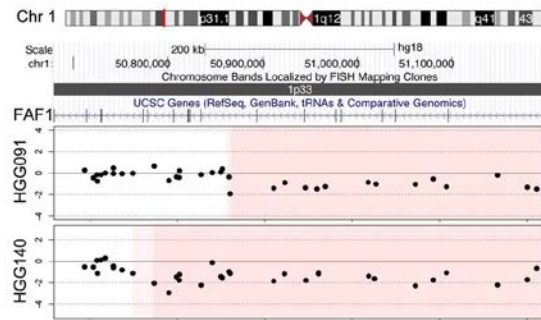
- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

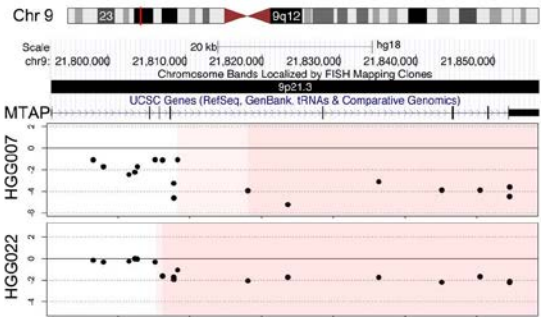


Supplementary Figure S3

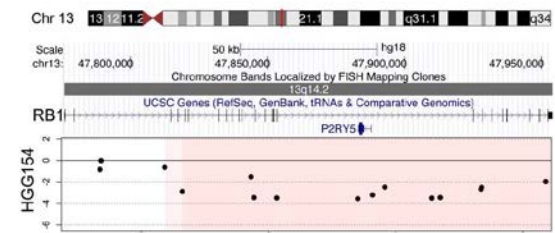
a



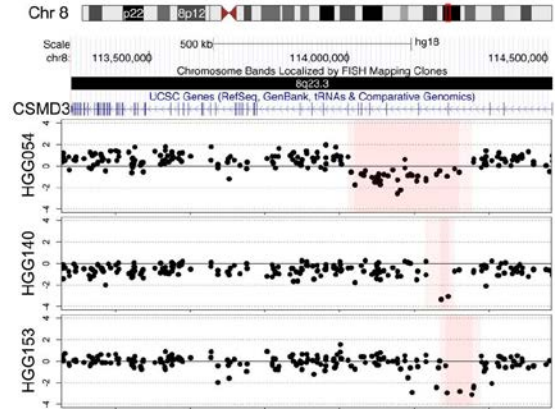
b



c

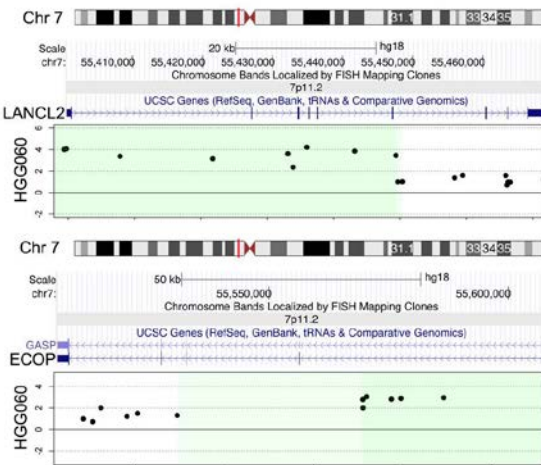


d

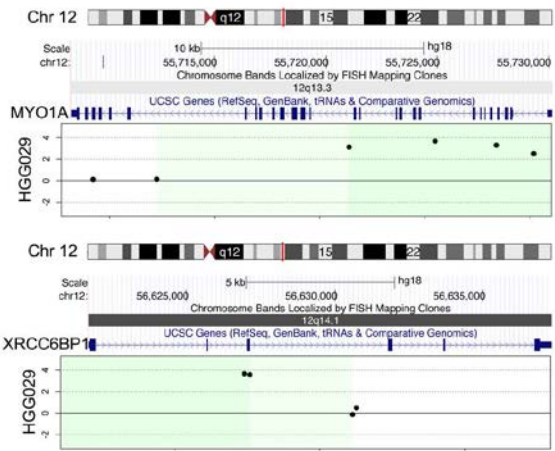


Supplementary Figure S4

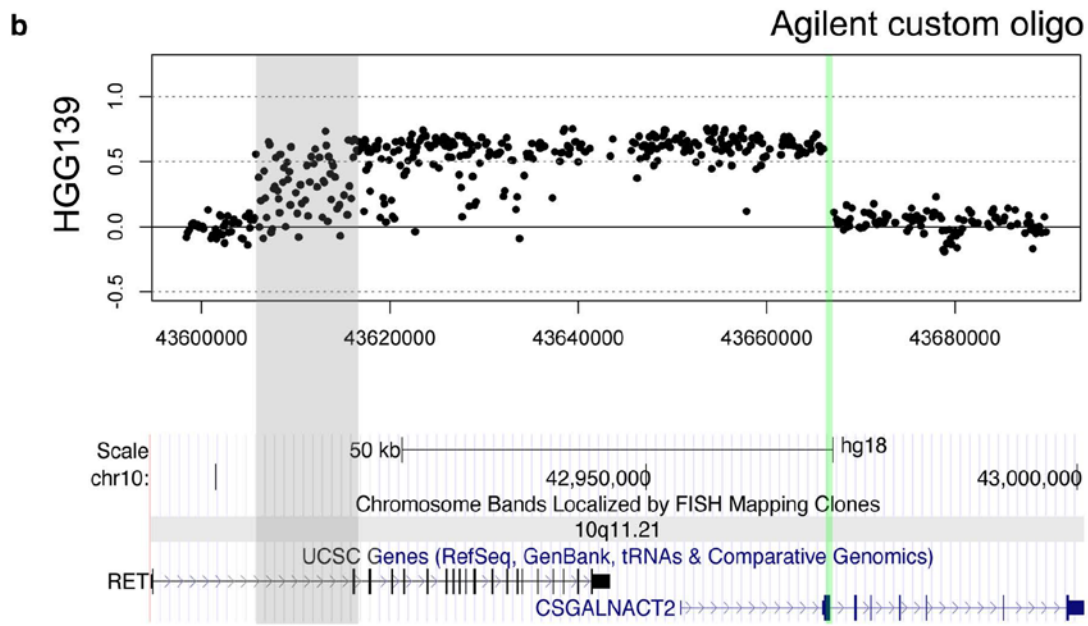
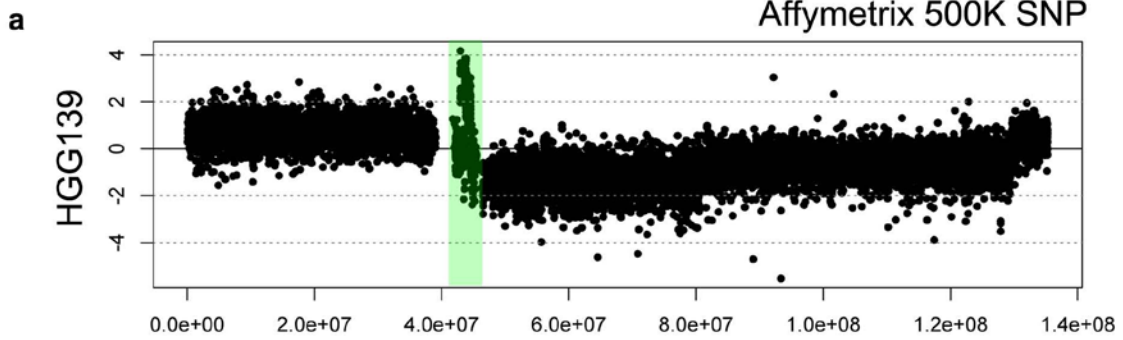
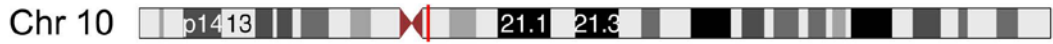
a



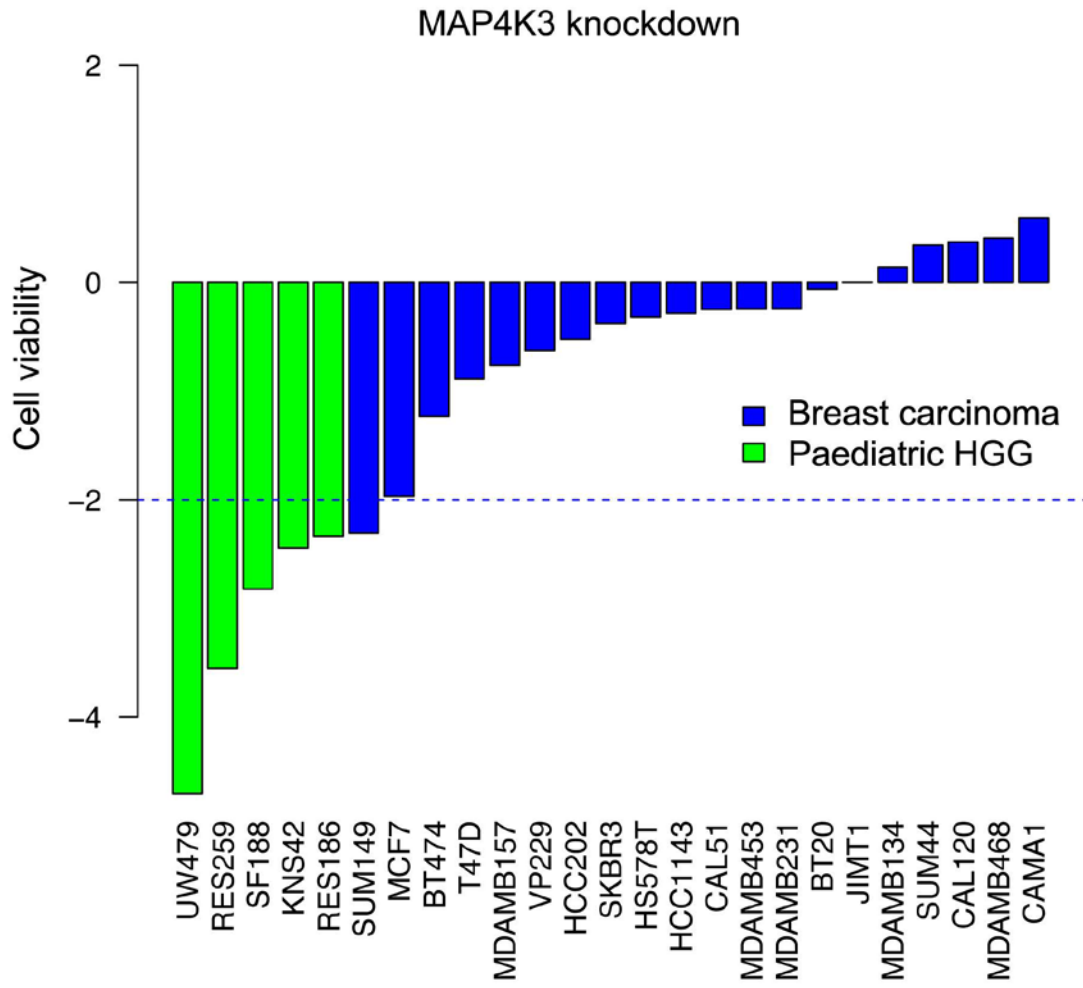
b



Supplementary Figure S5



Supplementary Figure S6



Appendix V

Table S5.1. Structural variants (SV) predicted by the Breakdancer. All SVs listed involve genes at either side of the rearrangement. Breakdancer score is >99 for all SVs. Genes were annotated with Ensembl references (columns: “Gene 1 name” and “Gene 2 name”).

Chr 1	Pos 1	Ori 1	Chr 2	Pos 2	Ori 2	Type	Insert Size	Reads	Cell Line	Gene 1 name	Gene 2 name
chr1	9121800	36+22-	chr14	93712302	38+23-	CTX	-404	57	KNS42	ENSG00000142583 SLC2A5	ENSG00000011114 BTBD7
chr1	28515580	22+0-	chr19	24027536	22+0-	CTX	-404	22	KNS42	ENSG00000169403 PTAFR	
chr1	38433028	15+15-	chr6	28863601	14+15-	CTX	-404	27	KNS42	ENSG00000183431 SF3A3	
chr1	58802590	1+16-	chr10	74176383	0+15-	CTX	-404	15	KNS42	ENSG00000173406 DAB1	ENSG00000107745 MICU1
chr1	59096783	10+1-	chr2	32048657	1+12-	CTX	-404	10	KNS42	ENSG00000185839	ENSG00000242272 AK2P2
chr1	109495128	30+28-	chr3	110413045	28+30-	CTX	-404	58	KNS42	ENSG00000121940 CLCC1	
chr1	119401099	13+2-	chr15	51189367	0+12-	CTX	-404	12	KNS42		ENSG00000247586
chr1	152276735	0+26-	chr1	152281500	26+1-	ITX	4064	26	KNS42	ENSG00000143631 FLG	ENSG00000237975
chr1	152387895	37+0-	chr11	132713260	2+43-	CTX	-404	37	KNS42	ENSG00000227415	ENSG00000183715 OPCML
chr1	156051752	12+0-	chr3	189542047	12+0-	CTX	-404	12	KNS42	ENSG00000254726 MEX3A	ENSG00000073282 TP63
chr1	168025506	22+0-	chr19	24033172	2+24-	CTX	-404	22	KNS42	ENSG00000143164 DCAF6	
chr1	196735841	4+18-	chr1	196745076	19+1-	ITX	8434	18	KNS42		ENSG00000116785 CFHR3
chr1	200067607	15+17-	chr2	23697849	17+18-	CTX	-404	31	KNS42	ENSG00000116833 NR5A2	ENSG00000119771 KLHL29
chr1	201070274	1+15-	chr1	201210022	15+2-	ITX	138949	14	KNS42	ENSG00000081248 CACNA1S	
chr2	43437013	0+20-	chr6	45891802	2+19-	CTX	-404	18	KNS42	ENSG00000115970 THADA	ENSG00000112782 CLIC5
chr2	131096289	0+12-	chr2	131110431	15+3-	ITX	13271	12	KNS42	ENSG00000136710 CCDC115	
chr2	131778059	0+12-	chr9	1407175	12+0-	CTX	-404	12	KNS42	ENSG00000136002 ARHGEF4	
chr2	132976153	1+12-	chr14	19040918	10+1-	CTX	-404	10	KNS42	ENSG00000163046 ANKRD30BL	
chr2	132987650	27+0-	chr15	20000010	31+14-	CTX	-404	26	KNS42	ENSG00000163046 ANKRD30BL	
chr2	133001848	3+34-	chr14	19000141	0+33-	CTX	-404	32	KNS42	ENSG00000163046 ANKRD30BL	

chr2	133005674	5+3-	chr21	10707432	162+185-	ITX	-312	68	KNS42	ENSG00000163046 ANKRD30BL	
chr2	133006389	19+8-	chr16	33970966	42+14-	CTX	-404	14	KNS42	ENSG00000163046 ANKRD30BL	
chr2	160149043	42+33-	chr3	58772063	42+31-	CTX	-404	72	KNS42		ENSG00000163689 C3orf67
chr2	171814878	21+1-	chr13	49848076	1+18-	CTX	-404	18	KNS42	ENSG00000115806 GORASP2	ENSG00000102543 CDADC1
chr2	172409811	2+18-	chr13	32800774	18+1-	CTX	-404	18	KNS42	ENSG00000071967 CYBRD1	ENSG00000073910 FRY
chr3	80325550	14+17-	chr6	121481236	14+16-	CTX	-404	28	KNS42		ENSG00000146350 C6orf170
chr3	84440279	2+21-	chr4	190840980	1+22-	CTX	-404	21	KNS42		ENSG00000245685
chr3	111274477	0+29-	chr8	128533826	0+29-	CTX	-404	29	KNS42	ENSG00000153283 CD96	
chr3	122653134	0+11-	chr12	49892490	11+1-	CTX	-404	10	KNS42	ENSG00000082684 SEMA5B	ENSG00000123352 SPATS2
chr3	164722685	1+36-	chr5	57093054	34+1-	CTX	-404	33	KNS42	ENSG00000090402 SI	
chr3	174480490	0+13-	chr20	54946398	0+12-	CTX	-404	12	KNS42	ENSG00000177694 NAALADL2	ENSG00000087586 AURKA
chr3	195729630	11+1-	chr5	1623025	2+10-	CTX	-404	11	KNS42		ENSG00000188002
chr3	196625623	6+2-	chr10	42597220	165+150-	ITX	-335	35	KNS42	ENSG00000119231 SENP5	
chr4	153556195	0+12-	chr21	23045425	0+13-	CTX	-404	12	KNS42	ENSG00000170006 TMEM154	ENSG00000237527
chr4	160666229	5+12-	chr21	17459342	6+12-	CTX	-404	17	KNS42	ENSG00000250488	ENSG00000215386 LINC00478
chr4	175454371	22+41-	chr11	127378377	22+13-	CTX	-404	12	KNS42	ENSG00000250596	
chr4	175719343	0+21-	chr4	176196699	20+0-	ITX	476537	19	KNS42	ENSG00000145451 GLRA3	
chr5	159349918	0+31-	chrX	66982460	31+0-	CTX	-404	31	KNS42	ENSG00000170214 ADRA1B	
chr6	14261299	16+0-	chr8	8718282	0+16-	CTX	-404	16	KNS42		ENSG00000147324 MFHAS1
chr6	15532273	0+11-	chr8	8667072	11+0-	CTX	-404	11	KNS42	ENSG00000047579 DTNBP1	ENSG00000147324 MFHAS1
chr6	20112705	2+11-	chr9	117857639	2+21-	CTX	-404	10	KNS42	ENSG00000172197 MBOAT1	ENSG00000041982 TNC
chr6	35754551	0+14-	chr6	35766786	15+1-	ITX	11406	14	KNS42	ENSG00000204140 C6orf127	
chr6	70386596	0+11-	chr6	92147626	12+14-	ITX	21759690	11	KNS42	ENSG00000168216 LMBRD1	
chr6	107307634	16+1-	chr8	105231691	19+17-	CTX	-404	15	KNS42		ENSG00000176406 RIMS2
chr6	119558597	7+18-	chr8	82754452	7+8-	CTX	-404	11	KNS42	ENSG00000111885 MAN1A1	ENSG00000104497 SNX16
chr6	134366639	2+15-	chr22	24198439	16+71-	CTX	-404	15	KNS42	ENSG00000146411 SLC2A12	

chr7	3023122	1+15-	chr7	3359981	17+1-	ITX	335662	15	KNS42	ENSG00000198286 CARD11	ENSG00000146555 SDK1
chr7	21935990	5+8-	chr22	22151474	8+5-	CTX	-404	13	KNS42	ENSG00000105877 DNAH11	ENSG00000100030 MAPK1
chr7	99998836	11+0-	chr11	74292835	0+11-	CTX	-404	11	KNS42	ENSG00000078487 ZCWPW1	ENSG00000077514 POLD3
chr7	103159742	18+0-	chr12	98598388	19+9-	CTX	-404	18	KNS42	ENSG00000189056 RELN	ENSG00000258312
chr7	110756337	2+24-	chr7	110757801	25+2-	ITX	682	24	KNS42	ENSG00000173114 LRRN3	ENSG00000184903 IMMP2L
chr7	136805167	0+21-	chr7	136917293	22+2-	ITX	111162	19	KNS42	ENSG00000234352	ENSG00000105894 PTN
chr7	149736180	1+13-	chr7	149739027	12+0-	ITX	2081	11	KNS42		ENSG00000241449
chr8	15289742	21+14-	chr13	74313749	21+15-	CTX	-404	35	KNS42	ENSG00000104723 TUSC3	ENSG00000118922 KLF12
chr8	52730546	2+60-	chr11	38812130	60+53-	CTX	-404	56	KNS42	ENSG00000168300 PCMTD1	
chr9	12681979	0+34-	chr9	12686650	35+0-	ITX	3470	34	KNS42		ENSG00000107165 TYRP1
chr9	68434547	30+6-	chr13	110874948	6+16-	CTX	-404	20	KNS42	ENSG00000215548	ENSG00000187498 COL4A1
chr10	1114363	8+4-	chr20	5020915	3+8-	CTX	-404	11	KNS42	ENSG00000047056 WDR37	
chr10	101588012	0+14-	chr22	24606864	0+14-	CTX	-404	14	KNS42	ENSG00000023839 ABCC2	
chr10	127575298	4+89-	chr13	20371183	88+2-	CTX	-404	84	KNS42	ENSG00000089876 DHX32	
chr10	127586269	25+4-	chrY	59033113	23+1-	CTX	-404	23	KNS42	ENSG00000203780 FANK1	
chr10	127594410	3+28-	chrY	59025715	8+31-	CTX	-404	27	KNS42	ENSG00000203780 FANK1	
chr11	49838891	1+21-	chr11	49873934	21+0-	ITX	33776	21	KNS42		ENSG00000254714
chr12	20704514	19+12-	chr16	33963841	24+32-	CTX	-404	15	KNS42	ENSG00000172572 PDE3A	
chr12	69291843	13+1-	chr14	59220388	10+1-	CTX	-404	10	KNS42	ENSG00000135678 CPM	
chr12	73239994	21+0-	chr15	94888226	21+1-	CTX	-404	21	KNS42		ENSG00000140563 MCTP2
chr12	96967216	0+14-	chrX	100735303	1+15-	CTX	-404	13	KNS42	ENSG00000188596 C12orf63	ENSG00000196440 ARMCX4
chr12	96977487	0+15-	chrX	98569531	17+3-	CTX	-404	14	KNS42	ENSG00000188596 C12orf63	
chr12	127650496	6+14-	chr17	31149593	7+12-	CTX	-404	10	KNS42	ENSG00000239776	ENSG00000176658 MYO1D
chr13	23487887	1+17-	chr13	24902264	18+2-	ITX	1413472	16	KNS42	ENSG00000248788	
chr17	39579140	1+11-	chr17	39595475	9+0-	ITX	14016	10	KNS42	ENSG00000108417 KRT37	ENSG00000171360 KRT38
chr18	110103	174+139-	chrX	29270354	10+14-	CTX	-404	10	KNS42		ENSG00000169306 IL1RAPL1

chr20	11281915	0+11-	chrX	81651240	0+11-	CTX	-404	11	KNS42	ENSG00000230990	
chr20	55038728	1+19-	chr20	57284430	19+1-	ITX	2244784	18	KNS42		ENSG00000215440 NPEPL1
chr21	11054605	0+29-	chrY	59001093	0+29-	CTX	-404	29	KNS42	ENSG00000187172 BAGE2	
chrX	1317601	1+29-	chrX	1689918	31+2-	ITX	371494	28	KNS42	ENSG00000205755 CRLF2	
chr1	28515699	12+1-	chr19	24027616	10+0-	CTX	-406	10	SF188	ENSG00000169403 PTAFR	
chr1	38433053	12+20-	chr6	28863616	12+20-	CTX	-406	32	SF188	ENSG00000183431 SF3A3	
chr1	91853069	4001+761-	chr8	70602300	2+1060-	CTX	-406	1054	SF188	ENSG00000162669 HFM1	ENSG00000137571 SLCO5A1
chr1	91853162	2939+753-	chr12	127650866	734+3-	CTX	-406	731	SF188	ENSG00000162669 HFM1	ENSG00000239776
chr1	119401206	22+1-	chr15	51189113	1+22-	CTX	-406	21	SF188		ENSG00000247586
chr1	153043779	0+16-	chr1	153066681	15+0-	ITX	22111	15	SF188	ENSG00000196805 SPRR2B	ENSG00000203785 SPRR2E
chr1	164240743	12+1-	chr2	190999553	0+11-	CTX	-406	11	SF188		ENSG00000187699 C2orf88
chr1	178665120	2+13-	chr5	173541191	4+13-	CTX	-406	12	SF188		ENSG00000170091
chr1	225133401	6+12-	chr1	225248304	12+1-	ITX	114194	12	SF188	ENSG00000143799 PARP1	ENSG00000185842 DNAH14
chr1	226056264	1+28-	chr1	226058049	27+1-	ITX	842	26	SF188	ENSG00000143799 PARP1	ENSG00000196187 TMEM63A
chr2	3931958	12+0-	chr12	124499519	12+1-	CTX	-406	11	SF188		ENSG00000179195 ZNF664
chr2	11930066	11+21-	chrX	42764633	0+16-	CTX	-406	13	SF188	ENSG00000134324 LPIN1	
chr2	28965181	1+20-	chr2	29052473	22+0-	ITX	86311	20	SF188		ENSG00000163806 SPDYA
chr2	68914697	10+0-	chr5	64467845	10+1-	CTX	-406	10	SF188	ENSG00000163219 ARHGAP25	ENSG00000049192 ADAMTS6
chr2	132981508	15+10-	chr14	19035116	30+4-	CTX	-406	15	SF188	ENSG00000163046 ANKRD30BL	
chr2	132987675	24+0-	chr15	20000013	26+8-	CTX	-406	21	SF188	ENSG00000163046 ANKRD30BL	
chr2	133006034	19+2-	chr16	33970522	44+10-	CTX	-406	16	SF188	ENSG00000163046 ANKRD30BL	
chr2	162139084	32+4-	chr21	9829856	5+31-	CTX	-406	29	SF188	ENSG00000225813	
chr3	66324954	0+12-	chr3	94513880	14+1-	ITX	28188174	12	SF188	ENSG00000144741 SLC25A26	
chr3	75722542	1+12-	chr4	190935603	0+13-	CTX	-406	12	SF188	ENSG00000242516	
chr3	75984130	1+19-	chr19	33444289	0+28-	CTX	-406	17	SF188	ENSG00000185008 ROBO2	ENSG00000121289 CEP89

chr3	84440288	4+23-	chr4	190840937	1+24-	CTX	-406	22	SF188		ENSG00000245685
chr3	111275066	25+1570-	chr8	128488625	1254+3460-	CTX	-406	1555	SF188	ENSG00000153283 CD96	ENSG00000246228
chr3	124176062	10+0-	chr7	63657750	1+11-	CTX	-406	10	SF188	ENSG00000160145 KALRN	
chr3	170340075	3+8-	chr17	74374118	5+10-	CTX	-406	11	SF188	ENSG00000013297 CLDN11	ENSG00000161542 PRPSAP1
chr3	196625619	16+6-	chr10	42597278	261+215-	ITX	-341	55	SF188	ENSG00000119231 SENP5	
chr3	196625676	16+26-	chr10	42390288	56+116-	CTX	-406	20	SF188	ENSG00000119231 SENP5	
chr4	1617097	0+13-	chr20	9816311	15+1-	CTX	-406	13	SF188		ENSG00000101349 PAK7
chr4	52798913	31+710-	chr4	54214944	712+31-	ITX	1408910	688	SF188		ENSG00000184178 SCFD2
chr4	53474896	0+17-	chr4	55131340	16+1-	ITX	1655630	16	SF188	ENSG00000109189 USP46	ENSG00000134853 PDGFRA
chr4	53503816	2+17-	chr4	55608654	18+1-	ITX	2104056	15	SF188	ENSG00000109189 USP46	
chr4	53882777	1+35-	chr9	88287132	1+35-	CTX	-406	35	SF188	ENSG00000184178 SCFD2	ENSG00000135049 AGTPBP1
chr4	54044675	28+670-	chr8	128265420	1041+446-	CTX	-406	643	SF188	ENSG00000184178 SCFD2	
chr4	54127071	2+36-	chr17	29741322	0+34-	CTX	-406	34	SF188	ENSG00000184178 SCFD2	ENSG00000131242 RAB11FIP4
chr4	54165402	216+618-	chr8	128043959	469+31-	CTX	-406	435	SF188	ENSG00000184178 SCFD2	
chr4	54211080	27+3-	chr9	71691856	2+23-	CTX	-406	22	SF188	ENSG00000184178 SCFD2	ENSG00000165060 FXN
chr4	54224715	0+15-	chr4	55298688	17+1-	ITX	1006086	16	SF188	ENSG00000184178 SCFD2	
chr4	54283618	56+743-	chr11	118277406	99+788-	CTX	-406	692	SF188	ENSG00000145216 FIP1L1	ENSG00000167283 ATP5L
chr4	54291376	661+75-	chr8	128942087	1470+1485-	CTX	-406	585	SF188	ENSG00000145216 FIP1L1	ENSG00000249859 PVT1
chr4	54327856	30+63-	chr4	54561209	21+0-	ITX	231409	21	SF188	ENSG00000072201 LNX1	ENSG00000145216 FIP1L1
chr4	55045259	1+23-	chr17	29881541	0+23-	CTX	-406	23	SF188	ENSG00000145216 FIP1L1	
chr4	55068948	0+25-	chr4	55763485	25+0-	ITX	693729	25	SF188	ENSG00000145216 FIP1L1	
chr4	55559032	1+36-	chr17	29677191	0+35-	CTX	-406	35	SF188	ENSG00000157404 KIT	ENSG00000196712 NF1
chr4	80894424	14+1-	chr5	21207753	0+14-	CTX	-406	13	SF188	ENSG00000163297 ANTXR2	
chr4	81060458	1+15-	chr4	81190083	15+0-	ITX	128524	15	SF188		ENSG00000138675 FGF5
chr4	88186116	49+35-	chr16	23176680	8+48-	CTX	-406	48	SF188	ENSG00000250572	
chr4	172619946	21+1-	chr7	23309549	15+21-	CTX	-406	21	SF188		ENSG00000136235 GPNMB

chr5	159350015	0+44-	chrX	66982357	44+2-	CTX	-406	43	SF188	ENSG00000170214 ADRA1B	
chr5	160531888	9+3-	chr10	96337717	3+7-	CTX	-406	10	SF188		ENSG00000119969 HELLS
chr5	169544250	11+15-	chr5	171214306	16+10-	ITX	1668039	15	SF188		ENSG00000185662 C5orf50
chr6	35754559	2+42-	chr6	35766982	42+3-	ITX	11404	41	SF188	ENSG00000204140 C6orf127	
chr6	36013499	10+0-	chr11	107984106	2+11-	CTX	-406	10	SF188	ENSG00000112062 MAPK14	ENSG00000255467
chr6	132122317	0+17-	chr6	154332127	16+10-	ITX	22208698	16	SF188		ENSG00000112038 OPRM1
chr6	158783682	22+35-	chr17	78736108	35+24-	CTX	-406	52	SF188	ENSG00000130338 TULP4	ENSG00000141564 RPTOR
chr7	111053658	12+8-	chr12	108202919	8+14-	CTX	-406	20	SF188	ENSG00000184903 IMMP2L	
chr7	149736176	4+12-	chr7	149739225	14+3-	ITX	2057	12	SF188		ENSG00000241449
chr7	154876549	36+40-	chr7	154881085	41+35-	ITX	3153	40	SF188	ENSG00000157219 HTR5A	
chr8	15289876	25+33-	chr13	74313847	24+32-	CTX	-406	56	SF188	ENSG00000104723 TUSC3	ENSG00000118922 KLF12
chr8	52730430	0+38-	chr11	38812243	39+50-	CTX	-406	37	SF188	ENSG00000168300 PCMTD1	
chr8	96265842	1+15-	chr19	18835272	15+0-	CTX	-406	14	SF188	ENSG00000156172 C8orf37	ENSG00000105662 CRTC1
chr8	128215568	42+1171-	chr12	58019610	250+1356-	CTX	-406	1124	SF188		ENSG00000135454 B4GALNT1
chr8	128997301	1390+791-	chr12	58150692	1391+792-	CTX	-406	614	SF188	ENSG00000249859 PVT1	ENSG00000139266 MARCH9
chr9	3107616	13+10-	chr9	4190288	10+0-	ITX	1081996	10	SF188		ENSG00000107249 GLIS3
chr9	73105725	0+15-	chr17	29767045	14+0-	CTX	-406	14	SF188		ENSG00000131242 RAB11FIP4
chr9	73130362	2+31-	chr17	29702985	33+1-	CTX	-406	31	SF188		ENSG00000196712 NF1
chr9	73158216	27+0-	chr17	29706699	28+1-	CTX	-406	27	SF188	ENSG00000083067 TRPM3	ENSG00000196712 NF1
chr9	88246706	24+1-	chr17	29877072	24+1-	CTX	-406	24	SF188	ENSG00000135049 AGTPBP1	
chr9	88293892	15+0-	chr17	29768963	1+14-	CTX	-406	13	SF188	ENSG00000135049 AGTPBP1	ENSG00000131242 RAB11FIP4
chr9	109305732	4+15-	chr9	109354148	15+0-	ITX	47573	15	SF188	ENSG00000234323	ENSG00000229297
chr9	124505824	1+15-	chr16	30630816	15+1-	CTX	-406	15	SF188	ENSG00000136848 DAB2IP	
chr9	124521292	10+0-	chr16	66292217	1+11-	CTX	-406	10	SF188	ENSG00000136848 DAB2IP	
chr10	101588047	4+10-	chr22	24606887	3+9-	CTX	-406	12	SF188	ENSG00000023839 ABCC2	
chr10	127575372	7+132-	chr13	20371210	131+1-	CTX	-406	129	SF188	ENSG00000089876 DHX32	

chr10	127586240	26+4-	chrY	59032305	25+4-	CTX	-406	19	SF188	ENSG00000203780 FANK1	
chr10	127594439	5+48-	chrY	59025664	13+61-	CTX	-406	48	SF188	ENSG00000203780 FANK1	
chr11	29534037	1+46-	chr11	69947170	47+1-	ITX	40412260	46	SF188	ENSG00000254530	ENSG00000131620 ANO1
chr11	57812425	19+1-	chr20	3753964	20+0-	CTX	-406	19	SF188	ENSG00000186509 OR9Q1	
chr12	73239952	28+0-	chr15	94888163	28+2-	CTX	-406	28	SF188		ENSG00000140563 MCTP2
chr14	105628776	64+66-	chr14	105628961	64+66-	ITX	-361	34	SF188	ENSG00000184916 JAG2	ENSG00000257622
chr15	75622782	0+12-	chr15	75739744	13+1-	ITX	116118	11	SF188		ENSG00000169375 SIN3A
chr16	542305	24+25-	chr16	542428	24+25-	ITX	-323	14	SF188	ENSG00000090565 RAB11FIP3	ENSG00000256323
chr16	23640617	0+40-	chr16	23914058	41+2-	ITX	272186	40	SF188	ENSG00000083093 PALB2	ENSG00000166501 PRKCB
chr16	23791248	0+48-	chr16	26084062	49+1-	ITX	2290900	45	SF188		ENSG00000182601 HS3ST4
chr16	48280936	10+11-	chr16	48281199	10+11-	ITX	-325	10	SF188	ENSG00000102910 LONP2	ENSG00000121270 ABCC11
chr16	56411916	2+18-	chr16	56829788	18+5-	ITX	416526	16	SF188	ENSG00000159461 AMFR	ENSG00000102900 NUP93
chr17	36636571	14+0-	chr19	10029848	15+6-	CTX	-406	13	SF188	ENSG00000225485 ARHGAP23	ENSG00000105088 OLFM2
chr18	110081	336+249-	chrX	29270196	27+58-	CTX	-406	26	SF188		ENSG00000169306 IL1RAPL1
chr19	33444427	0+11-	chr20	26220364	10+1-	CTX	-406	10	SF188	ENSG00000121289 CEP89	
chr20	278393	0+18-	chr21	17652744	22+35-	CTX	-406	18	SF188	ENSG00000177764 ZCCHC3	ENSG00000215386 LINC00478
chr20	1840412	14+0-	chr21	17125470	1+15-	CTX	-406	14	SF188		ENSG00000155313 USP25
chr20	4895215	1+32-	chr20	25459154	31+0-	ITX	20562950	31	SF188	ENSG00000089057 SLC23A2	ENSG00000101004 NINL
chr21	11054613	1+28-	chrY	59000805	14+35-	CTX	-406	27	SF188	ENSG00000187172 BAGE2	
chrX	1236190	1+20-	chrX	1321896	18+0-	ITX	84927	18	SF188		ENSG00000205755 CRLF2
chr1	569070	14+0-	chr8	63015045	20+35-	CTX	-453	14	UW479	ENSG00000230021	
chr1	20371660	1+16-	chr3	87620063	15+25-	CTX	-453	15	UW479	ENSG00000127472 PLA2G5	
chr1	27103911	1+19-	chr1	27340815	19+2-	ITX	235149	17	UW479	ENSG00000117713 ARID1A	
chr1	28515712	96+1-	chr19	24027325	108+12-	CTX	-453	96	UW479	ENSG00000169403 PTAFR	
chr1	38328968	2+25-	chr1	38428686	24+0-	ITX	98648	23	UW479	ENSG00000204084 INPP5B	ENSG00000183431 SF3A3
chr1	47836334	1+34-	chr14	53815391	0+34-	CTX	-453	33	UW479	ENSG00000162368 CMPK1	ENSG00000237356

chr1	55029942	27+0-	chr22	19435945	0+29-	CTX	-453	27	UW479	ENSG00000162390 ACOT11	ENSG00000185065
chr1	59096776	17+0-	chr2	32048520	0+20-	CTX	-453	17	UW479	ENSG00000185839	ENSG00000242272 AK2P2
chr1	91853126	3069+471-	chr8	70602300	10+1115-	CTX	-453	1106	UW479	ENSG00000162669 HFM1	ENSG00000137571 SLCO5A1
chr1	91853219	1963+459-	chr12	127650864	454+3-	CTX	-453	451	UW479	ENSG00000162669 HFM1	ENSG00000239776
chr1	103311283	20+5-	chr5	102750041	18+2-	CTX	-453	15	UW479	ENSG00000230864	
chr1	109495349	16+13-	chr3	110413011	13+18-	CTX	-453	28	UW479	ENSG00000121940 CLCC1	
chr1	112960870	24+1-	chr12	52615093	23+1-	CTX	-453	23	UW479	ENSG00000143079 CTTNBP2NL	ENSG00000258279
chr1	113149456	18+0-	chr20	3902968	17+18-	CTX	-453	18	UW479	ENSG00000007341 ST7L	ENSG00000125779 PANK2
chr1	153299201	1+46-	chr1	153641153	52+8-	ITX	333306	45	UW479		ENSG00000143621 ILF2
chr1	154914163	0+42-	chr1	155232085	41+2-	ITX	316908	41	UW479		ENSG00000116521 SCAMP3
chr1	161668108	0+14-	chr1	161800589	13+0-	ITX	131721	13	UW479		ENSG00000118217 ATF6
chr1	164240569	12+14-	chr2	190998938	14+1-	CTX	-453	14	UW479		ENSG00000187699 C2orf88
chr1	168025510	102+0-	chr19	24031551	11+113-	CTX	-453	102	UW479	ENSG00000143164 DCAF6	
chr1	170926316	0+14-	chr6	123171229	24+1-	CTX	-453	13	UW479	ENSG00000117501 C1orf129	
chr1	173909966	1+22-	chr1	174187941	24+2-	ITX	276849	21	UW479	ENSG00000135870 RC3H1	ENSG00000152061 RABGAP1L
chr1	182275010	21+0-	chr17	74319270	0+21-	CTX	-453	21	UW479	ENSG00000228918	ENSG00000161542 PRPSAP1
chr1	189715582	2+13-	chr11	32042611	0+13-	CTX	-453	13	UW479		ENSG00000049449 RCN1
chr1	200068784	16+5-	chr2	23697827	18+4-	CTX	-453	18	UW479	ENSG00000116833 NR5A2	ENSG00000119771 KLHL29
chr1	207579542	21+1-	chr6	57493597	25+6-	CTX	-453	21	UW479		ENSG00000146143 PRIM2
chr1	219923945	15+4-	chr12	98034739	17+1-	CTX	-453	15	UW479	ENSG00000196660 SLC30A10	
chr1	225133402	3+12-	chr1	225248296	11+0-	ITX	114141	11	UW479	ENSG00000143799 PARP1	ENSG00000185842 DNAH14
chr1	240091360	15+25-	chr22	29065227	144+50-	CTX	-453	40	UW479		ENSG00000100154 TTC28
chr2	11432966	0+22-	chr2	11545236	25+3-	ITX	111348	22	UW479	ENSG00000134318 ROCK2	
chr2	18253884	2+20-	chr2	19077190	20+2-	ITX	822294	18	UW479	ENSG00000170745 KCNS3	
chr2	42052748	35+8-	chr4	66413204	9+35-	CTX	-453	37	UW479		ENSG00000145242 EPHA5
chr2	68914690	11+0-	chr5	64467726	12+1-	CTX	-453	11	UW479	ENSG00000163219 ARHGAP25	ENSG00000049192 ADAMTS6

chr2	79376836	10+0-	chr13	107845299	22+3-	CTX	-453	10	UW479		ENSG00000204442 FAM155A
chr2	91937509	13+6-	chr12	123873181	11+2-	CTX	-453	10	UW479	ENSG00000223703	ENSG00000183955 SETD8
chr2	99167934	23+0-	chr11	46252923	23+0-	CTX	-453	23	UW479	ENSG00000040933 INPP4A	
chr2	132977273	4+13-	chr14	19034839	39+38-	CTX	-453	12	UW479	ENSG00000163046 ANKRD30BL	
chr2	132987614	28+0-	chr15	20000009	57+31-	CTX	-453	27	UW479	ENSG00000163046 ANKRD30BL	
chr2	133001831	1+36-	chr14	19000169	7+46-	CTX	-453	36	UW479	ENSG00000163046 ANKRD30BL	
chr2	133004923	7+0-	chr21	10707760	184+173-	ITX	-394	62	UW479	ENSG00000163046 ANKRD30BL	
chr2	162138495	26+4-	chr21	9829856	0+22-	CTX	-453	22	UW479	ENSG00000225813	
chr2	187514641	2+11-	chr2	188784801	11+1-	ITX	1269112	10	UW479	ENSG00000138448 ITGAV	ENSG00000231689
chr3	49442178	1+18-	chr12	86120307	2+19-	CTX	-453	18	UW479	ENSG00000067560 RHOA	
chr3	50879613	28+5-	chr5	81987180	25+6-	CTX	-453	30	UW479	ENSG00000088538 DOCK3	
chr3	52435832	19+0-	chr19	56217031	19+3-	CTX	-453	17	UW479	ENSG00000163930 BAP1	
chr3	53300460	1+14-	chr19	53495576	15+3-	CTX	-453	12	UW479		ENSG00000242779 ZNF702P
chr3	53335607	17+1-	chr19	52478810	2+19-	CTX	-453	16	UW479	ENSG00000162290 DCP1A	ENSG00000246125
chr3	55024288	37+0-	chr19	47734383	1+37-	CTX	-453	37	UW479	ENSG00000157445 CACNA2D3	ENSG00000105327 BBC3
chr3	55380592	0+16-	chr19	52614592	20+3-	CTX	-453	15	UW479	ENSG00000242568	
chr3	56766012	19+21-	chr7	147538955	22+20-	CTX	-453	40	UW479	ENSG00000163947 ARHGEF3	ENSG00000174469 CNTNAP2
chr3	85686452	12+4-	chr16	11673269	4+15-	CTX	-453	11	UW479	ENSG00000175161 CADM2	ENSG00000189067 LITAF
chr3	87410003	21+4-	chr7	78953929	13+13-	CTX	-453	13	UW479		ENSG00000187391 MAGI2
chr3	99389332	0+27-	chr15	26705465	29+13-	CTX	-453	27	UW479	ENSG00000144810 COL8A1	
chr3	110889290	14+0-	chrX	57292578	15+1-	CTX	-453	14	UW479	ENSG00000177707 PVRL3	
chr3	111274994	2+110-	chr8	128533275	8+117-	CTX	-453	110	UW479	ENSG00000153283 CD96	
chr3	114154875	0+18-	chr3	168102681	18+0-	ITX	53946956	18	UW479	ENSG00000181722 ZBTB20	ENSG00000206120 EGFEM1P
chr3	134299172	36+1-	chr6	36913072	36+1-	CTX	-453	34	UW479		ENSG00000164530 PI16
chr3	139226960	2+21-	chr19	53106091	4+24-	CTX	-453	19	UW479	ENSG00000248932	
chr3	142553308	0+83-	chr14	22776228	1+84-	CTX	-453	83	UW479	ENSG00000163710 PCOLCE2	

chr3	178769835	1+16-	chr15	87685223	18+13-	CTX	-453	15	UW479	ENSG00000172667 ZMAT3	
chr4	80894474	21+0-	chr19	31113491	3+23-	CTX	-453	21	UW479	ENSG00000163297 ANTXR2	
chr4	95518653	21+10-	chr7	46904209	13+22-	CTX	-453	31	UW479	ENSG00000163110 PDLIM5	
chr4	112258318	0+14-	chr4	112816034	14+1-	ITX	556933	14	UW479		ENSG00000249815
chr4	186117225	40+31-	chr13	40674751	3+33-	CTX	-453	31	UW479	ENSG00000164323 KIAA1430	
chr5	100497538	0+11-	chr21	26214157	11+0-	CTX	-453	11	UW479		ENSG00000226983
chr6	150670	7+31-	chr11	191186	34+13-	CTX	-453	31	UW479		ENSG00000177951 BET1L
chr6	20112078	10+1-	chr9	117857715	14+11-	CTX	-453	10	UW479	ENSG00000172197 MBOAT1	ENSG00000041982 TNC
chr6	45229405	26+2-	chr8	62847098	26+3-	CTX	-453	23	UW479	ENSG00000196284 SUPT3H	ENSG00000254119
chr6	58779262	790+768-	chr7	53801572	6+22-	CTX	-453	17	UW479		ENSG00000205628
chr6	116774524	4+12-	chr7	26251457	32+22-	CTX	-453	11	UW479	ENSG00000217241 CBX3P9	ENSG00000122565 CBX3
chr6	126239149	5+49-	chr20	34416968	56+4-	CTX	-453	48	UW479	ENSG00000111912 NCOA7	ENSG00000025293 PHF20
chr7	22120450	1+22-	chr17	73963834	2+23-	CTX	-453	22	UW479		ENSG00000161533 ACOX1
chr7	68695954	0+22-	chr7	106200173	22+0-	ITX	37503284	22	UW479	ENSG00000225718	ENSG00000243797
chr7	71425213	0+21-	chr7	99282578	23+1-	ITX	27856510	12	UW479	ENSG00000183166 CALN1	
chr7	88827600	30+2-	chr8	143781332	35+7-	CTX	-453	30	UW479	ENSG00000182348 ZNF804B	
chr7	101667589	1+14-	chr8	91903678	14+16-	CTX	-453	12	UW479	ENSG00000257923 CUX1	ENSG00000123119 NECAB1
chr7	111053660	44+41-	chr12	108202877	43+47-	CTX	-453	82	UW479	ENSG00000184903 IMMP2L	
chr7	147472510	5+8-	chr17	35071532	9+6-	CTX	-453	11	UW479	ENSG00000174469 CNTNAP2	
chr8	14946075	21+26-	chrX	124433507	25+21-	CTX	-453	46	UW479	ENSG00000185053 SGCZ	
chr8	40433239	0+27-	chr14	55716413	29+5-	CTX	-453	27	UW479	ENSG00000165061 ZMAT4	
chr8	47159193	31+0-	chr9	137280905	30+0-	CTX	-453	30	UW479		ENSG00000186350 RXRA
chr8	51886715	5+6-	chr12	108002659	6+4-	CTX	-453	10	UW479		ENSG00000151136 BTBD11
chr8	52730477	1+57-	chr11	38812217	56+64-	CTX	-453	56	UW479	ENSG00000168300 PCMTD1	
chr8	57896780	1+16-	chr19	31716239	20+8-	CTX	-453	14	UW479	ENSG00000104331 IMPAD1	
chr8	62579337	1+32-	chr8	62708134	31+0-	ITX	127822	31	UW479	ENSG00000198363 ASPH	ENSG00000254119

chr8	96643384	20+19-	chr18	4086683	19+19-	CTX	-453	38	UW479	ENSG00000253773	ENSG00000170579 DLGAP1
chr8	117230013	20+9-	chr17	49354767	15+0-	CTX	-453	15	UW479	ENSG00000249917	ENSG00000011260 UTP18
chr8	134879116	0+10-	chr8	135683085	11+0-	ITX	803270	10	UW479		ENSG00000066827 ZFAT
chr8	141832537	22+3-	chr15	42650218	1+19-	CTX	-453	19	UW479	ENSG00000169398 PTK2	ENSG00000092529 CAPN3
chr8	145554246	0+21-	chr8	145655751	23+0-	ITX	100345	21	UW479		ENSG00000160949 TONSL
chr8	146058423	1+12-	chr17	27112965	0+11-	CTX	-453	11	UW479	ENSG00000147789 ZNF7	ENSG00000173065 C17orf63
chr9	22741727	0+13-	chr9	128368012	16+4-	ITX	105624816	13	UW479	ENSG00000234840	ENSG00000119487 MAPKAP1
chr9	25865280	21+0-	chr18	31751559	27+21-	CTX	-453	21	UW479		ENSG00000101746 NOL4
chr9	27935991	0+21-	chr9	135661990	23+3-	ITX	-96797295	21	UW479		ENSG00000165695 AK8
chr9	27945060	12+1-	chr18	30711907	12+0-	CTX	-453	12	UW479		ENSG00000166960 C18orf34
chr9	28000942	0+14-	chr9	86632969	14+1-	ITX	58631084	14	UW479	ENSG00000174482 LINGO2	
chr9	30392631	0+12-	chr9	86610459	14+2-	ITX	56216928	12	UW479		ENSG00000178966 RMI1
chr9	32381210	0+21-	chr9	33849115	20+0-	ITX	1466998	20	UW479		ENSG00000107341 UBE2R2
chr9	78899983	22+0-	chr17	18077236	7+27-	CTX	-453	22	UW479	ENSG00000099139 PCSK5	ENSG00000091536 MYO15A
chr9	86638539	0+14-	chr18	30729821	2+15-	CTX	-453	13	UW479		ENSG00000166960 C18orf34
chr9	102086745	0+17-	chr18	31750899	20+2-	CTX	-453	17	UW479		ENSG00000101746 NOL4
chr9	102146807	14+0-	chr18	30736184	17+2-	CTX	-453	14	UW479		ENSG00000166960 C18orf34
chr9	131081825	0+20-	chr18	14823395	20+15-	CTX	-453	20	UW479	ENSG00000167112 TRUB2	ENSG00000180777 ANKRD30B
chr9	136327728	1+22-	chr14	56290681	4+24-	CTX	-453	22	UW479	ENSG00000160325 C9orf7	
chr10	1114901	12+0-	chr20	5021199	2+14-	CTX	-453	11	UW479	ENSG00000047056 WDR37	
chr10	5202158	0+10-	chr10	5324856	10+1-	ITX	121967	10	UW479	ENSG00000196326 AKR1CL1	ENSG00000215267
chr10	26226198	7+4-	chr17	6554228	4+7-	CTX	-453	11	UW479	ENSG00000095777 MYO3A	ENSG00000108590 MED31
chr10	34822411	9+26-	chr12	55472537	26+0-	CTX	-453	26	UW479	ENSG00000148498 PARD3	
chr10	45422946	25+35-	chr10	45424592	37+25-	ITX	290	35	UW479	ENSG00000187783 TMEM72	ENSG00000224812
chr10	109325960	27+23-	chr13	107846599	48+70-	CTX	-453	47	UW479		ENSG00000204442 FAM155A
chr10	127575172	1+52-	chr13	20371005	52+4-	CTX	-453	52	UW479	ENSG00000089876 DHX32	

chr10	127586296	16+5-	chrY	59032932	11+1-	CTX	-453	10	UW479	ENSG00000203780 FANK1	
chr11	49839008	2+22-	chr11	49873895	19+0-	ITX	33689	19	UW479		ENSG00000254714
chr11	55373517	31+3-	chr18	76876931	30+0-	CTX	-453	29	UW479		ENSG00000166377 ATP9B
chr12	12453997	1+19-	chr20	62851844	3+20-	CTX	-453	18	UW479		ENSG00000196132 MYT1
chr12	20704524	8+25-	chr16	33963005	28+1259-	CTX	-453	20	UW479	ENSG00000172572 PDE3A	
chr13	23487980	1+25-	chr13	24902217	25+0-	ITX	1413371	25	UW479	ENSG00000248788	
chr13	89981969	0+18-	chr15	39504892	1+18-	CTX	-453	16	UW479		ENSG00000259345
chr13	100589796	1+17-	chr21	34847166	0+18-	CTX	-453	16	UW479	ENSG00000255768	ENSG00000142188 TMEM50B
chr14	31036241	0+17-	chr15	30984655	0+16-	CTX	-453	16	UW479	ENSG00000092140 G2E3	ENSG00000175344 CHRNA7
chr14	32066014	1+47-	chr14	33251221	46+1-	ITX	1183569	45	UW479	ENSG00000151413 NUBPL	ENSG00000151320 AKAP6
chr14	67834188	1+12-	chr14	67862968	13+2-	ITX	28007	12	UW479	ENSG00000134001 EIF2S1	ENSG00000100558 PLEK2
chr14	88035460	3+92-	chr19	33010389	104+8-	CTX	-453	90	UW479	ENSG00000258859	
chr15	34326914	1+24-	chr15	34419814	28+4-	ITX	91589	23	UW479	ENSG00000169857 AVEN	
chr15	35352133	1+16-	chr17	8516724	5+43-	CTX	-453	15	UW479		ENSG00000133026 MYH10
chr15	85867818	2+26-	chr15	85913538	27+4-	ITX	44137	24	UW479	ENSG00000218052	ENSG00000170776 AKAP13
chr17	8745523	1+17-	chr17	9557484	20+1-	ITX	810986	15	UW479	ENSG00000174083 PIK3R6	ENSG00000154914 USP43
chr17	17916041	5+20-	chr17	18055066	35+22-	ITX	137263	15	UW479	ENSG00000171962 LRRC48	ENSG00000091536 MYO15A
chr17	41713672	24+1-	chr22	18470346	28+5-	CTX	-453	22	UW479		ENSG00000093100
chr19	23934384	470+472-	chr19	23943716	470+472-	ITX	-413	354	UW479	ENSG00000196172 ZNF681	
chr19	52505400	2+21-	chr19	56253579	22+1-	ITX	3746818	19	UW479	ENSG00000197619 ZNF615	
chr19	53186921	0+27-	chr19	53482483	29+8-	ITX	294327	22	UW479	ENSG00000167766 ZNF83	ENSG00000242779 ZNF702P
chr19	55475788	0+14-	chr19	55476878	14+0-	ITX	299	14	UW479	ENSG00000167634 NLRP7	ENSG00000022556 NLRP2
chr20	31521723	2+34-	chr20	31553810	32+0-	ITX	31082	32	UW479	ENSG00000215529 EFCAB8	
chr21	11055054	2+25-	chrY	59001084	0+21-	CTX	-453	19	UW479	ENSG00000187172 BAGE2	

Abbreviations : Chr – chromosome; Pos – position ; Ori – reads orientation;

Appendix VI

Table S5.2. Genic structural variants predicted by the Breakdancer in which either region involved has copy number gain or loss has.

Cell Line Fusion ID	Gene 1	Gene 2	Gene1.diff.smo	Gene2.diff.smo
KNS42_1_156051752_3_189542047		TP63		0.4177
KNS42_2_171814878_13_49848076	GORASP2	CDADC1	0.02305	0.3368
KNS42_1_58802590_10_74176383	DAB1	CBARA1	0.2695	0.3825
KNS42_2_43437013_6_45891802	THADA	CLIC5	0.1975	0.2072
KNS42_6_14261299_8_8718282		MFHAS1		0.3237
KNS42_6_15532273_8_8667072	DTNBP1	MFHAS1	0.2771	0.3237
KNS42_1_152387895_11_132713260		OPCML		0.5185
KNS42_12_96967216_X_100735303	C12orf63		0.2565	
KNS42_7_103159742_12_98598388	RELN		0.2606	
KNS42_3_164722685_5_57093054	SI		0.4412	
KNS42_12_96977487_X_98569531	C12orf63		0.2565	
KNS42_5_64964650_5_65433006	SGTB		0.2663	
SF188_3_44466690_9_73176666		TRPM3		0.405
SF188_20_4895215_20_25459154	SLC23A2		0.3601	
SF188_3_75984130_19_33444289	ROBO2		0.2728	
SF188_9_73105725_17_29767045		RAB11FIP4		0.4491
SF188_3_44853404_17_29848790	KIF15	RAB11FIP4	0.8036	0.4491
SF188_9_88293892_17_29768963	AGTPBP1	RAB11FIP4	0.4196	0.4491
SF188_4_54127071_17_29741322	SCFD2	RAB11FIP4	0.1755	0.4491
SF188_3_44844784_17_29795166	KIF15	RAB11FIP4	0.8036	0.4491
SF188_11_29534037_11_69947170		ANO1		1.6909
SF188_4_53882777_9_88287132	SCFD2	AGTPBP1	0.0166	0.4196
SF188_6_158783682_17_78736108	TULP4		0.2787	
SF188_4_54327856_4_54561209	LNX1	FIP1L1	1.976	2.1076
SF188_20_1840412_21_17125470		USP25		0.2904
SF188_4_54211080_9_71691856	SCFD2		0.5206	
SF188_11_47249941_11_118298167		ATP5L		2.2766
SF188_4_54283618_11_118277406	FIP1L1	ATP5L	2.1076	2.2766
SF188_1_144680231_1_144954567		PDE4DIP		0.2049
SF188_11_88909186_15_27561047		GABRG3		0.1039
SF188_3_44468743_4_53879701		SCFD2		0.5206
SF188_4_52798913_4_54214944		SCFD2		0.5206
SF188_9_73158216_17_29706699	TRPM3		0.405	
SF188_20_278393_21_17652744		C21orf34		0.4874
SF188_4_54291376_8_128942087	FIP1L1		2.1076	
SF188_16_542305_16_542428	RAB11FIP3		0.3206	
SF188_4_54165402_8_128043959	SCFD2		0.5206	
SF188_9_88246706_17_29877072	AGTPBP1		0.4196	
SF188_4_55045259_17_29881541	FIP1L1		2.1076	
SF188_4_54224715_4_55298688	SCFD2		0.5206	
SF188_4_55068948_4_55763485	FIP1L1		2.1076	
SF188_4_54044675_8_128265420	SCFD2		0.0166	
SF188_4_54224715_4_55298688	SCFD2		0.0166	
SF188_4_54165402_8_128043959	SCFD2		0.0166	
SF188_4_54224715_4_55298688	SCFD2		0.1755	
SF188_11_57812425_20_3753964	OR9Q1		0.4632	
UW479_8_137751825_17_49361503		UTP18		0.4898

UW479_8_117230013_17_49354767		UTP18		0.4898
UW479_8_141832537_15_42650218	PTK2	CAPN3	0.01635	0.2261
UW479_3_47028066_22_19936322	NBEAL2		0.3468	
UW479_3_55024288_19_47734383	CACNA2D3		0.6856	
UW479_9_32381210_9_33849115		UBE2R2		0.5006
UW479_1_113149456_20_3902968	ST7L		0.2146	
UW479_14_32066014_14_33251221	NUBPL	AKAP6	0.3064	0.3727
UW479_17_8745523_17_9557484	PIK3R6		0.3554	
UW479_9_27945060_18_30711907		C18orf34		0.4284
UW479_9_102146807_18_30736184		C18orf34		0.4284
UW479_9_86638539_18_30729821		C18orf34		0.4284
UW479_1_82756062_14_67423798		GPHN		0.4351
UW479_14_31036241_15_30984655	G2E3		0.2964	
UW479_12_1695918_12_9246077		A2M		0.2579
UW479_8_47159193_9_137280905		RXRA		0.4528
UW479_12_3690145_20_62854780	PRMT8		0.3924	
UW479_3_114154875_3_168102681	ZBTB20		0.2724	
UW479_20_15544438_22_35396912	MACROD2		0.2331	
UW479_5_93556435_22_35366567	C5orf36		0.2471	
UW479_1_47836334_14_53815391	CMPK1		0.3541	
UW479_6_45229405_8_62847098	SUPT3H		0.1742	
UW479_8_62579337_8_62708134	ASPH		0.2395	
UW479_8_57896780_19_31716239	IMPAD1		0.2229	
UW479_7_71425213_7_99282578	CALN1		0.4129	
UW479_8_40433239_14_55716413	ZMAT4		0.4246	
UW479_9_28000942_9_86632969	LINGO2		0.7717	
UW479_7_88827600_8_143781332	ZNF804B		0.516	

Gene1.diff.smo and Gene2.diff.smo refer to smoothed log ratios covering each gene (Gene 1 and Gene 2) involved in the SV.

9. Additional publications in the PhD time course

9. ADDITIONAL PUBLICATIONS IN THE PHD TIME COURSE

Human Cancer Biology

Clinical
Cancer
Research

A Distinct Spectrum of Copy Number Aberrations in Pediatric High-Grade Gliomas

Dorine A. Bax¹, Alan Mackay³, Suzanne E. Little¹, Diana Carvalho^{1,6,7}, Marta Viana-Pereira^{1,6}, Narinder Tamber³, Anita E. Grigoriadis⁴, Alan Ashworth³, Rui M. Reis⁶, David W. Ellison⁸, Safa Al-Sarraj⁵, Darren Hargrave², and Chris Jones^{1,2}

Abstract

Purpose: As genome-scale technologies begin to unravel the complexity of the equivalent tumors in adults, we can attempt detailed characterization of high-grade gliomas in children, that have until recently been lacking. Toward this end, we sought to validate and extend investigations of the differences between pediatric and adult tumors.

Experimental Design: We carried out copy number profiling by array comparative genomic hybridization using a 32K bacterial artificial chromosome platform on 63 formalin-fixed paraffin-embedded cases of high-grade glioma arising in children and young people (<23 years).

Results: The genomic profiles of these tumors could be subclassified into four categories: those with stable genomes, which were associated with a better prognosis; those with aneuploid and those with highly rearranged genomes; and those with an amplifier genotype, which had a significantly worse clinical outcome. Independent of this was a clear segregation of cases with 1q gain (more common in children) from those with concurrent 7 gain/10q loss (a defining feature of adults). Detailed mapping of all the amplification and deletion events revealed numerous low-frequency amplifications, including *IGF1R*, *PDGFRB*, *PIK3CA*, *CDK6*, *CCND1*, and *CCNE1*, and novel homozygous deletions encompassing unknown genes, including those at 5q35, 10q25, and 22q13. Despite this, aberrations targeting the “core signaling pathways” in adult glioblastomas are significantly underrepresented in the pediatric setting.

Conclusions: These data highlight that although there are overlaps in the genomic events driving gliomagenesis of all ages, the pediatric disease harbors a distinct spectrum of copy number aberrations compared with adults. *Clin Cancer Res*; 16(13); 3368–77. ©2010 AACR.

The use of genome-scale profiling techniques to identify the key genetic aberrations underlying various tumor types has led to fundamental discoveries about the drivers of oncogenesis, and provides the rationale for specific targeted therapies in these lesions. Until recently, the application of such studies to the fields of high-grade glioma specifically,

and childhood tumors in general, have lagged behind the adult epithelial cancers. This is now rapidly changing, with large screens of adult glioblastoma through collaborative networks (1) or single institutions (2) joining an increasing number of smaller independent studies (3–7) in comprehensively mapping the glioblastoma genome.

There are also beginning to emerge genomic studies specifically addressing childhood cancers, and there is mounting evidence that the pediatric high-grade glioma genome has certain key differences with that of histologically similar adult tumors. An early study using metaphase comparative genomic hybridization (CGH; ref. 8) highlighted distinct chromosomal changes in 23 childhood cases, a result borne out in a later 10K single nucleotide polymorphism (SNP) array study of a further 14 high-grade tumors (9), and more recent studies of 18 pediatric glioblastoma on Illumina 100K arrays (10), and 20 high-grade tumors using molecular inversion probes (11). Most recently, we participated in a collaborative effort to carry out molecular profiling of 78 pediatric high-grade gliomas by Affymetrix 500K SNP and U133 Plus2.0 expression arrays (12). From all these studies, it seems clear that although there are many large-scale chromosomal and

Authors' Affiliations: ¹Section of Paediatric Oncology, The Institute of Cancer Research, and ²Paediatric Oncology, Royal Marsden Hospital, Sutton, United Kingdom; ³Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, and ⁴Breakthrough Breast Cancer Unit and ⁵Clinical Neuropathology, Kings College Hospital, London, United Kingdom; ⁶Life and Health Sciences Research Institute, University do Minho, Braga, Portugal; ⁷Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal; and ⁸St Jude Children's Research Hospital, Memphis, Tennessee

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

D.A. Bax and A. Mackay contributed equally to this work.

Corresponding Author: Chris Jones, Section of Paediatric Oncology, Institute of Cancer Research, Sutton, Surrey, SM2 5NG, United Kingdom. Phone: 44-0-20-8722-4416; Fax: 44-0-20-8722-4321; E-mail: chris.jones@icr.ac.uk

doi: 10.1158/1078-0432.CCR-10-0438

©2010 American Association for Cancer Research.

Translational Relevance

Pediatric high-grade gliomas represent clinically devastating and biologically understudied tumors of the central nervous system. Little is known about the key genomic alterations that arise in childhood cases, nor of the specific differences between these and the adult disease. We present the copy number profiling of a large series of these rare tumors, and identify numerous low-frequency events previously unreported in pediatric high-grade glioma, including the potential therapeutic target *IGF1R*. Tumors could be classified into distinct genomic subtypes, with marked differences in clinical outcome, and an idealized *PDGFRA*^{amp}, 1q+, 16q- genotype was considerably enriched in pediatric cases, in contrast to the *EGFR*^{amp}, 7+, 10q- cases more commonly associated with adults. We further highlight the importance of platelet-derived growth factor (PDGF) signaling in this context, through the most commonly observed genomic amplification of *PDGFRA*, as well as a unique amplification of *PDGFRB*, providing strong rationale for clinically targeting this pathway in children with this disease.

specific genetic amplification/deletion events common to tumors from patients of all ages, there are certain events found at significantly different frequencies in pediatric versus adult lesions.

One of the most immediately apparent differences was the high frequency of chromosome 1q gains and 16q losses, and the lower frequency of (often concurrent) gain of chromosome 7 and loss of 10q in childhood cases compared with adults. Although there were numerous low-frequency amplifications and deletions such as *MYC/MYCN*, *CCND2*, *KRAS*, and *CDKN2C*, which seemed to show the pediatric high-grade glioma genome to be similar to clinical secondary adult glioblastomas (13, 14), a lack of *IDH1* mutations in the childhood setting showed the distinct biological pathways active during pathogenesis (12).

The most common amplification in the pediatric cases was at 4q12, with shortest region of overlap (SRO) and expression analyses identifying the amplicon driver to be *PDGFRA* (12). This was present in up to 17% of primary pediatric glioblastoma, and 29% of diffuse intrinsic pontine glioma, and was also found in 50% of cases of high-grade glioma arising as a secondary malignancy after cranio-spinal radiation (post-IR). Many cases without *PDGFRA* amplification were still found to show overexpression of a specific *PDGFRA*-associated gene signature, which was itself distinct from that observed in adult cases with the 4q12 amplification. Taken together, platelet-derived growth factor (PDGF)-driven signaling seems to be preferentially activated in the majority of pediatric tumors, in contrast to adults, where epidermal growth factor receptor (*EGFR*) is implicated as the predominant target (12).

Although these studies are beginning to unravel the key features of the pediatric high-grade glioma genome, the total number of cases studied remains considerably smaller than for adult tumors. This is of particular importance given the lower frequency of the majority of genetic aberrations detected in childhood cases. Validating these low-frequency events in independent cohorts as being recurrent abnormalities, as well as the likely identification of novel isolated copy number changes will aid our understanding of the key pathways underlying the diversity of high-grade gliomas in children. To this end we carried out an array CGH study of 63 cases of pediatric high-grade glioma from formalin-fixed, paraffin-embedded (FFPE) archival pathology specimens on a 32K tiling-path bacterial artificial chromosome (BAC) platform.

Materials and Methods

Samples and DNA extraction

High-grade glioma samples from 63 patients (<23 years old) treated at the Royal Marsden Hospital (RMH), Sutton, and the Newcastle Royal Infirmary, United Kingdom, were obtained after approval by local and multicenter ethical review committees. The collection consisted of 37 glioblastoma multiforme, 14 anaplastic astrocytomas, 4 anaplastic oligodendrogliomas, 4 diffuse intrinsic (brain stem) gliomas, 2 astroblastoma, 1 oligoastrocytoma, and 1 gliosarcoma. All cases were archival FFPE tissues. The presence of tumor tissue in these samples and the tumor type were verified on a H&E-stained section independently by two neuropathologists (DWE and SA-S). Nine of the cases were previously profiled from a frozen tumor specimen in the collaborative SNP study (12). DNA was extracted using the DNeasy Tissue Kit (Qiagen) according to the manufacturer's protocol and quantitated on a NanoDrop spectrophotometer (Thermo Scientific).

Array CGH

All raw and processed data have been deposited in Array Express (<http://www.ebi.ac.uk/microarray-as/ae/E-TABM-857>). The array CGH platform used in this study was constructed at the Breakthrough Breast Cancer Research Centre and comprises 31,619 overlapping BAC probes covering the human genome at an approximate resolution of 50 kb (A-MEXP-1734). Hybridizations were carried out as previously described (15) and slides were scanned using an Axon 4000B scanner (Axon Instruments) with images analyzed using Genepix Pro 4.1 software (Axon Instruments). The median localized background slide signal for each clone was subtracted and each clone Cy5/Cy3 ratio was normalized by local regression (loess) against fluorescence intensity and spatial location. Clones overlapping known copy number variants were removed for statistical and visualization purposes, but not for mapping of specific amplifications

and deletions, which was done according to the March 2006 build of the human genome sequence (hg18).

Data analysis

All data transformation and statistical analysis were carried out in R 2.9.0 (<http://www.r-project.org/>) and BioConductor 2.4 (<http://www.bioconductor.org/>), making extensive use of modified versions of the package aCGH in particular (15). For identification of DNA copy number alterations, data were smoothed using a local polynomial adaptive weights procedure for regression problems with additive errors, with thresholds for assigning "gain" and "loss" set at 0.1 ($3 \times$ SD of control hybridizations). For visualization purposes, the processed \log_2 ratios were colored green (gain) or red (loss) after segmentation and copy number determination.

To assess the significance of the genomic alterations, we applied an algorithm similar to those previously described, namely, Genomic Identification of Significant Targets in Cancer (GISTIC; ref. 13) and Genome Topography Scanning (GTS; ref. 16), taking into account the frequency, amplitude, and focality of the observed amplifications (\log_2 ratio >1.0) and deletions (\log_2 ratio <-0.75). This was calculated as the product of the absolute \log_2 ratio, the number of clones in each segment, and the frequency within the entire cohort, scaled to the absolute maximum for amplifications/deletions separately, and overplotted on the frequency histogram for gains and losses described above.

Fluorescent *in situ* hybridization

Fluorescent *in situ* hybridization (FISH) analysis was carried out on FFPE sections as previously described (17). Probes directed against *MYCN* (pool of clones RP11-1183P10, RP11-674F13 and RP11-754G14), *PIK3CA* (RP11-4B14, RP11-642A13, RP11-379M20), *PDGFRA* (RP11-819D11, RP11-58C6), *SKP2* (RP11-749P08, CTD-2010F22), *PDGFRB* (RP11-211F05, RP11-211I20), *MYC* (RP11-440N18, RP11-237F24, CTD-2034C18), *CDK4* (RP11-66N19, RP11-277A02, RP11-672O16), *MDM2* (RP11-611O02, RP13-618A08, CTD-2067J14), and *IGF1R* (CTD-2015I17, RP11-203H14, RP11-189B22) were labeled with Cy3 (GE Healthcare), whereas chromosome-specific control probes at loci of no copy number change were labeled with fluorescein (GE Healthcare). Hybridized preparations were counterstained with 4', 6-diamidino-2-phenylindole in antifade (Vector Laboratories Inc.). Images were captured using a cooled charge-coupled device camera (Photometrics).

Statistics

All statistical tests were done in R2.9.0. Correlations between categorical values were done using the χ^2 and Fisher's exact tests. Correlations between continuous variables were done using Student's *t* test or the Mann-Whitney *U* test. Cumulative survival probabilities were calculated using the Kaplan-Meier method on uniformly treated patients within our cohort from the same institution (RMH),

with differences between survival rates analyzed with the log-rank test. Important prognostic information (including extent of resection, Karnofsky performance score) was not available for all cases in this retrospective study, so multivariate analysis could not be done. All tests were two-tailed, with a confidence interval of 95%. *P* values of <0.05 were considered statistically significant.

Results

Distinct patterns of copy number change in the pediatric high-grade glioma genome

Previously we utilized whole genome amplification strategies for array CGH studies of tumors extracted from FFPE specimens (18). In this study, however, we were able to utilize a cohort of samples for which sufficient material was available to avoid the previous approach. We were able to generate high-quality copy number profiles from an unselected series of 63 pediatric high-grade gliomas using 32K tiling-path BAC arrays from which the tumor cell purity could be verified as $>90\%$ without the need for additional steps.

We observed a mean number of large-scale (whole chromosome or chromosomal arms) gains and losses of 5.8 per sample (median, 4; range, 0-22), with more losses (mean, 3.5; median, 3; range, 0-14) than gains (mean, 2.3; median, 2; range, 0-11). There was a further mean of 1.8 focal amplifications/deletions per sample (median, 1; range, 0-11), again with a slightly increased number of deletions (mean, 1.0; median, 0; range, 0-8) compared with amplifications (mean, 0.8; median, 0; range, 0-4). The list of observed alterations is given for the full dataset in Supplementary Table S1.

We were able to subtype the samples into four groups based upon the pattern of their genomic profiles. First was a group of tumors that had a very stable genome, with few (<3), low-level, focal changes. This subtype comprised 13 of 63 (20.6%) cases, and included 8 tumors (12.7%) that harbored no detectable copy number alterations on our 32K BAC platform (Fig. 1A). The second type contained only large, single copy alterations involving whole chromosomes or chromosomal arms, resulting in aneuploidy in the absence of any high-level amplifications in 22 of 63 (34.9%) cases, the largest subgroup we observed (Fig. 1B). The third type harbored numerous, low-level, intrachromosomal breaks resulting in multiple gains and losses and a highly rearranged genome. This group was also defined for this purpose by exclusion of cases with bona fide amplicons, and comprised 11 of 63 (17.5%) of the cohort (Fig. 1C). Finally, we considered those tumors with single or multiple high-level (\log_2 ratio >1.0) amplifications, regardless of the genomic background, as belonging to the fourth, "amplifier" subtype. This group consisted of 17 of 63 (27.0%) of cases (Fig. 1D).

There were no significant correlations between genomic subtype and WHO grade or histology ($P > 0.05$, Fisher's exact test), with glioblastomas, anaplastic astrocytomas,

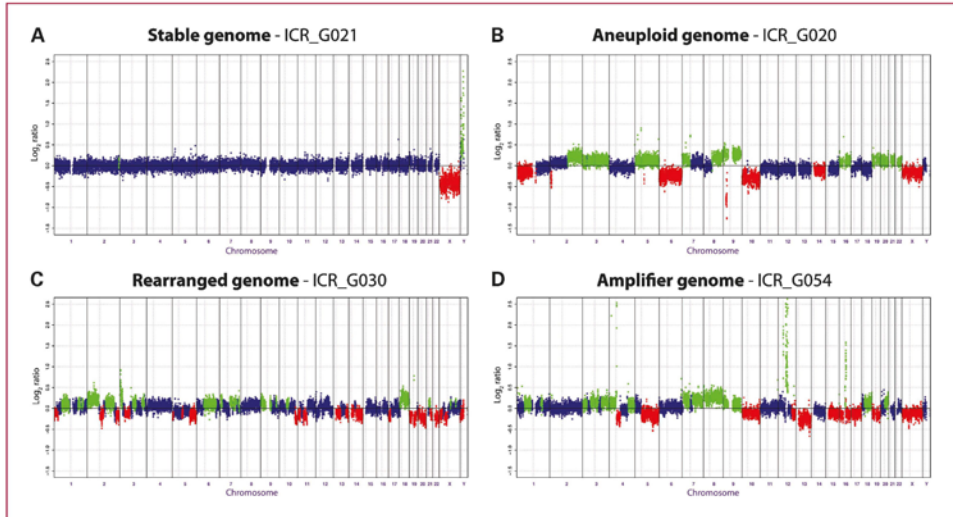
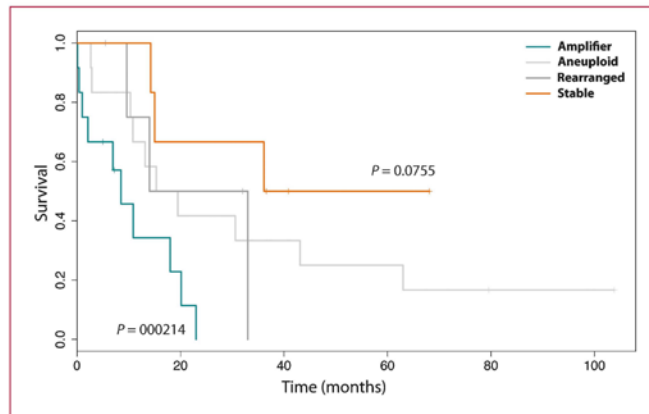


Fig. 1. Pediatric high-grade gliomas comprise different subtypes of copy number profiles. Sample genome plots are given for stable (A), aneuploid (B), rearranged (C), and amplifier (D) genomes within our sample cohort. Log₂ ratios for each clone (Y-axis) are plotted according to chromosomal location (X-axis). Vertical lines, centromeres; green points, gains; red points, losses.

and anaplastic oligodendrogliomas spread across all subtypes. Of note, there were no “stable” genomic cases among the series of five patients that were treated for a previous malignancy by cranio-spinal radiation (post-IR; Supplementary Table S1). There was also no association of copy number profiles with age at diagnosis ($P > 0.05$, Mann-Whitney U test), although the amplifier group did not include any infant tumors (<3 years). However, when

we investigated the overall survival of the patients treated at a single institution (RMH), we detected significant differences by retrospective univariate analysis in the clinical outcome of cases according to the genomic profile of the tumor. The stable genome cases showed a trend towards better prognosis when compared with all other cases ($P = 0.0755$, log-rank test), whereas the samples with an amplifier genome had a significantly shorter time to death

Fig. 2. Genomic subtypes of pediatric high-grade glioma have prognostic relevance. Kaplan-Meier plot for overall survival of pediatric high-grade gliomas treated at a single institution stratified according to genomic subtype. The stable genome cases showed a trend towards better prognosis when compared with all other cases ($P = 0.0755$, log-rank test), whereas the samples with an amplifier genome had a significantly shorter time to death ($P = 0.00214$, log-rank test).



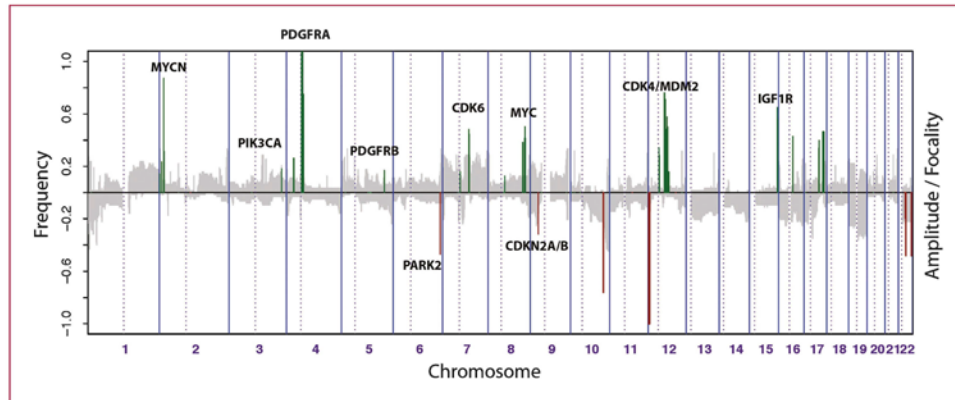


Fig. 3. Summary and significance of genomic aberrations in pediatric high-grade glioma. The proportion of tumors in which each clone is gained or lost is plotted in grey (Y-axis) for each BAC clone according to genomic location (X-axis). A measure of the frequency, amplitude, and focality of high-level events was calculated for each affected clone and was overplotted for amplifications (green) and deletions (red), scaled to the absolute maximum for each.

($P = 0.00214$, log-rank test; Fig. 2). The aneuploid and rearranged cases fell in between, and were representative of the survival characteristics of the cohort as a whole, suggesting that they may need to be considered together as falling between the extremes of the other two groups.

One of the defining features of pediatric high-grade glioma is the frequent gain of chromosome 1q (12 of 63, or 19.0%, versus 17 of 189, or 9.0%, of adult cases; ref. 1; $P = 0.039$, Fisher's exact test) and loss of 16q (11 of 63, or 17.5%, versus 14 of 189, or 7.4%; $P = 0.028$, Fisher's exact test); in contrast to adult glioblastoma cases, in which gains of chromosome 7 (12 of 63, or 19.0%, versus 140 of 189, or 74.1%; $P < 0.0001$, Fisher's exact test) and losses of 10q (10 of 63, or 15.9%, versus 152 of 189, or 80.4%; $P < 0.0001$, Fisher's exact test) predominate. In our FFPE cohort, we noticed a clear distinction of 1q gain cases from those with concurrent 7 gain/10q loss (7+/10q-, 8 of 63, or 12.7%), with only a single case harboring both abnormalities. Neither event was significantly associated with any clinicopathologic parameters, although there was a trend towards shorter survival in the 1q+ cases ($P = 0.0865$, log-rank test). Neither abnormality was seen in any infant cases.

Mapping of focal amplifications and deletions to known oncogenes and novel loci

As we had with large-scale alterations, we observed numerous focal amplifications and deletions. In summary, we identified 47 unique amplification and 32 unique deletions. All these events are detailed in full in Supplementary Table S2 (amplifications) and Supplementary Table S3 (deletions).

The most common amplicon was at 4q12 (10 of 63, or 15.9%), and deletion at 9p21 (10 of 63, or 15.9%, consisting of 8 homozygous, 2 hemizygous). Mapping the SRO in these cases narrowed these regions specifically to

PDGFRA and *CDKN2A*, respectively, confirming the initial observations that these are by far the most common amplifications/deletions in pediatric high-grade glioma (12). Other common events included amplification of *MYCN* at 2p24 (3 of 63, or 4.7%) or *MYC* at 8q24 (2 of 63, or 3.2%), together giving a frequency of 7.9% (5 of 63) of cases with genomic *MYC* family dysregulation; and 3 of 63 (4.7%) *EGFR* amplification at 7p12 – a lower frequency than observed in our recent chromogenic *in situ* hybridization study of a larger cohort of which this series is a subset, reflecting the focal nature of the amplification event in a small number of tumors identified by molecular pathology (19).

For the remaining aberrations, we highlighted the SROs where they were found to be recurrent. However, as most were present only in a single case, and we were unable to narrow down gained/lost regions, the result was that we identified a total of 1,026 amplified and 1,243 deleted genes across our series. To facilitate the identification of key oncogenic events in pediatric high-grade glioma, we sought to assign significance to the genomic aberrations we observed. Inspired by algorithms such as GISTIC (13) and GTS (16), we developed a simple measure based upon three key features of our data for each clone on the array: (a) frequency of high-level amplification/homozygous deletion, (b) absolute magnitude of the change, and (c) focality of the segmented copy number change. This amplitude/focality measure was then scaled to the maximum and minimum for amplifications/deletions, respectively, and plotted over the frequency of low-level gains and losses on the same histogram (Fig. 3).

As well as *PDGFRA* (the highest scoring gene) and *CDKN2A*, this analysis highlighted the importance of several known oncogenes, amplified at low frequency in our series, but at high magnitude, and in a focally restricted

manner. These included *PIK3CA* (3q26), *CDK6* (7q21), and *CDK4* (12q14), the first two previously reported in adult glioblastoma, but not in pediatric cases, and present here in a single case. We also identified amplifications of two additional receptor tyrosine kinases: *IGF1R* at 15q26 (Fig. 4A) and *PDGFRB* at 5q33 (Fig. 4B). Such an approach further highlighted the potential significance of known deletions targeting *PARK2* at 6q6 and *MGMT*, *PITPRE*, and others at 10q26, as well as unique events for which the candidate gene is unknown at 10q25 (Fig. 4C) and 11q14 (Fig. 4D).

We were able to validate nine of these lower-frequency amplification events by carrying out FISH on our FFPE sections using specific probes against *MYCN*, *PIK3CA*, *PDGFRB*, *SKP2*, *PDGFRB*, *MYC*, *CDK4*, *MDM2*, and *IGF1R* (Fig. 5).

Glioblastoma core signaling pathways are not commonly activated by copy number changes in pediatric patients

One of the most important findings from recent large-scale genomic profiling studies of adult glioblastoma was the identification of three core signaling pathways that were abrogated by amplification, deletion, and/or mutations of key genes in the vast majority of cases. Considering only the copy number data from these studies, 59%, 70%, and 66% of cases were found to have at least one genetic event targeting the receptor tyrosine kinase/phosphoinositide 3-kinase (RTK/PI3K), p53, or RB pathways, respectively (1, 2).

We mapped the copy number changes in our pediatric cases to the same pathways, which included many of the genes described above, as well as others described in adult

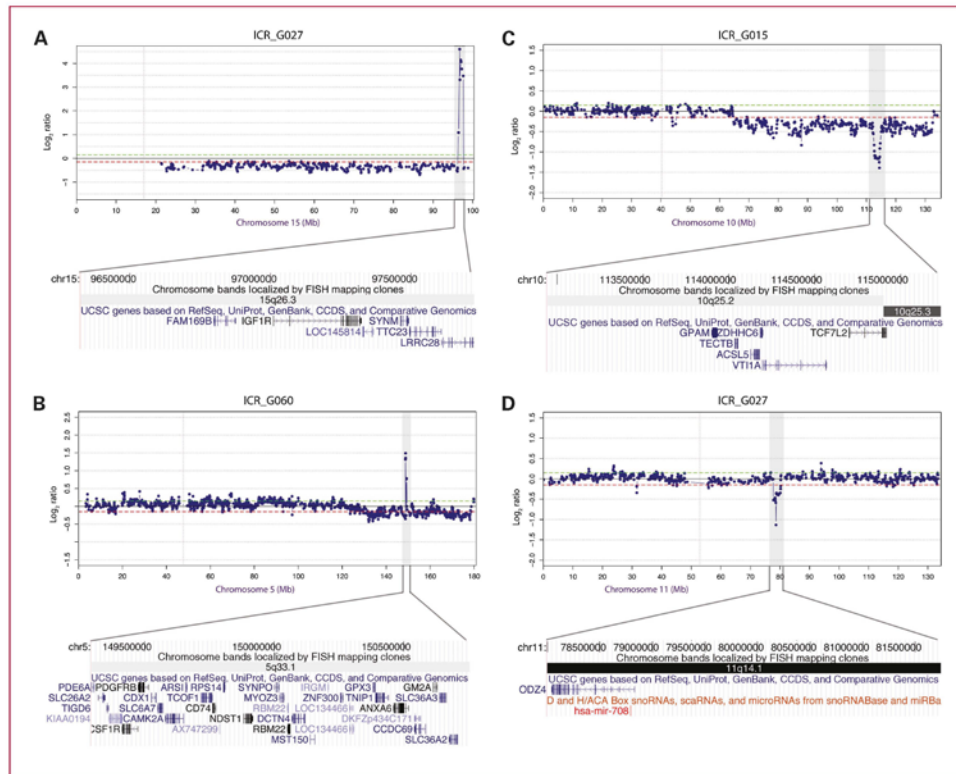


Fig. 4. Novel low-frequency amplifications and fine-mapping focal deletions in pediatric high-grade glioma. Chromosome plots for chromosome 15, targeting *IGF1R* (A); chromosome 5, targeting *PDGFRB/CSF1R* (B); chromosome 10, mapping a deletion at 10q25.2-q25.3 (C); and chromosome 11, resolving a deletion at 11q14 to *ODZ4* and *hsa-mir-708* (D). Log₂ ratios for each clone are plotted (Y-axis) for each BAC clone according to location (X-axis) along the length of the chromosome, with genes and microRNA within the minimal regions plotted underneath according to positional information from the UCSC Genome Browser (hg18).

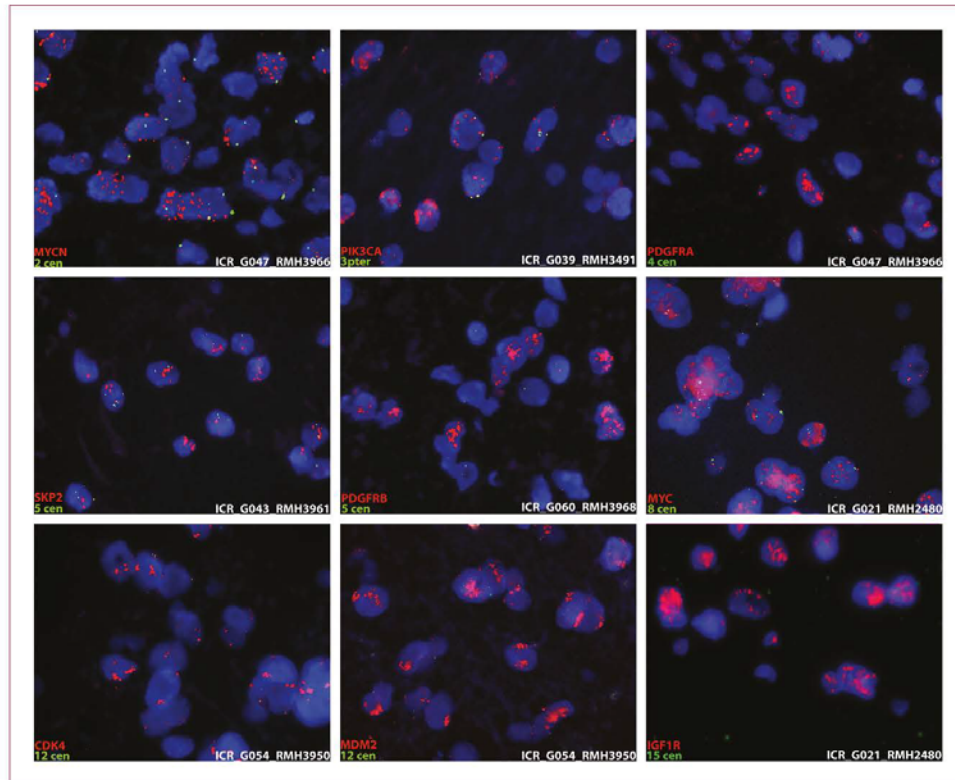


Fig. 5. FISH validation of low-frequency amplifications in pediatric high-grade glioma. Specific probes for *MYCN*, *PIK3CA*, *PDGFRA*, *SKP2*, *PDGFRB*, *MYC*, *CDK4*, *MDM2*, and *IGF1R* were labeled with Cy3 (red) and cohybridized to interphase nuclei on FFPE specimens with chromosome-specific control probes labeled with fluorescein.

glioblastoma, including *MET*, *KRAS*, and *AKT2* (RTK/PI3K), *MDM2* (p53), and *CCND2* (RB). Despite this, we observed a significantly lower frequency of pathway dysregulation compared with that reported in adults: 16 of 63 (25%) RTK/PI3K, 12 of 63 (19%) p53, and 14 of 63 (22%) RB (all $P < 0.0001$, Fisher's exact test; Fig. 6). Even after removing the stable genome subtype from this analysis, it is apparent that pediatric tumors show targeting of these core pathways by copy number alterations in less than half as many instances than in adults.

To explore whether other canonical pathways may be activated by this mechanism preferentially in childhood tumors, we mapped amplified/deleted genes in those tumors without core pathway targeting via GenMAPP. Although there were isolated cases with clear genomic events linked to activation of the Sonic Hedgehog (*GLI2* amplification, *HHIP* deletion) and Notch (*DLL3* amplification, *DLK1* deletion)

pathway activation, there was no consistently targeted pathway in these cases, nor was there specific enrichment of any additional pathway across the entire cohort.

Discussion

We were previously part of a collaborative study setting out to comprehensively map the copy number alterations present in the pediatric high-grade glioma genome, in which we used Affymetrix 500K SNP arrays on a series of 78 cases available as frozen tumor samples (12). Those data revealed an overlapping, but distinct, underlying molecular genetics of the childhood disease when compared with recent large-scale genomic analyses of adult high-grade glioma (1, 2). Along with the common amplification/deletion targets of *PDGFRA* and *CDKN2A/B*, there were numerous low-frequency events targeting both well-recognized

oncogenes and novel loci. The present study had three purposes: (a) to validate the high-frequency events in an independent set of samples, analyzed on an independent microarray platform; (b) to extend the sample set to provide evidence of recurrence of the low-frequency events previously reported; and (c) to identify novel low-frequency events, which by their nature may have been missed in the earlier study.

The most frequent focal events were *PDGFRA* amplification and *CDKN2A/B* deletion, and the most common large-scale gains and losses included chromosomes 1q and 16q, respectively. The *PDGFRA*^{amp}, 1q+, 16q- events were signifi-

cantly more common in the childhood setting (10, 11, 20), although it is important to note that they are present in a proportion of adult tumors. Similarly, we observed a group of tumors in our cohort containing aberrations more commonly associated with the adult disease, namely *EGFR*^{amp}, 7+, 10q-, albeit at significantly reduced frequencies. That they tended towards exclusivity suggests they represent archetypes for different ends of the spectrum of the disease.

One of the most intriguing differences observed in the pediatric setting was the presence of a proportion of cases of high-grade tumors with very few, or even no detectable copy number alterations. This was true on both BAC

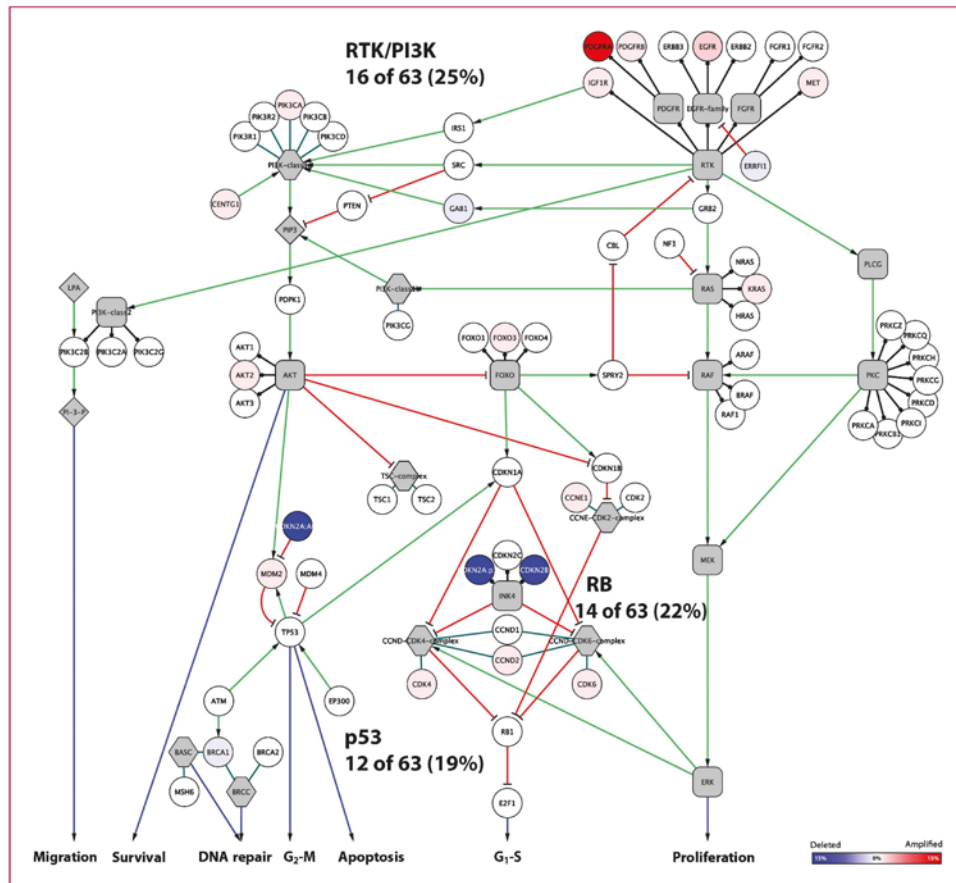


Fig. 6. Glioblastoma core signaling pathways are dysregulated by copy number changes less frequently in pediatric than in adult tumors. Signaling pathway heatmap of interactions defined by the cancer genome atlas (1). Red, genes with amplifications; blue, genes with focal deletion. The overall frequency of copy number alteration in pediatric high-grade glioma for each pathway is listed, and is significantly lower than in adults (25% RTK/PI3K, 19% p53, and 22% RB versus 59%, 70%, and 66% for adult glioblastoma, respectively).

(approximately 32,000 probes, 100 kb resolution) and SNP (approximately 500,000 probes, 6 kb resolution) platforms (12), and is in direct contrast to data from adult tumors (1, 2). This stable genomic profile is independent of histologic grade or type, and seems to convey an improved survival in patients with high-grade glioma, in contrast to those patients with an amplifier genomic profile, who do significantly worse.

Another of the defining features of the pediatric high-grade glioma genome is the numerous low-frequency amplifications and deletions present only in isolated cases in any given study. By nearly doubling the number of these rare tumors for which we have genomic data, we have been able to ascertain aberrations as recurrent across 132 cases. These include amplifications of known oncogenes within the core signaling pathways described in adult glioblastoma, such as *CDK6* (10), *MET*, and *CCND2*, as well as novel targets. These include *ID2* at 2p25, previously found in association with the *MYCN* amplicon at 2p24, possibly part of a single event, identified here as an independent target in its own right. *ID2* is a helix-loop-helix transcription factor that has previously been shown to be widely expressed in astrocytic tumors (21, 22), and may play a role in negatively regulating cell differentiation and promoting cell survival (23, 24). Another amplicon at 17q22 was also confirmed in the FFPE series, with a SRO analysis identifying *RNF43* as the most likely target. *RNF43* is a ubiquitin ligase that promotes cell growth and is upregulated in colon cancer (25, 26), but has not previously been implicated in gliomagenesis.

Homozygous deletions now apparent as recurrent lesions include those at 14q32, encompassing a large number of microRNAs, as well as the gene *DLK1*. *DLK1* is a δ -like homolog that acts to inhibit Notch signaling through specific binding interactions with the receptor (27), and may play diverse roles in cellular transformation and differentiation (28). Although we have now observed two cases of homozygous deletion, other mechanisms of downregulation may be active, as *DLK1* is present at an imprinted locus, with increased methylation upstream of *GTI2* leading to reduced expression in other tumor types (29). Other deletions may have a more complicated role in gliomagenesis such as those on chromosome 16q. The SNP study identified a large deletion in a single tumor that is present as two separate events at 16q12 and 16q21 in two independent cases here, targeting numerous candidates including clusters of Iroquois homeobox genes, metallothioneins, and coiled-coil domain containing genes. By contrast, a homozygous deletion observed in the present study overlaps two independent loci previously reported at 11q14 to target a single microRNA, *hsa-mir-708*, and a single gene, *ODZ4*. Although little is known about *mir-708*, the odd *Oz/ten-m* homolog 4 is expressed in the developing and adult central nervous system, and seems to act as an important transcriptional regulator associated with neurodevelopment (30, 31).

Finally, we were also able to identify several novel amplifications and deletions, the significance of many of which is not yet clear. There were some genes identified that were

also present in adult glioblastoma studies which had not previously been reported in pediatric high-grade glioma, such as *AKT2*, *CCNE1*, *GLI2*, *MDM2*, *PARK2*, and *PIK3CA*. There were other previously unreported genes that may be associated with specific glioblastoma-related signaling pathways such as *AKTIP* (16q12), an Akt-interacting protein that acts as an activator of the PI3K pathway (32), and *PIK3C3* (18q12), also known as *Vps34*, a member of the PI3K family associated with autophagy (33). There were numerous others with potential functional relevance unknown.

We also noted rare amplifications at receptor tyrosine kinases considered less likely to be driven by copy number gain. Firstly was a very high level gain of *IGF1R* at 15q26 (11). Insulin-like growth factor (IGF) signaling has previously been implicated in gliomagenesis, primarily on the basis of high levels of the ligand *IGF2* in glioblastoma specimens (34). The growth-promoting effects of *IGF2* that were shown were mediated via *IGF1R* and the PI3K regulatory subunit *PIK3R3*. Of particular relevance to the childhood setting was the observation of a mutual exclusivity between *IGF2*-associated tumors and *EGFR*-driven cases, suggesting that the IGF pathway may play a prominent role in pediatric tumors, possibly in concert with PDGF receptor (PDGFR)-related signaling.

Secondly was an amplicon at 5q33 which included *PDGFRB* (with another receptor tyrosine kinase *CSF1R*). Given the clear importance of PDGFR signaling on pediatric high-grade gliomas, it is perhaps unsurprising that there may be multiple mechanisms active in driving tumorigenesis through a common pathway. To this end, we also observed recurrent amplification of the ligand *PDGFB* (22q13) in the previous SNP study (12), and here further observed focal copy number gain at 7p22 encompassing *PDGFA*. That these unique genomic events have thus far been found to be restricted to pediatric tumors adds further evidence to a distinct underlying genetics driving archetypal high-grade gliomas in children, one that is largely PDGF-driven, and forms a discrete pole within the diversity of glioma biology. Understanding the most appropriate ways of efficaciously targeting these pathways in the most appropriate patient populations will hopefully overcome the disappointing early-phase clinical trials observed thus far with PDGFR inhibitors.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

We acknowledge National Health Service funding to the NIHR Biomedical Research Centre. This work was supported by The Royal Marsden Children's Department Fund, Fundação para a Ciência e Tecnologia, Portugal, and Breakthrough Breast Cancer.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 02/18/2010; revised 03/31/2010; accepted 04/26/2010; published OnlineFirst 06/22/2010.

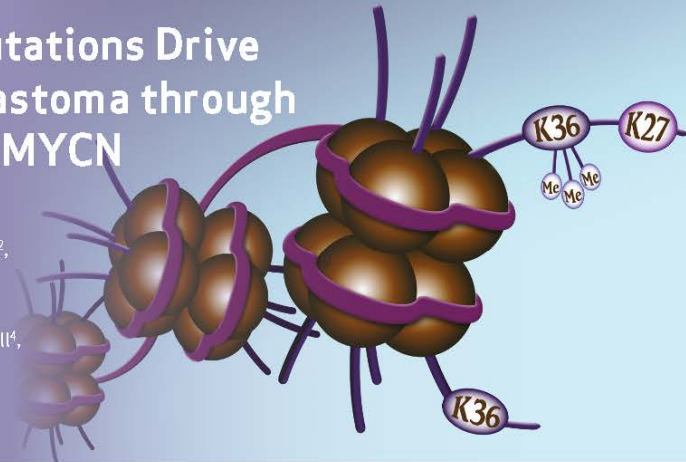
References

- McLendon R, Friedman A, Bigner D, et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 2008;455:1061–8.
- Parsons DW, Jones S, Zhang X, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science* 2008;321:1807–12.
- Rao SK, Edwards J, Joshi AD, Siu IM, Riggins GJ. A survey of glioblastoma genomic amplifications and deletions. *J Neurooncol* 2010; 96:169–79.
- Gardina PJ, Lo KC, Lee W, Cowell JK, Turpaz Y. Ploidy status and copy number aberrations in primary glioblastomas defined by integrated analysis of allelic ratios, signal ratios and loss of heterozygosity using 500K SNP Mapping Arrays. *BMC Genomics* 2008;9:489.
- de Tayrac M, Etcheverry A, Aubry M, et al. Integrative genome-wide analysis reveals a robust genomic glioblastoma signature associated with copy number driving changes in gene expression. *Genes Chromosomes Cancer* 2009;48:55–68.
- Ruano Y, Molloje M, Ribalta T, et al. Identification of novel candidate target genes in amplicons of Glioblastoma multiforme tumors detected by expression and CGH microarray profiling. *Mol Cancer* 2006;5:39.
- Korshunov A, Sycheva R, Golanov A. Genetically distinct and clinically relevant subtypes of glioblastoma defined by array-based comparative genomic hybridization (array-CGH). *Acta Neuropathol* 2006; 111:465–74.
- Rickert CH, Strater R, Kaatsch P, et al. Pediatric high-grade astrocytomas show chromosomal imbalances distinct from adult cases. *Am J Pathol* 2001;158:1525–32.
- Wong KK, Tsang YT, Chang YM, et al. Genome-wide allelic imbalance analysis of pediatric gliomas by single nucleotide polymorphic allele array. *Cancer Res* 2006;66:11172–8.
- Qu HQ, Jacob K, Fatet S, et al. Genome-wide profiling using single-nucleotide polymorphism arrays identifies novel chromosomal imbalances in pediatric glioblastomas. *Neuro-oncol* 2010;12:153–63.
- Schiffman JD, Hodgson JG, VandenBerg SR, et al. Oncogenic BRAF mutation with CDKN2A inactivation is characteristic of a subset of pediatric malignant astrocytomas. *Cancer Res* 2010;70:512–9.
- Paugh BS, Qu C, Jones C, et al. Integrated molecular profiling of pediatric high grade gliomas reveals key differences with the adult disease. *J Clin Oncol* 2010 May 17.
- Beroukhi R, Getz G, Nghiemphu L, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* 2007;104:20007–12.
- Maher EA, Brennan C, Wen PY, et al. Marked genomic differences characterize primary and secondary glioblastoma subtypes and identify two distinct molecular and clinical secondary glioblastoma entities. *Cancer Res* 2006;66:11502–13.
- Natrajan R, Williams RD, Hing SN, et al. Array CGH profiling of favourable histology Wilms tumours reveals novel gains and losses associated with relapse. *J Pathol* 2006;210:49–58.
- Wiedemeyer R, Brennan C, Heffernan TP, et al. Feedback circuit among INK4 tumor suppressors constrains human glioblastoma development. *Cancer Cell* 2008;13:355–64.
- Lambros MB, Simpson PT, Jones C, et al. Unlocking pathology archives for molecular genetic studies: a reliable method to generate probes for chromogenic and fluorescent *in situ* hybridization. *Lab Invest* 2006;86:390–406.
- Vuononvirta R, Sebire NJ, Dallosso AR, et al. Perilobar nephrogenic rests are nonobligate molecular genetic precursor lesions of insulin-like growth factor-II-associated Wilms tumors. *Clin Cancer Res* 2008; 14:7635–44.
- Bax DA, Gaspar N, Little SE, et al. EGFRvIII deletion mutations in pediatric high-grade glioma and response to targeted therapy in pediatric glioma cell lines. *Clin Cancer Res* 2009;15:5753–61.
- Zarghooni M, Bartels U, Lee E, et al. Whole-genome profiling of pediatric diffuse intrinsic pontine gliomas highlights platelet-derived growth factor receptor (α) and poly (ADP-ribose) polymerase as potential therapeutic targets. *J Clin Oncol* 2010;28:1337–44.
- Mikami S, Hirose Y, Yoshida K, et al. Predominant expression of OLIG2 over ID2 in oligodendroglial tumors. *Virchows Arch* 2007; 450:575–84.
- Vandeputte DA, Troost D, Leenstra S, et al. Expression and distribution of id helix-loop-helix proteins in human astrocytic tumors. *Glia* 2002;38:329–38.
- Gray MJ, Dallas NA, Van Buren G, et al. Therapeutic targeting of Id2 reduces growth of human colorectal carcinoma in the murine liver. *Oncogene* 2008;27:192–200.
- Obayashi S, Tabunoki H, Kim SU, Satoh J. Gene expression profiling of human neural progenitor cells following the serum-induced astrocyte differentiation. *Cell Mol Neurobiol* 2009;29:423–38.
- Sugiura T, Yamaguchi A, Miyamoto K. A cancer-associated RING finger protein, RNF43, is a ubiquitin ligase that interacts with a nuclear protein, HAP95. *Exp Cell Res* 2008;314:1519–28.
- Yagyu R, Furukawa Y, Lin YM, Shimokawa T, Yamamura T, Nakamura Y. A novel oncoprotein RNF43 functions in an autocrine manner in colorectal cancer. *Int J Oncol* 2004;25:1343–8.
- Nueda ML, Baladron V, Sanchez-Solana B, Ballesteros MA, Laborda J. The EGF-like protein Dlk1 inhibits notch signaling and potentiates adipogenesis of mesenchymal cells. *J Mol Biol* 2007;367: 1281–93.
- Espina AG, Mendez-Vidal C, Moreno-Mateos MA, et al. Induction of Dlk1 by PTTG1 inhibits adipocyte differentiation and correlates with malignant transformation. *Mol Biol Cell* 2009;20:3353–62.
- Kawakami T, Chano T, Minami K, Okabe H, Okada Y, Okamoto K. Imprinted DLK1 is a putative tumor suppressor gene and inactivated by epimutation at the region upstream of GTL2 in human renal cell carcinoma. *Hum Mol Genet* 2006;15:821–30.
- Tucker RP, Chiquet-Ehrismann R. Teneurin: a conserved family of transmembrane proteins involved in intercellular signaling during development. *Dev Biol* 2006;290:237–45.
- Zhou XH, Brandau O, Feng K, et al. The murine Ten-m/Odz genes show distinct but overlapping expression patterns during development and in adult brain. *Gene Expr Patterns* 2003;3:397–405.
- Remy I, Michnick SW. Regulation of apoptosis by the Ft1 protein, a new modulator of protein kinase B/Akt. *Mol Cell Biol* 2004;24: 1493–504.
- Simonsen A, Tooze SA. Coordination of membrane events during autophagy by multiple class III PI3-kinase complexes. *J Cell Biol* 2009;186:773–82.
- Soroceanu L, Kharbanda S, Chen R, et al. Identification of IGF2 signaling through phosphoinositide-3-kinase regulatory subunit 3 as a growth-promoting axis in glioblastoma. *Proc Natl Acad Sci U S A* 2007;104:3466–71.

RESEARCH BRIEF

Histone H3.3 Mutations Drive Pediatric Glioblastoma through Upregulation of MYCN

Lynn Bjerke^{1,2}, Alan Mackay^{1,2},
Meera Nandhabalan^{1,2}, Anna Burford^{1,2},
Alexa Jury^{1,2}, Sergey Popov^{1,2},
Dorine A. Bax^{1,2}, Diana Carvalho^{1,2,6,7},
Kathryn R. Taylor^{1,2}, Maria Vinci^{1,2},
Ilijana Bajrami^{1,3}, Imelda M. McGonnell⁴,
Christopher J. Lord^{1,2}, Rui M. Reis^{7,8},
Darren Hargrave⁵, Alan Ashworth^{1,3},
Paul Workman², and Chris Jones^{1,2}



ABSTRACT

Children and young adults with glioblastoma (GBM) have a median survival rate of only 12 to 15 months, and these GBMs are clinically and biologically distinct from histologically similar cancers in older adults. They are defined by highly specific mutations in the gene encoding the histone H3.3 variant *H3F3A*, occurring either at or close to key residues marked by methylation for regulation of transcription—K27 and G34. Here, we show that the cerebral hemisphere-specific G34 mutation drives a distinct expression signature through differential genomic binding of the K36 trimethylation mark (H3K36me₃). The transcriptional program induced recapitulates that of the developing forebrain, and involves numerous markers of stem-cell maintenance, cell-fate decisions, and self-renewal. Critically, *H3F3A* G34 mutations cause profound upregulation of *MYCN*, a potent oncogene that is causative of GBMs when expressed in the correct developmental context. This driving aberration is selectively targetable in this patient population through inhibiting kinases responsible for stabilization of the protein.

SIGNIFICANCE: We provide the mechanistic explanation for how the first histone gene mutation in human disease biology acts to deliver *MYCN*, a potent tumorigenic initiator, into a stem-cell compartment of the developing forebrain, selectively giving rise to incurable cerebral hemispheric GBM. Using synthetic lethal approaches to these mutant tumor cells provides a rational way to develop novel and highly selective treatment strategies. *Cancer Discov*; 3(5): 512–19. © 2013 AACR.

See related commentary by Huang and Weiss, p. 484.

INTRODUCTION

The clinical and molecular differences observed in glioblastoma (GBM) of children and young adults compared with the

more common, histologically similar lesions in older adults is strongly suggestive of a distinct underlying biology (1). The identification of unique and highly specific mutations in the gene *H3F3A*, encoding the variant histone H3.3A in GBM of

Authors' Affiliations: ¹Divisions of Molecular Pathology and ²Cancer Therapeutics, and ³Breakthrough Breast Cancer Research Centre and CRUK Gene Function Laboratory, Division of Breast Cancer Research, The Institute of Cancer Research, Sutton, Surrey, SM2 5NG, United Kingdom; ⁴Royal Veterinary College; ⁵Great Ormond Street Hospital, London, United Kingdom; ⁶University of Coimbra, Coimbra; ⁷ICVS, University of Minho, Braga, Portugal; and ⁸Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos SP, Brazil

Note: Supplementary data for this article are available at Cancer Discovery Online (<http://cancerdiscovery.aacrjournals.org/>).

Corresponding Author: Chris Jones, Glioma Team, Divisions of Molecular Pathology and Cancer Therapeutics, The Institute of Cancer Research, Sutton, Surrey, SM2 5NG, United Kingdom. Phone: 44-020-8722-4416; Fax: 44-020-8722-4321; E-mail: chris.jones@icr.ac.uk

doi: 10.1158/2159-8290.CD-12-0426

© 2013 American Association for Cancer Research.

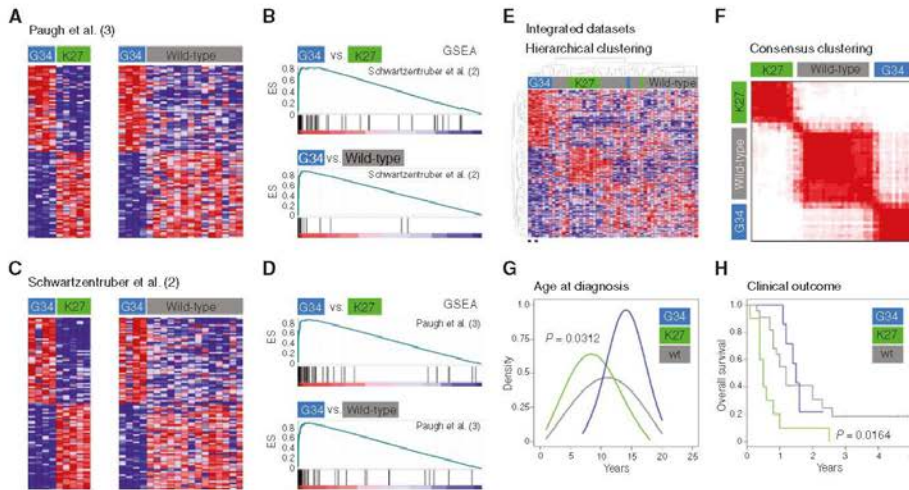


Figure 1. Distinct molecular and clinical correlates of *H3F3A* mutation subgroups. **A**, heatmap representing differential gene expression signatures between G34 versus K27, and G34 versus wild-type, pediatric GBM specimens identified by Paugh and colleagues (3). Top 100 differentially expressed genes are shown for each comparison. **B**, gene set enrichment analysis (GSEA) for differential gene expression signatures identified by Schwartzentruber and colleagues (2) versus those from Paugh and colleagues (3). Top, G34 versus K27: enrichment score (ES) = 0.833, P [family-wise error rate (FWER)] = 0.0, q [false discovery rate (FDR)] = 0.0. Bottom, G34 versus wild-type: ES = 0.94, FWER P = 0.0, FDR q = 0.0. **C**, heatmap representing differential gene expression signatures between G34 versus K27, and G34 versus wild-type, pediatric GBM specimens from (2). Top 100 differentially expressed genes are shown for each comparison. **D**, GSEA for differential gene expression signatures identified in (3) versus those in (2). Top, G34 versus K27: ES = 0.98, FWER P = 0.03, FDR q = 0.04. Bottom, G34 versus wild-type: ES = 0.90, FWER P = 0.0, FDR q = 0.0. **E**, heatmap representing differential gene expression signatures, highlighting specific clusters of G34- and K27-mutant tumors, distinct from a more heterogeneous group of wild-type cases. G34V tumors are represented by asterisks. **F**, K-means consensus clustering finds the most stable number of subgroups to be 3, marked by *H3F3A* mutation status. **G**, K27- and G34V mutant pediatric GBM in our integrated dataset have distinct age incidence profiles, with K27 tumors peaking at 7 years in contrast to G34 at age 14. The 2 G34V tumors were diagnosed at age 14 and 20. **H**, Kaplan-Meier plot for overall survival of pediatric patients with GBM stratified by *H3F3A* status. K27-mutant tumors have significantly shorter survival than G34 (P = 0.0164, log-rank test). A single G34V case for which data were available had an overall survival of 1.4 years. wt, wild-type.

children and young adults has recently provided definitive proof of this hypothesis (2). However, a mechanism was lacking for how mutations at or close to key residues associated with posttranslational modification of the histone tail led to tumorigenesis.

We have sought to address this by examining how the differences in clinical presentation, anatomic location, and gene expression associated with the different *H3F3A* mutations are manifested. By exploiting the only known G34-mutant model system, we show that differential binding of the H3K36 trimethyl mark underpins these processes and identify MYCN as the oncogenic driver during forebrain development, providing a novel avenue for targeted therapy in children with these tumors.

RESULTS

Initial evidence suggested a distinct gene expression signature associated with mutations at the K27 (lysine to methionine, K27M) versus G34 (glycine to either arginine, G34R, or valine, G34V) residues (2). We validated these data by identifying differential expression patterns for mutations with G34 versus K27 mutations in 2 independent datasets

for which mutation data were either publicly available or were ascertained in our laboratory (refs. 2, 3; Fig. 1). In both instances, highly significant differential gene expression was noted between G34-mutant tumors and K27 or wild-type cases (Fig. 1A and C), which was consistent across the datasets as assessed by gene set enrichment analysis (GSEA; Figs. 1B and D) with enrichment scores (ES) of 0.833 to 0.943 and P (family-wise error rate; FWER) and q (false discovery rate; FDR) values of 0.0 to 0.04. Given the considerable overlap in gene expression signatures between studies, we subsequently utilized an integrated dataset (Supplementary Table S1), where hierarchical clustering resolved G34- and K27-mutant tumors from a more heterogeneous wild-type subgroup (Fig. 1E), confirmed by k-means consensus clustering (Fig. 1F). These subgroups also showed important clinical differences, as previously described (2), with K27-mutant tumors arising in younger children (peak age 7 years; P = 0.0312, t test; Fig. 1G) and having a worse clinical outcome (P = 0.0164, log-rank test; Fig. 1H) compared with G34 tumors (peak age 14 years) and *H3F3A* wild-type tumors. There were no significant transcriptional or clinicopathologic differences between G34R and G34V tumors, although a lack of samples of the latter (n = 2) precludes robust analyses.

To understand the functional significance of *H3F3A* mutations in cerebral hemispheric tumors, we turned to a well-characterized (4) model of pediatric GBM, the KNS42 cell line, which was derived from a 16-year-old patient and harbors the G34V mutation (Fig. 2A). In contrast to the reported data in a single pediatric GBM sample with G34R (2), KNS42 cells did not show increased levels of total histone H3K36 trimethylation compared with a panel of *H3F3A* wild-type pediatric glioma cells (Fig. 2B, Supplementary Fig. S1). KNS42 cells harbor a nonsynonymous coding change of *ATRX* (Q891E) that appears in the single-nucleotide polymorphism databases (rs3088074), and Western blot analysis shows no diminution of protein levels. As *ATRX* is a known chaperone of histone H3.3 to the telomeres, a wild-type protein would not be expected to convey the alternative lengthening of telomeres (ALT) phenotype, as observed (Supplementary Fig. S2); however, this ought not play a significant role in gene transcription as deposition of H3.3 in euchromatin is carried out by alternative chaperones such as HIRA.

We conducted chromatin immunoprecipitation linked to next-generation whole genome sequencing (ChIP-Seq) for H3K36me3 to test the hypothesis that, rather than total H3K36me3, the G34V mutation may instead result in differential binding of the trimethyl mark throughout the genome. Compared with *H3F3A* wild-type SF188 pediatric GBM cells, H3K36me3 was found to be significantly differentially bound in KNS42 cells at 5,130 distinct regions of the genome corresponding to 156 genes (DESeq $P < 0.05$, overall fold change > 2 , contiguous median coverage > 2 ; Supplementary Table S2). These observations were not due to differential gene amplification, as concurrent whole genome DNA sequencing showed that these bound genes were not found in regions on cell line-specific copy number alterations (Fig. 2C; Supplementary Fig. S3 and Supplementary Table S2). As trimethyl H3K36 is regarded as an activating mark for gene expression (5), we concurrently conducted ChIP-Seq for RNA polymerase II to produce a readout of transcriptional activity, and observed a significant correlation between H3K36me3 and RNA polymerase II binding for the 156 differentially bound genes ($R^2 = 0.923$, $P < 0.0001$; Fig. 2D). By integrating the H3K36me3 and RNA polymerase II data, we derived a ranked list of differentially trimethyl-bound and expressed genes (Fig. 2E). Interrogating this ranked list using our integrated pediatric GBM expression dataset showed highly significant enrichment for G34-associated gene signatures in the differentially bound and expressed genes in G34-mutant KNS42 cells (ES = 0.84–0.86, FWER $P = 0.02$ –0.03; FDR $q = 0.03$ –0.04; Fig. 2F).

To investigate the functions of the transcriptional programs targeted by this novel mechanism, we conducted gene ontology analysis of the differentially bound and expressed genes. These data revealed highly significant enrichment of the processes involved in forebrain and cortex development, as well as differentiation of neurons and regulation of cell proliferation (Fig. 2G). We identified a subset of 16 genes to be part of the core enrichment group showing significant overlap between G34-mutant pediatric GBM specimens and transcription driven by differential binding of H3K36me3 in KNS42 cells (Supplementary Table S3). By mapping the expression of these genes to published signa-

tures of restricted spatiotemporal areas of brain development (6), we noted highly elevated levels at embryonic and early fetal time-points, which rapidly tailed off through mid-late fetal development and postnatal and adult periods (Fig. 2H). Expression of the G34 core enrichment genes was particularly pronounced in the early fetal amygdala, inferior temporal cortex, and the caudal, medial, and lateral ganglionic eminences (Fig. 2H). Developmental expression patterns of G34 mutation-associated genes were in contrast to those observed with K27 mutation signatures derived from pediatric GBM specimens, which correlated with those of the embryonic upper rhombic lip, early-mid fetal thalamic, and cerebellar structures, and peaked during the mid-late fetal period (Supplementary Fig. S4).

Specifically, the G34 mutation drives expression of numerous highly developmentally regulated transcription factors, including as an exemplar *DLX6* (distal-less homeobox 6), which encodes a homeobox transcription factor that plays a role in neuronal differentiation in the developing forebrain (7). The highly significant differential H3K36me3 and RNA polymerase II binding observed by ChIP-Seq (Fig. 3A) was validated by ChIP-quantitative real-time PCR (qPCR) (Fig. 3B), and expression of *DLX6* was noted to be significantly higher in G34 pediatric GBM samples than K27-mutant or wild-type tumors in the integrated gene expression datasets at the mRNA level (Fig. 3C), and at the protein level in a tissue microarray comprising 46 pediatric and young adult GBM cases (Fig. 3D and Supplementary Table S4). Other similarly validated forebrain development-associated transcription factor genes included *ARX* (8), *DLX5* (7), *FOXA1* (9), *NR2E1* (10), *POU3F2* (11), and *SP8* (ref. 12; Supplementary Fig. S5–S10). Moreover, a number of key determinants of cell fate were also found to be differentially bound by H3K36me3 and expressed in G34-mutant cells. These included *MSH1* (Musashi-1; ref. 13; Supplementary Fig. S11); *EYA4* (eyes absent homolog 4; ref. 14; Supplementary Fig. S12); and *SOX2*, which is required for stem cell maintenance (Fig. 3E–H).

Strikingly, the most significant differentially bound and expressed gene in our G34-mutant KNS42 cells was *MYCN* (33-fold H3K36me3 compared with SF188, DESeq $P = 7.94 \times 10^{-8}$; 60-fold RNA Pol II, DESeq $P = 1.59 \times 10^{-8}$; Fig. 4A–D). Of note, a small number of *H3F3A* wild-type tumors also expressed high levels of *MYCN*, and were found to be *MYCN* gene amplified (Fig. 4C). However, amplification was not seen in G34-mutant tumors, which parallels observations in diffuse intrinsic pontine glioma where *MYCN* amplification was found in wild-type, but not K27-mutant, tumors (15). Transduction of the G34V mutation into normal human astrocytes (NHA) and transformed human fetal glial cells (SVG) conferred an approximately 2- to 3-fold increase in *MYCN* transcript levels over wild-type-transduced controls, validating these observations (Supplementary Fig. S13). *H3F3A* G34 mutation may therefore represent an alternative mechanism of enhancing expression levels of *MYCN* in pediatric GBM.

Targeting *MYCN* is an attractive therapeutic intervention in tumors harboring gene mutation such as neuroblastoma (16), and direct inhibition by siRNA knockdown in KNS42 cells reduced cell viability in proportion to the reduction of protein levels observed (Fig. 4E). Pharmacologic agents that

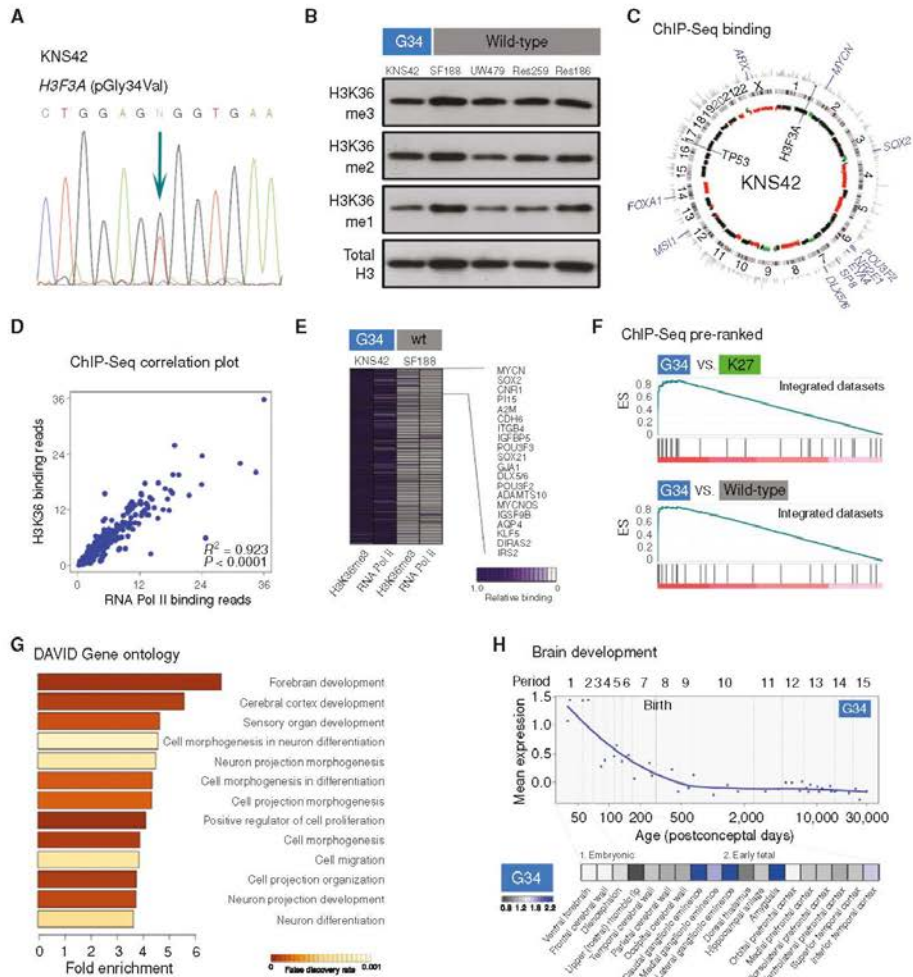


Figure 2. Differential binding of H3K36me3 in G34-mutant KNS42 cells drives pediatric GBM expression signatures. **A**, Sanger sequencing trace for KNS42 pediatric GBM cells reveals a heterozygous c.1046G>T p.(Gly34Val)H3F3A mutation. **B**, Western blot analysis for mono-(me1), di-(me2) and tri-(me3) methylated histone H3 in G34-mutant KNS42 and wild-type (wt) pediatric glioma cell lines. Total H3 is used as an extracted histone loading control. **C**, Circos plot representing the KNS42 genome, aligned with chromosomes 1 to X running clockwise from 12 o'clock. Outer ring, H3K36me3 ChIP-Seq binding. Grey, all binding; blue, differential binding in KNS42 versus SF188. Selected differentially bound developmental transcription factors and pluripotency genes are labeled. Inner ring, DNA copy number. Green points, copy number gains; black points, normal copy number; red points, copy number loss. Single base mutations in selected genes (H3F3A;G34V and TP53;R342*) are labeled inside the circle. **D**, correlation plot of RNA polymerase II versus H3K36me3 for 65 differentially trimethyl-bound regions by ChIP-Seq in KNS42 cells. $R^2 = 0.66$; $P < 0.0001$. **E**, heatmap representing a ranked list of differentially bound H3K36me3 and RNA polymerase II in G34V KNS42 versus wild-type SF188 cells, with top 20 genes listed. **F**, GSEA for pre-ranked differentially bound genes identified in ChIP-Seq versus those in the integrated gene expression datasets. Top, G34 versus K27; ES = 0.96; FWER $P = 0.03$, FDR $q = 0.03$. Bottom, G34 versus wild-type; ES = 0.84, FWER $P = 0.02$, FDR $q = 0.04$. **G**, DAVID gene ontology analysis for pre-ranked list of differentially bound genes identified in ChIP-Seq. Fold enrichment of processes are plotted and colored by FDR q value. **H**, top, mean expression of the G34 core enrichment signature in a temporal gene expression dataset of human brain development. Period 1, embryonal; periods 2–7, fetal; periods 8–12, postnatal; periods 13–15, adulthood. Bottom, heatmap representing spatial differences in G34 core enrichment signature expression in structures within embryonic and early fetal development, with highest levels mapping to the ganglionic eminences and amygdala.

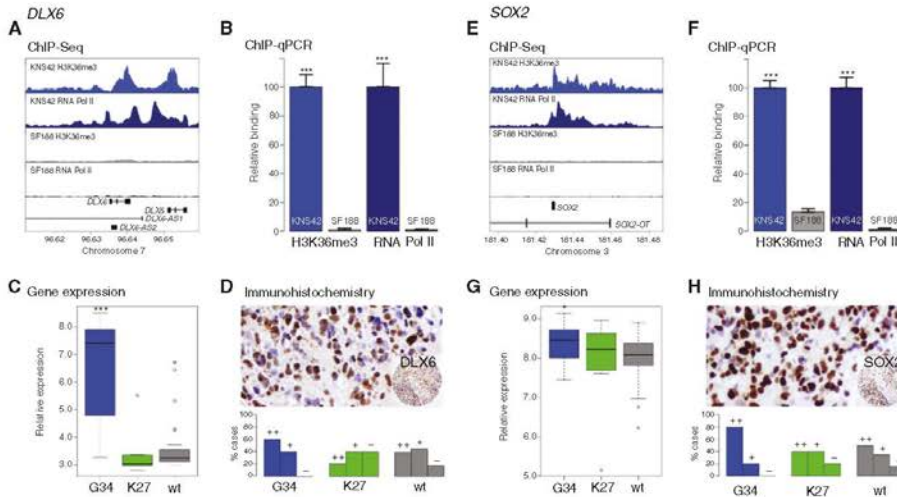


Figure 3. G34 induces a transcriptional program linked to forebrain development and self-renewal. **A**, ChIP-Seq of H3K36me3 and RNA polymerase II binding for G34-mutant KNS42 (blue) and wild-type (wt) SF188 cells (gray) for the *DLX6* locus, which also encompasses the transcripts *DLX5*, *DLX6-AS1*, and *DLX6-AS2*. **B**, validation of ChIP-Seq data by ChIP-qPCR using specific primers targeting *DLX6*. Blue bars, KNS42; gray, SF188. **C**, boxplot of *DLX6* expression in the integrated pediatric GBM samples stratified by *H3F3A* status. Blue box, G34; green, K27; gray, wild-type. **D**, top, immunohistochemistry for *DLX6* protein in a G34 mutant pediatric GBM sample RMH2465. Bottom, barplot of *DLX6* expression in a pediatric GBM tissue microarray stratified by *H3F3A* status. Blue bars, G34; green, K27; gray, wild-type. **E**, ChIP-Seq of H3K36me3 and RNA polymerase II binding for G34-mutant KNS42 (blue) and wild-type SF188 cells (gray) for the *SOX2* locus, which also encompasses the *SOX2-OT* transcript. **F**, validation of ChIP-Seq data by ChIP-qPCR using specific primers targeting *SOX2*. Blue bars, KNS42; gray, SF188. **G**, boxplot of *SOX2* expression in the integrated pediatric GBM samples stratified by *H3F3A* status. Blue box, G34; green, K27; gray, wild-type. **H**, top, immunohistochemistry for *SOX2* protein in a G34-mutant pediatric GBM sample RMH2465. Bottom, barplot of *SOX2* expression in a pediatric GBM tissue microarray stratified by *H3F3A* status. Blue bars, G34; green, K27; gray, wild-type. **+**, strong expression; **+**, moderate expression; **-**, negative.

directly inhibit Myc transcription factors, however, remain elusive. We therefore carried out a synthetic lethal screen to ascertain how we might target these *H3F3A* G34-mutant, MYCN-driven tumors in the clinic. We utilized a series of siRNAs directed against 714 human kinases against our panel of pediatric glioma cell lines to identify those which conferred selective cell death to the MYCN-expressing KNS42 cells versus wild-type, non-MYCN-expressing controls (Fig. 4F). The most significant synthetically lethal hits in the G34-mutant cells compared with *H3F3A* wild-type were kinases that have been previously associated with stabilization of MYCN protein, specifically CHK1 (checkpoint kinase 1; ref. 17) and AURKA (aurora kinase A; ref. 18). Knockdown of AURKA by an independent set of 4 individual oligonucleotides targeting the gene led to a concurrent reduction of MYCN protein in KNS42 cells (Fig. 4G). This destabilization of MYCN was also observed in a dose-dependent manner using a highly selective small-molecule inhibitor of AURKA, VX-689 (also known as MK-5108; ref. 19), which in addition led to a significant reduction in viability of the G34-mutant cells (Fig. 4H). Together, these data show the use of targeting MYCN stability in *H3F3A* G34-mutant pediatric GBM as a means of treating this subgroup of patients.

DISCUSSION

Emerging evidence strongly suggests that pediatric GBMs with *H3F3A* mutations can be subclassified into distinct entities. Our data indicate key molecular and clinical differences between G34- and K27-mutant tumors, reflecting the anatomic specificity (K27 tumors restricted to the pons and thalamus and G34 to the cerebral hemispheres; ref. 15; Supplementary Table S4) and likely distinct developmental origins of these disease subgroups. Using the only known model of *H3F3A*-mutant cells to date, we propose that the gene expression signature associated with G34 mutation in pediatric GBM patient samples is likely driven by a genomic differential binding of the transcriptionally activating H3K36me3 mark.

Mapping these gene expression signatures to publicly available datasets of human brain development shows a strong overlap with the ganglionic eminences of the embryonic and early fetal periods. These structures represent a transiently proliferating cell mass of the fetal subventricular zone, are the source of distinct neuroglial progenitors (20), and are therefore strong candidates for the location of the cells of origin of cerebral hemispheric G34-driven pediatric GBM. As with other pediatric brain tumors (21, 22),

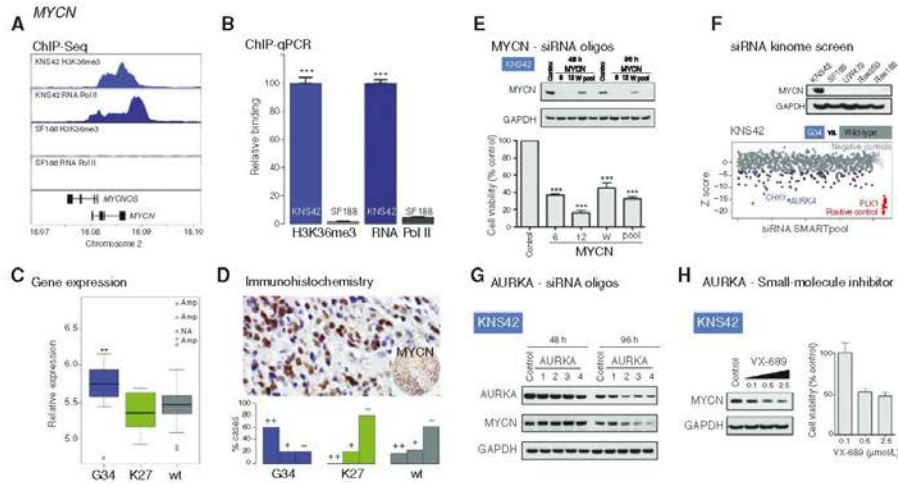


Figure 4. G34 H3K36me3 upregulates MYCN which is selectively targetable by kinases that destabilize the protein. **A**, ChIP-Seq of H3K36me3 and RNA polymerase II binding for G34-mutant KNS42 (blue) and wild-type (wt) SF198 cells (gray) for the MYCN locus, which also encompasses the MYCNOS transcript. **B**, validation of ChIP-Seq data by ChIP-qPCR using specific primers targeting MYCN. Blue bars, KNS42; gray, SF198. **C**, boxplot of MYCN expression in the integrated pediatric GBM samples stratified by H3F3A status. Blue box, G34; green, K27; gray, wild-type. **D**, top, immunohistochemistry for MYCN protein in a G34-mutant pediatric GBM sample, RMH2465 bottom, barplot of MYCN expression in a pediatric GBM tissue microarray stratified by H3F3A status. Blue bars, G34; green, K27; gray, wild-type. **E**, effects on cell viability after siRNA knockdown at 7 days. **F**, siRNA screen for 714 human kinases in KNS42 cells. Western blot analysis showing expression of MYCN protein in G34-mutant KNS42 cells in contrast to a panel of wild-type pediatric glioma lines. GAPDH is used as a loading control. Kinase targets are plotted in plate well order along the x-axis, and Z-scores along the y-axis. PLK1 is used as a positive control and is plotted in red. Negative controls are colored light gray, and kinases with Z-scores greater than -2.0 (no effect on cell viability) are colored gray. "Hits" (Z score < -2.0) are colored dark gray or blue, the latter if the effect on cell viability is specific to KNS42 cells and not in panel of 4 H3F3A wild-type pediatric glioma cell lines. The most significant and selective hits were for CHK1 and AURKA. **G**, effect of knockdown of AURKA on MYCN levels in KNS42 cells. Western blot analysis for AURKA and MYCN in KNS42 cells treated with individual oligonucleotides directed against AURKA for 48 and 96 hours. GAPDH is used as a loading control. **H**, effect of a selective small-molecule inhibitor of AURKA on MYCN protein levels and cell viability. Left, Western blot analysis for MYCN protein in KNS42 cells after exposure to 0.1, 0.5, and 2 to 5 $\mu\text{mol/L}$ VX-689 (triangle). GAPDH is used as a loading control. Right, barplot showing effects on cell viability of KNS42 cells exposed to 0.1, 0.5, and 2 to 5 $\mu\text{mol/L}$ VX-689. **C**, **D**, **E**, **F**, **G**, **H**, **P** < 0.001, t test versus control.

mutation-driven subgroups of GBM retain gene expression signatures related to discrete cell populations from which these distinct tumors may arise. In addition, this mutation-driven differential H3K36me3 binding leads to a significant upregulation of numerous genes associated with cell fate decisions. Thus, we have identified a transcriptional readout of the likely developmental origin of G34-mutant GBM coupled with a self-renewal signature we previously identified in KNS42 cells (23) driven by mutation-induced differential binding of H3K36me3.

Significantly, the G34 mutation additionally upregulates MYCN through H3K36me3 binding. It was recently reported that the forced overexpression of stabilized MYCN protein in neural stem cells of the developing mouse forebrain gave rise to GBMs (24), and thus we provide the mechanism by which the initiating tumorigenic insult is delivered at the correct time and place (25) during neurogenesis. Targeting stabilization of MYCN protein via synthetic lethality approaches in H3F3A G34-mutant pediatric

GBM provides a potential novel means of treating this subgroup of patients.

METHODS

Primary Pediatric Glioblastoma Expression Profiling

Expression data from the Schwartzentruber and colleagues (ref. 2; GSE34824) and Paugh and colleagues (ref. 3; GSE19578) studies were retrieved from the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) and analyzed in GenePattern using a signal-to-noise metric. GSEA was implemented for testing of enrichment of gene lists. Pediatric GBM expression signatures were mapped to specific developmental stages and anatomic locations using a spatiotemporal gene expression dataset of human brain development in Kang and colleagues (ref. 6; GSE25219).

Tissue Microarrays

Immunohistochemistry for DLX6 (NBP1-85929, Novus Biologicals), SOX2 (EPR3131, Epitomics), and MYCN (#9405, Cell Signaling) was carried out on tissue microarrays consisting of 46 cases of

RESEARCH BRIEF

Bjerke et al.

pediatric and young adult GBM ascertained for *H3F3A* mutation by Sanger sequencing.

Cell Line Analysis

Pediatric GBM KNS42 cells were obtained from the JCRB (Japan Cancer Research Resources) cell bank. Pediatric SF188 cells were kindly provided by Dr. Daphne Haas-Kogan (University of California San Francisco, San Francisco, CA), and UW479, Res259, and Res186 were kindly provided by Dr. Michael Bobola (University of Washington, Seattle, WA). All cells have been extensively characterized previously (4), and were authenticated by short tandem repeat (STR) profiling. Western blot analysis was carried out for total histone H3 (ab97968, Abcam), as well as H3K36 trimethylation (ab9050, Abcam), dimethylation (ab9049, Abcam), and monomethylation (ab9050, Abcam), after histone extraction using a histone purification minikit (ActiveMotif), and quantitated by scanning on the Storm 860 Molecular Imager (GE Healthcare) and analyzed using ImageQuant software (GE Healthcare). Additional Western blots for MYCN (#9405, Cell Signaling), ATRX (sc-15408, Santa Cruz), and glyceraldehyde-3-phosphate dehydrogenase (GAPDH; #2118, Cell Signaling) were carried out according to standard procedures.

Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) was carried out employing antibodies against H3K36me3 and RNA polymerase II using the HistonePath and TranscriptionPath assays by ActiveMotif. Whole genome sequencing was carried out using an Illumina HiSeq2000 instrument with more than 30-fold coverage. Validation of active regions was carried out by ChIP-quantitative PCR (qPCR).

siRNA Screening and Validation

siRNA screening was carried out on a library of 714 human kinases using Dharmacon SMARTpools (Dharmacon), with cell viability estimated via a highly sensitive luminescent assay measuring cellular ATP levels (CellTiter-Glo, Promega). Z-scores were calculated using the median absolute deviation of all effects in each cell line. Individual ON-TARGETplus oligonucleotides for validation were obtained from Dharmacon and knockdown validated by Western blot analysis for AURKA (#4718, Cell Signaling) according to standard procedures for up to 96 hours. The AURKA-selective small-molecule inhibitor VX-689 (MK-5108) was obtained from Selleckchem and assayed for up to 5 days. Effects on cell viability were assessed by CellTiter-Glo (Promega). siRNAs targeting human MYCN were custom designs and kindly provided by Janet Shipley (The Institute of Cancer Research, London, United Kingdom).

Disclosure of Potential Conflicts of Interest

L. Bjerke, Alan Mackay, M. Nandhabalan, A. Burford, A. Jury, S. Popov, D.A. Bax, D. Carvalho, K.R. Taylor, M. Vinci, I. Bajrami, I.M. McGonnell, C.J. Lord, A. Ashworth, P. Workman, and C. Jones are employees of The Institute of Cancer Research, which has a commercial interest in AURKA and CHK1 inhibitors. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: L. Bjerke, A. Mackay, M. Nandhabalan, D. Bax, D. Hargrave, P. Workman, C. Jones
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): L. Bjerke, A. Mackay, M. Nandhabalan, A. Burford, A. Jury, S. Popov, D.A. Bax, D. Carvalho, K.R. Taylor, M. Vinci, I. Bajrami, C.J. Lord, D. Hargrave, P. Workman, C. Jones
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L. Bjerke, A. Mackay, M. Nandhabalan, A. Burford, A. Jury, S. Popov, D. Bax, K.R. Taylor,

M. Vinci, I.M. McGonnell, D. Hargrave, A. Ashworth, P. Workman, C. Jones

Writing, review, and/or revision of the manuscript: L. Bjerke, A. Mackay, M. Nandhabalan, I.M. McGonnell, R.M. Reis, D. Hargrave, A. Ashworth, P. Workman, C. Jones

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): L. Bjerke, M. Nandhabalan, I. Bajrami

Study supervision: P. Workman, C. Jones

Acknowledgments

The authors acknowledge NHS funding to the National Institute for Health Research Biomedical Research Centre.

Grant Support

This work is supported by Cancer Research UK, the Wellcome Trust, the Samantha Dickson Brain Tumour Trust, and The Stravros Niarchos Foundation.

Received September 20, 2012; revised January 25, 2013; accepted January 28, 2013; published OnlineFirst March 28, 2013.

REFERENCES

- Jones C, Perryman L, Hargrave D. Paediatric and adult malignant glioma: close relatives or distant cousins? *Nat Rev Clin Oncol* 2012;9:400-13.
- Schwartzentruber J, Korshunov A, Liu XY, Jones DT, Pfaff E, Jacob K, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* 2012;482:226-31.
- Paugh BS, Qu C, Jones C, Liu Z, Adamowicz-Brice M, Zhang J, et al. Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease. *J Clin Oncol* 2010;28:3061-8.
- Bax DA, Mackay A, Little SE, Carvalho D, Viana-Pereira M, Tamber N, et al. Molecular and phenotypic characterisation of paediatric glioma cell lines as models for preclinical drug development. *PLoS ONE* 2009;4:e5209.
- Wagner EJ, Carpenter PB. Understanding the language of Lys36 methylation at histone H3. *Nat Rev Mol Cell Biol* 2012;13:115-26.
- Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, et al. Spatio-temporal transcriptome of the human brain. *Nature* 2011;478:483-9.
- Panganiban G, Rubenstein JL. Developmental functions of the Distal-less/Dlx homeobox genes. *Development* 2002;129:4371-86.
- Kitamura K, Yanazawa M, Sugiyama N, Miura H, Iizuka-Kogo A, Kuaka M, et al. Mutation of ARX causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans. *Nat Genet* 2002;32:359-69.
- Chatterjee S, Bourque G, Lufkin T. Conserved and non-conserved enhancers direct tissue specific transcription in ancient germ layer specific developmental control genes. *BMC Dev Biol* 2011;11:63.
- Monaghan AP, Bock D, Gass P, Schwaeger A, Wolfer DP, Lipp HP, et al. Defective limbic system in mice lacking the tailless gene. *Nature* 1997;390:515-7.
- McBilley RJ, de Diaz MO, Schonemann MD, Hooshmand F, Rosenfeld MG. Transcriptional regulation of cortical neuron migration by POU domain factors. *Science* 2002;295:1528-32.
- Zembrzycki A, Griesel G, Stoykova A, Mansouri A. Genetic interplay between the transcription factors Sp8 and Emx2 in the patterning of the forebrain. *Neural Dev* 2007;2:8.
- Okano H, Kawahara H, Toriya M, Nakao K, Shibata S, Imai T. Function of RNA-binding protein Musashi-1 in stem cells. *Exp Cell Res* 2005;306:349-56.
- Li X, Oghi KA, Zhang J, Krone A, Bush KT, Glass CK, et al. Bta protein phosphatase activity regulates Six1-Dach-Bta transcriptional effects in mammalian organogenesis. *Nature* 2003;426:247-54.

15. Khuong-Quang DA, Buczkowicz P, Rakopoulos P, Liu XY, Fontebasso AM, Bouffet E, et al. K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas. *Acta Neuropathol* 2012;124:439-7.
16. Gustafson WC, Weiss WA. Myc proteins as therapeutic targets. *Oncogene* 2010;29:1249-59.
17. Cole KA, Huggins J, Laquaglia M, Hulderman CE, Russell MR, Bosse K, et al. RNAi screen of the protein kinome identifies checkpoint kinase 1 (CHK1) as a therapeutic target in neuroblastoma. *Proc Natl Acad Sci U S A* 2011;108:3336-41.
18. Otto T, Horn S, Brockmann M, Eilers U, Schuettrumpf L, Popov N, et al. Stabilization of N-Myc is a critical function of Aurora A in human neuroblastoma. *Cancer Cell* 2009;15:67-73.
19. Shimomura T, Hasako S, Nakatsuru Y, Mita T, Ichikawa K, Kodera T, et al. MK-5108, a highly selective Aurora-A kinase inhibitor, shows antitumor activity alone and in combination with docetaxel. *Mol Cancer Ther* 2010;9:157-66.
20. Miyoshi G, Hjerling-Lefler J, Karayannis T, Sousa VH, Butt SJ, Bartsch J, et al. Genetic fate mapping reveals that the caudal ganglionic eminence produces a large and diverse population of superficial cortical interneurons. *J Neurosci* 2010;30:1582-94.
21. Johnson RA, Wright KD, Poppleton H, Mohankumar KM, Finkelstein D, Pounds SB, et al. Cross-species genomics matches driver mutations and cell compartments to model ependymoma. *Nature* 2010;466:632-6.
22. Gibson P, Tong Y, Robinson G, Thompson MC, Currie DS, Eden C, et al. Subtypes of medulloblastoma have distinct developmental origins. *Nature* 2010;468:1095-9.
23. Gaspar N, Marshall L, Perryman L, Bax DA, Little SE, Viana-Pereira M, et al. MGMT-independent temozolomide resistance in pediatric glioblastoma cells associated with a PI3-kinase-mediated HOX/stem cell gene signature. *Cancer Res* 2010;70:9243-52.
24. Swartling FJ, Savov V, Persson AI, Chen J, Hackett CS, Northcott PA, et al. Distinct neural stem cell populations give rise to disparate brain tumors in response to N-MYC. *Cancer Cell* 2012;21:601-13.
25. Phoenix TN, Gilbertson RJ. There's a time and a place for MYCN. *Cancer Cell* 2012;21:593-5.

Recurrent activating *ACVR1* mutations in diffuse intrinsic pontine glioma

Kathryn R Taylor^{1,13}, Alan Mackay^{1,13}, Nathalie Truffaux², Yaron S Butterfield³, Olena Morozova⁴, Cathy Philippe², David Castel², Catherine S Grasso⁵, Maria Vinci¹, Diana Carvalho¹, Angel M Carcaboso⁶, Carmen de Torres⁶, Ofelia Cruz⁶, Jaume Mora⁶, Natacha Entz-Werle⁷, Wendy J Ingram⁸, Michelle Monje⁹, Darren Hargrave¹⁰, Alex N Bullock¹¹, Stéphanie Puget¹², Stephen Yip³, Chris Jones¹ & Jacques Grill²

Diffuse intrinsic pontine gliomas (DIPGs) are highly infiltrative malignant glial neoplasms of the ventral pons that, due to their location within the brain, are unsuitable for surgical resection and consequently have a universally dismal clinical outcome. The median survival time is 9–12 months, with neither chemotherapeutic nor targeted agents showing substantial survival benefit in clinical trials in children with these tumors¹. We report the identification of recurrent activating mutations in the *ACVR1* gene, which encodes a type I activin receptor serine/threonine kinase, in 21% of DIPG samples. Strikingly, these somatic mutations (encoding p.Arg206His, p.Arg258Gly, p.Gly328Glu, p.Gly328Val, p.Gly328Trp and p.Gly356Asp substitutions) have not been reported previously in cancer but are identical to mutations found in the germ line of individuals with the congenital childhood developmental disorder fibrodysplasia ossificans progressiva (FOP)² and have been shown to constitutively activate the BMP–TGF- β signaling pathway. These mutations represent new targets for therapeutic intervention in this otherwise incurable disease.

Recent high-throughput sequencing approaches have found a striking prevalence of mutations in the genes for the histone variants H3.3 (*H3F3A*) or H3.1 (*HIST1H3B*) that encode p.Lys27Met substitutions in the childhood brain tumor DIPG³. This lysine-to-methionine substitution confers a *trans*-dominant ablation of global trimethylation at lysine 27 of histone H3 (H3K27me3), which likely profoundly alters gene expression through the derepression of Polycomb repressive complex 2 (PRC2) target genes⁴. Despite these advances in understanding of the distinct biology of these tumors¹, approaches for desperately needed specific therapeutic interventions are not clear, and little has been reported of the additional mutations accompanying these changes.

We carried out whole-genome sequencing on a unique series of 20 pretreatment biopsy samples of DIPG, for which the patients underwent a safe stereotactic procedure⁵, and whole-exome sequencing on a further biopsy case as well as 5 samples obtained at autopsy (Supplementary Table 1). Histone H3 gene mutations encoding p.Lys27Met were observed in 23 of 26 cases (88%), comprising 15 (58%) *H3F3A* mutations and 8 (31%) *HIST1H3B* mutations (Fig. 1a). These were not found in concert with mutations in the chaperones *ATRX* or *DAXX* as has been described for supratentorial pediatric glioblastoma (pGBM)⁶. There was also an absence of other known glioma-related molecular abnormalities such as *IDH1*, *IDH2*, *BRAF* or *FGFR1* mutations and gene fusions. The mutational spectrum of the untreated biopsy cases was not significantly different from that of the autopsy cases ($P > 0.05$; Fig. 1b), although the treatment-naïve samples had a low overall mutation rate, with a mean of 14.8 somatic single-nucleotide variants (SNVs) per sample (range of 0–25), that was significantly lower than observed in the radiation-treated autopsy cases (mean = 32.0, range = 14–50; $P = 0.004$, *t* test). Similarly, there was a significantly lower overall mutation rate in untreated samples taken at biopsy than in autopsy samples (mean of 0.76 versus 1.2 mutations per megabase; $P = 0.023$, *t* test).

Eleven of 26 DIPGs (42%) harbored somatic *TP53* mutations, with a further 6 cases (23%) shown to have SNVs in *PPM1D*, which encodes a regulator of p38 mitogen-activated protein kinase (p38-MAPK)-p53 signaling in response to cellular stress, and an additional case with a somatic *ATM* mutation (Supplementary Fig. 1), demonstrating non-overlapping targeting of a DNA damage response pathway in 18 of 26 DIPGs (69%) (Supplementary Fig. 2). We further identified non-overlapping recurrent alterations in the phosphoinositide 3-kinase (PI3K) pathway targeting *PIK3CA*, *PIK3R1* and *PTEN* through SNVs and microdeletion (Supplementary Fig. 3), in addition to amplification of *MET* (1/26; 4%) as previously described^{7,8} and truncating

¹Institute of Cancer Research, London, UK. ²CNRS UMR 8203, Gustave Roussy, University Paris Sud, Villejuif, France. ³BC Cancer Agency, Vancouver, British Columbia, Canada. ⁴Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, California, USA. ⁵Department of Molecular and Medical Genetics, Oregon Health and Science University, Portland, Oregon, USA. ⁶Pediatric Hematology and Oncology, Hospital Sant Joan de Déu, Barcelona, Spain. ⁷Centre Hospitalier Régional et Universitaire Hautepierre, Strasbourg, France. ⁸Queensland Children's Tumour Bank, Queensland Children's Medical Research Institute, The University of Queensland, Brisbane, Queensland, Australia. ⁹Stanford Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, California, USA. ¹⁰Neuro-oncology and Experimental Therapeutics, Great Ormond Street Hospital, London, UK. ¹¹Structural Genomics Consortium, University of Oxford, Oxford, UK. ¹²Pediatric Neurosurgery, Necker Sick Children's Hospital, Paris, France. ¹³These authors contributed equally to this work. Correspondence should be addressed to C.J. (chris.jones@icr.ac.uk) or J.G. (grill@igr.fr).

Received 30 April 2013; accepted 21 February 2014; published online 6 April 2014; doi:10.1038/ng.2925



LETTERS

mutation of *NF1* (1/26; 4%) (Fig. 1c). We also identified new recurrent somatic mutations in *IGF2R* (2/26; 8%), although these mutations were concurrent with others in the pathway, such that their consequences are unknown. In total, 12 of 26 DIPG cases (46%) harbored some form of alteration predicted to activate the receptor tyrosine kinase (RTK)-PI3K-MAPK pathways (Supplementary Fig. 4).

Heterozygous somatic coding mutations in the *ACVR1* gene, which encodes the activin A type I receptor ALK2, were observed in 7 of

26 cases (27%) (Fig. 1c). These were restricted specifically to codons 328 (c.983G>T, p.Gly328Val, two cases; c.983G>A, p.Gly328Glu, two cases), 258 (c.772C>T, p.Arg258Gly, one case) and 356 (c.1067G>A, p.Gly356Asp, one case), all within the serine/threonine kinase domain, and 206 (c.617G>A, p.Arg206His, one case), within the glycine-serine-rich (GS) domain. Screening an extended series of 26 DIPG biopsy samples by Sanger sequencing identified further recurrences of these mutations and an additional variant at

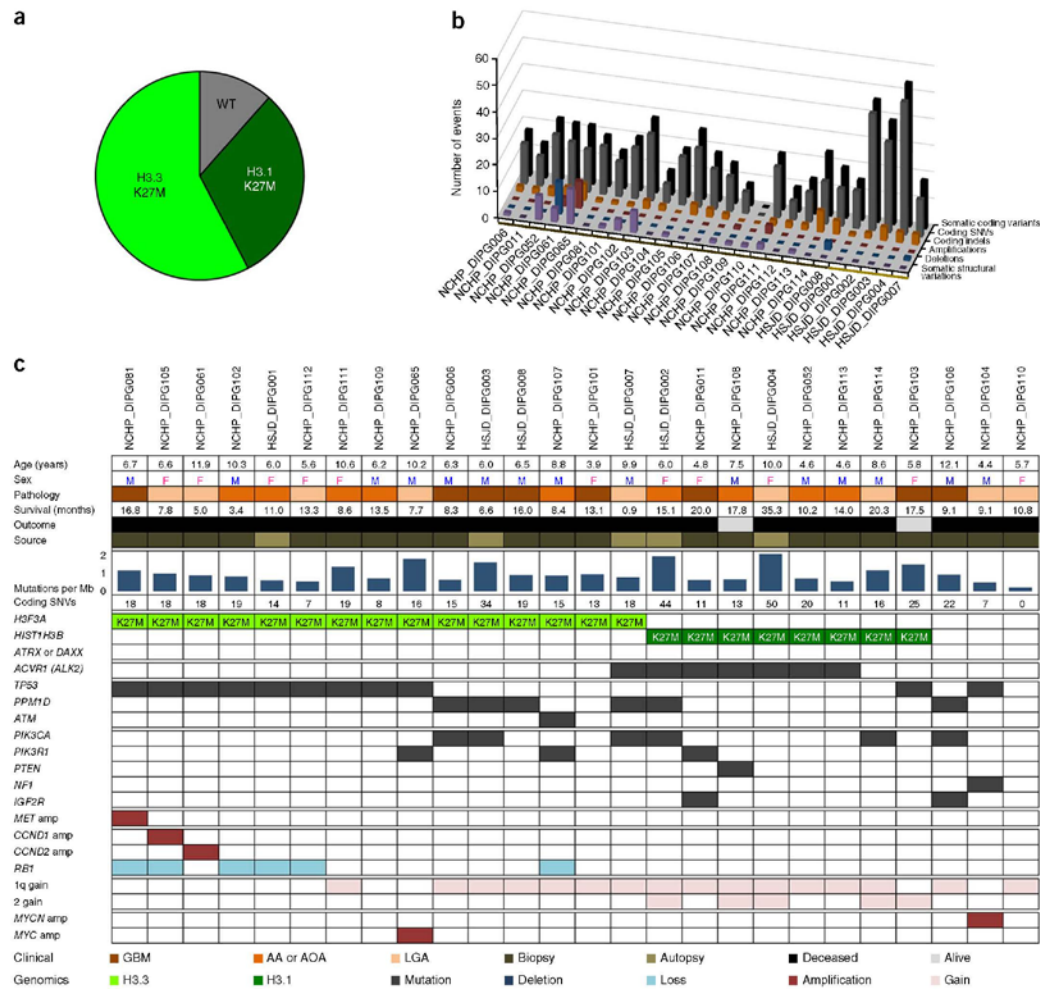


Figure 1 Genomic landscape of DIPG. (a) Pie chart showing the breakdown of histone H3 mutations in our series of 26 DIPG samples (*H3F3A* mutation encoding p.Lys27Met: 15/26, 58%; *HIST1H3B* mutation encoding p.Lys27Met: 8/26, 31%, wild-type histone H3: 3/26, 11%). (b) Mutational spectrum of DIPG. Bar chart showing the total number of somatic coding variants, coding SNVs, indels, amplifications, deletions and structural variations for each DIPG case. Biopsy cases are marked by the dark brown bar along the x axis, and autopsy cases are marked by the light brown bar. (c) Summary of major alterations found. Clinicopathological information for the 26 DIPG samples is provided along with the mutation rate and number of somatic coding SNVs. Mutations, amplifications and deletions are noted for the histone H3 genes and *ATRX* or *DAXX*, the *ATM-TP53-PPM1D* axis, members of the PI3K-MAPK signaling pathways, RTKs and members of the RB pathway as are chromosome 1q and 2 single-copy gains and amplification (amp) of *MYC* or *MYCN*. GBM, glioblastoma multiforme; AA, anaplastic astrocytoma; AOA, anaplastic oligoastrocytoma; LGA, low-grade astrocytoma; M, male; F, female.

© 2014 Nature America, Inc. All rights reserved.



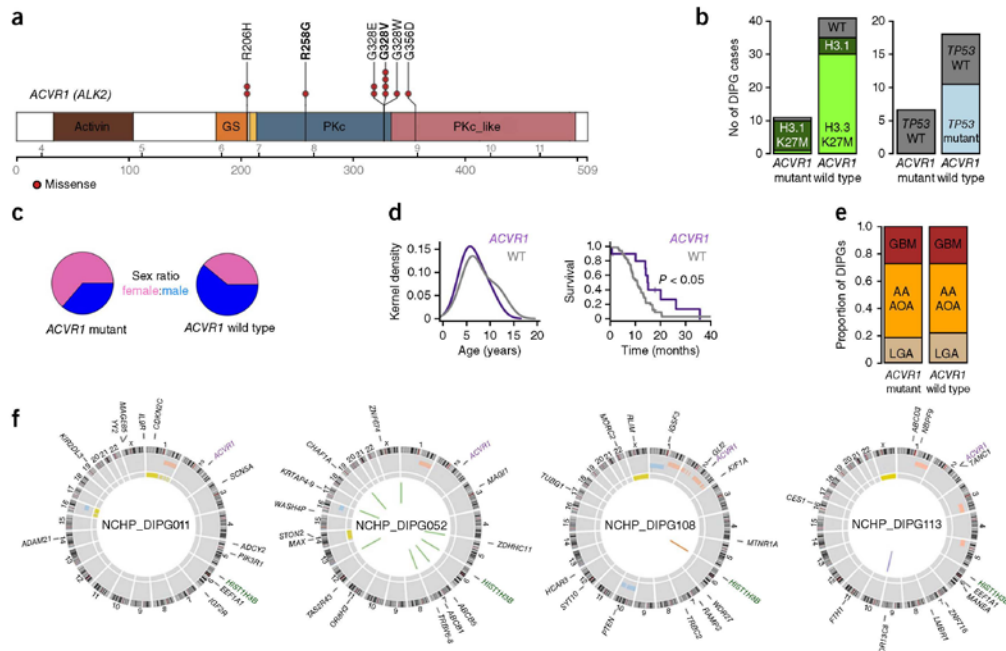
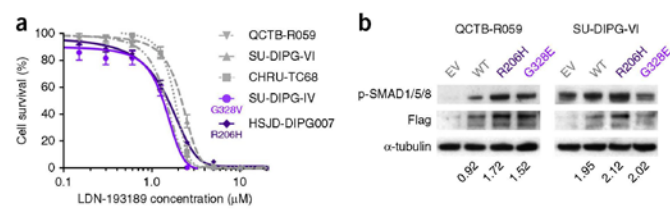


Figure 2 Recurrent *ACVR1* mutations in DIPG. **(a)** Schematic showing recurrent missense mutations in *ACVR1*, overlaid with functional protein domains and exon boundaries. In total, 11 of 52 DIPGs (21%) harbored somatic alterations at 4 residues, all of which have previously been described in the germ line of individuals with FOP. Specific base changes that may be unique to DIPG are highlighted in bold. Activin, activin type I and II receptor domain; GS, TGF- β glycine-serine-rich domain; PKc, protein kinase catalytic domain; PKc_like, protein kinase catalytic domain like. **(b)** Bar graphs showing segregation of activating mutations in *ACVR1* with *HIST1H3B* mutations encoding p.Lys27Met ($P < 0.0001$, Fisher's exact test) and with wild-type (WT) *TP53* ($P = 0.0103$, Fisher's exact test) in our extended series of 52 DIPG cases. **(c)** Sex distribution of cases with *ACVR1* mutations, showing a strong predominance of females in mutant samples. **(d)** Age distribution (left) and overall survival (right) of DIPG cases with *ACVR1* mutations compared to cases with wild-type *ACVR1*. **(e)** Bar plots representing the histological breakdown of samples with mutant and wild-type *ACVR1*. Brown, World Health Organization (WHO) grade 4; orange, WHO grade 3; tan, WHO grade 2. **(f)** Circos plots representing the whole-genome sequences of the four DIPG cases with concurrent *ACVR1* mutation, *HIST1H3B* mutation encoding p.Lys27Met and wild-type *TP53*. The outer ring contains chromosomal ideograms, annotated for somatic SNVs in coding genes. The middle ring plots copy number derived from coverage data: dark red, amplification; pink, gain; dark blue, deletion; light blue, loss. The inner ring represents loss of heterozygosity (LOH; yellow). Drawn inside the circle are structural variations: red, interchromosomal translocation; blue, intrachromosomal translocation; orange, deletion; purple, inversion.

position 328 (c.982G>T, p.Gly328Trp) (Supplementary Fig. 5). Overall, we identified 11 of 52 DIPG samples (21%) harboring mutation in *ACVR1* at 4 different codons (Fig. 2a). These mutations appear highly specific to DIPG. SNVs in the *ACVR1* coding

region are present in the Catalogue of Somatic Mutations in Cancer (COSMIC) database⁹ at an overall frequency of 20 in 5,965 (0.3%), with no individual tumor type harboring them with a frequency of greater than 2% and no mutations observed to affect any of the

Figure 3 *ACVR1* mutations are weakly activating and responsive to targeted inhibition. **(a)** *In vitro* cytotoxicity of the ALK2 inhibitor LDN-193189. Primary cultures were treated with inhibitor for 72 h, and cell viability was measured by CellTiter-Glo in triplicate experiments. Error bars, s.d. The cells used were HSJD-DIPG007 (DIPG; *ACVR1* Arg206His, *H3F3A* Lys27Met), SU-DIPG-IV (DIPG; *ACVR1* Gly328Val, *HIST1H3B* Lys27Met), CHRU-TC68 (DIPG; *ACVR1* wild type, *H3F3A* Lys27Met), SU-DIPG-VI (DIPG; *ACVR1* wild type, *H3F3A* Lys27Met) and QCTB-R059 (thalamic pediatric GBM; *ACVR1* wild type, *H3F3A* Lys27Met). **(b)** *ACVR1* mutations confer increased signaling through phosphorylated SMAD1/5/8. QCTB-R059 and SU-DIPG-VI cells were transfected with construct encoding Flag-tagged Arg206His or Gly328Glu mutant ALK2 and assessed for phosphorylated SMAD1/5/8 (p-SMAD1/5/8) by protein blot. EV, empty vector; WT, wild-type ALK2. α -tubulin is included as a loading control. Numbers below each lane represent phosphorylated SMAD1/5/8 levels quantified relative to the levels of Flag expression.



LETTERS

residues described in the present study, suggestive of a ‘passenger’ effect in other cancers.

ACVR1 mutations were found to cosegregate with the less common *HIST1H3B* mutation encoding p.Lys27Met in the canonical histone H3.1 variant ($P < 0.0001$, Fisher’s exact test) (Fig. 2b), as well as with wild-type *TP53* ($P = 0.0103$, Fisher’s exact test). There was also an association between H3.1 alteration and chromosome 2 gain (on which *ACVR1* is found at 2q24.1; $P = 0.0009$, Fisher’s exact test). *ACVR1* mutations seem to mark a distinct subset of DIPG cases (Supplementary Table 2). There was a marked predominance of females in the *ACVR1*-mutant tumor group compared to cases with wild-type *ACVR1* (1.75:1 versus 0.64:1 female to male ratios, respectively; $P = 0.05$, Fisher’s exact test) (Fig. 2c), as well as a relatively restricted age of onset (Fig. 2d). Cases with tumors harboring *ACVR1* mutations also had longer overall survival times (median of 14.9 versus 10.9 months; $P = 0.05$, log-rank test) (Fig. 2d), although outcome remained very poor. There were no significant differences in histology between the groups (Fig. 2e). Whole-genome sequencing of biopsy samples exemplifying the genotype with concurrent *ACVR1* and *HIST1H3B* mutations harbored an additional 10–19 somatic SNVs and 0–9 structural variations (Fig. 2f).

Remarkably, these somatic mutations in *ACVR1* affected identical residues as the ones described in the germ line of individuals with the autosomal dominant congenital childhood developmental disorder FOP (MIM 135100)². This debilitating disease is characterized by heterotopic ossification of soft connective tissue resulting in severe skeletal abnormalities¹⁰. Individuals with classical clinical features of FOP carry heterozygous mutations in *ACVR1* encoding p.Arg206His that affect the GS activation domain¹¹, whereas atypical cases with a less severe phenotype have been shown to harbor other mutations encoding p.Arg258Ser¹², p.Gly328Glu, p.Gly328Arg or p.Gly328Trp¹³ or p.Gly356Asp¹⁴ or other heterozygous mutations affecting the GS and kinase domains^{2,15}. This latter series of alterations may be exposed to be present at the interface with the GS domain and may abrogate interactions with the negative regulator FKBP12 (refs. 12,13,15). These mutations have been shown to constitutively activate the bone morphogenetic protein (BMP)-dependent transforming growth factor (TGF)- β pathway in the absence of ligand binding, as evidenced by increased phosphorylation of SMAD1, SMAD5 and SMAD8 (SMAD1/5/8) *in vitro*^{4,16}.

To investigate the specific role of *ACVR1* mutations in the context of DIPG, we assembled a panel of four DIPG case-derived primary cultures (and one thalamic pediatric GBM (glioblastoma multiforme) culture harboring an *H3F3A* mutation encoding p.Lys27Met), representing two *ACVR1* mutations (encoding p.Arg206His and p.Gly328Val) and three wild-type lines (Supplementary Table 3). In these models, RNA sequencing (RNA-seq) data demonstrated that the mutant allele was expressed in approximately half of the reads, which was also evidenced by Sanger sequencing of cDNA from case sample NCHP_DIPG011 (Supplementary Fig. 6). Treatment with the selective ALK2 inhibitor LDN-193189 (ref. 17) resulted in marked inhibition of cell viability in all cells, with GI_{50} values (concentrations required to inhibit cell growth by 50%) ranging from 0.86–2.1 μ M, approximately tenfold lower than with the less potent parent compound dorsomorphin, with a trend toward increased sensitivity in the mutant cultures ($P = 0.10$, F test) (Fig. 3a). Transfection of thalamic GBM and DIPG cells with wild-type *ACVR1* (both with *H3F3A* mutation encoding p.Lys27Met) with constructs encoding Flag-tagged mutants conferred increased signaling through phosphorylated SMAD1/5/8, particularly for the Arg206His mutant and, to a lesser extent, for the Gly328Glu mutant (Fig. 3b). *ACVR1* mutation may

only be one mechanism by which this pathway is activated in DIPG, however, as high basal levels of phosphorylated SMAD1/5/8 were also observed for the cells with *H3F3A* mutation encoding p.Lys27Met and wild-type *ACVR1* that were used in this study (Supplementary Fig. 7). This finding may explain the lack of a more robust genotype-dependent response to the inhibitor and may also expand the population of patients who might benefit from targeting of the receptor.

There are no reports to our knowledge of coincident FOP and DIPG, although the clinical features of both typical and atypical cases of FOP can commonly include neurological symptoms and have been reported in children to include cerebellar and brain stem abnormalities^{15,18}, including demyelinated lesions in the pons, both in human cases and mouse models¹⁹. It will nonetheless be a challenge to identify the mechanism by which the temporal and spatial context of BMP-TGF- β pathway activation confers such differing clinical phenotypes. In experimental models of FOP, *ACVR1* mutations are associated with defects in stem cell maintenance, reprogramming and differentiation, offering links with cancer-related cellular processes. First-generation ALK2 inhibitors such as dorsomorphin²⁰ and LDN-193189 (ref. 17) have been shown to downregulate intracellular BMP-TGF- β signaling and to reduce heterotypic ossification, opening the tantalizing possibility of central nervous system (CNS)-penetrant compounds showing a similar potential in a childhood brain tumor otherwise devoid of efficacious treatment options.

URLs. Burrows-Wheeler Aligner (BWA), <http://bio-bwa.sourceforge.net/>; Picard, <http://picard.sourceforge.net/>; Genome Analysis Toolkit, <http://www.broadinstitute.org/gatk/>; SVDetect, <http://svdetect.sourceforge.net/>; Ensembl Variant Effect Predictor, <http://www.ensembl.org/info/docs/variation/vep/>; SIFT, <http://sift.jcvi.org/>; PolyPhen, <http://genetics.bwh.harvard.edu/pph2/>; Catalogue of Somatic Mutations in Cancer (COSMIC) v64, <http://www.sanger.ac.uk/genetics/CGP/cosmic/>; dbSNP Build 137, <http://www.ncbi.nlm.nih.gov/sites/ SNP>; circular binary segmentation, <http://www.biocompare.com/>; APOLLOH, <http://compbio.bccrc.ca/software/apolloh/>; St. Jude Washington University Protein Paint tool, <http://www.explorepcpg.org/>; R3.0.1, <http://www.r-project.org/>.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Raw data have been submitted to the European Genome-phenome Archive under accession EGAS00001000572.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This study was funded by the Cancer Research UK Genomics Initiative (A14078) and makes use of data generated by the St. Jude Children’s Research Hospital–Washington University Pediatric Cancer Genome Project. We are grateful to the DIPG Preclinical Consortium funded by The Cure Starts Now and the Iyla Nsouli Foundation for RNA-seq data. This work is supported by the Stavros Niarchos Foundation, Abbie’s Army, the Iyla Nsouli Foundation, the Royal Marsden Hospital Children’s Department Fund and Fondo Alicia Pueyo. M.M. gratefully acknowledges funding by the National Institutes of Neurological Disease and Stroke (NINDS; grant K08NS070926), Alex’s Lemonade Stand Foundation, the McKenna Claire Foundation and the Dylan Jewett, Elizabeth Stein, Connor Johnson and Zoey Ganesh Memorial Funds. C.P. acknowledges funding from the Agence National de la Recherche. N.T., C.P. and J.G. acknowledge funding from the charity l’Etoile de Martin, and N.E.-W. acknowledges support from Enfants et Santé. A.M.C. acknowledges funding from the Fundación Científica de la Asociación Española Contra el Cáncer. W.J.I. acknowledges funding from the Children’s Health Foundation Queensland and the Brainchild Foundation. The Structural Genomics Consortium is a registered charity (1097737) that receives funds from AbbVie, Boehringer Ingelheim, the Canada



Foundation for Innovation, the Canadian Institutes for Health Research, Genome Canada, GlaxoSmithKline, Janssen, Lilly Canada, the Novartis Research Foundation, the Ontario Ministry of Economic Development and Innovation, Pfizer, Takeda and the Wellcome Trust (092809/Z/10/Z). K.R.T., A.M., M.V., D. Carvalho, D.H. and C.J. acknowledge National Health Service (NHS) funding to the National Institute of Health Research Biomedical Research Centres.

AUTHOR CONTRIBUTIONS

C.J., J.G., D.H. and S.Y. designed the study. C.J. wrote the manuscript. K.R.T., A.M. and C.J. designed and reviewed experiments, and designed and reviewed statistical and bioinformatic analyses. K.R.T. performed experiments. A.M. performed bioinformatic analyses. N.T., D. Castel, M.V. and D. Carvalho performed sample preparation and experiments. Y.S.B., O.M., C.P., C.S.G. and S.Y. performed and reviewed bioinformatic analyses. A.M.C., C.d.T., O.C., J.M., N.E.-W., W.J.L., M.M., A.N.B., S.P. and I.G. provided and prepared samples and experimental materials. All authors reviewed the manuscript during its preparation.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Jones, C., Perryman, L. & Hargrave, D. Paediatric and adult malignant glioma: close relatives or distant cousins? *Nat. Rev. Clin. Oncol.* **9**, 400–413 (2012).
- Katagiri, T. Recent topics in fibrodysplasia ossificans progressiva. *J. Oral Biosciences* **54**, 119–123 (2012).
- Wu, G. *et al.* Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat. Genet.* **44**, 251–253 (2012).
- Lewis, P.W. *et al.* Inhibition of PRG2 activity by a gain-of-function H3 H3 mutation found in pediatric glioblastoma. *Science* **340**, 867–861 (2013).
- Roujeau, T. *et al.* Stereotactic biopsy of diffuse pontine lesions in children. *J. Neurosurg.* **107**, 1–4 (2007).
- Schwartzentruber, J. *et al.* Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482**, 226–231 (2012).
- Pugot, S. *et al.* Mesenchymal transition and *PDGFRA* amplification/mutation are key distinct oncogenic events in pediatric diffuse intrinsic pontine gliomas. *PLoS ONE* **7**, e30313 (2012).
- Paugh, B.S. *et al.* Genome-wide analyses identify recurrent amplifications of receptor tyrosine kinases and cell-cycle regulatory genes in diffuse intrinsic pontine glioma. *J. Clin. Oncol.* **29**, 3999–4006 (2011).
- Forbes, S.A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
- Shore, E.M. & Kaplan, F.S. Inherited human diseases of heterotopic bone formation. *Nat. Rev. Rheumatol.* **6**, 518–527 (2010).
- Shore, E.M. *et al.* A recurrent mutation in the BMP type I receptor *ACVR1* causes inherited and sporadic fibrodysplasia ossificans progressiva. *Nat. Genet.* **38**, 525–527 (2006).
- Bocciardi, R., Bordo, D., Di Duca, M., Di Rocco, M. & Rava220lo, R. Mutational analysis of the *ACVR1* gene in Italian patients affected with fibrodysplasia ossificans progressiva: confirmations and advancements. *Eur. J. Hum. Genet.* **17**, 311–318 (2009).
- Petrie, K.A. *et al.* Novel mutations in *ACVR1* result in atypical features in two fibrodysplasia ossificans progressiva patients. *PLoS ONE* **4**, e5005 (2009).
- Fukuda, T. *et al.* A unique mutation of *ALK2*, G356D, found in a patient with fibrodysplasia ossificans progressiva is a moderately activated BMP type I receptor. *Biochem. Biophys. Res. Commun.* **377**, 905–909 (2008).
- Kaplan, F.S. *et al.* Classic and atypical fibrodysplasia ossificans progressiva (FOP) phenotypes are caused by mutations in the bone morphogenetic protein (BMP) type I receptor *ACVR1*. *Hum. Mutat.* **30**, 379–390 (2009).
- Chaikuad, A. *et al.* Structure of the bone morphogenetic protein receptor *ALK2* and implications for fibrodysplasia ossificans progressiva. *J. Biol. Chem.* **287**, 36990–36998 (2012).
- Yu, P.B. *et al.* BMP type I receptor inhibition reduces heterotopic ossification. *Nat. Med.* **14**, 1363–1369 (2008).
- Tumolo, M., Moscatelli, A. & Silvestri, G. Anaesthetic management of a child with fibrodysplasia ossificans progressiva. *Br. J. Anaesth.* **97**, 701–703 (2006).
- Kan, L. *et al.* CNS demyelination in fibrodysplasia ossificans progressiva. *J. Neurol.* **259**, 2644–2655 (2012).
- Yu, P.B. *et al.* Dorsomorphin inhibits BMP signals required for embryogenesis and iron metabolism. *Nat. Chem. Biol.* **4**, 33–41 (2008).





ONLINE METHODS

Tumor cohort. DIPG samples and matched peripheral blood samples were available from 21 patients who underwent stereotactic biopsy at the Neurosurgery Department of Necker Sick Children's Hospital (Paris, France), 20 of which were subjected to whole-genome sequencing. All patients were clinically diagnosed as having DIPG on the basis of clinical presentation and radiography as part of a multidisciplinary assessment. These patients had diffuse intrinsic tumor centered in the pons and occupying at least 50% of the volume of this structure and had an associated short clinical history of less than 3 months. DNA from an additional 26 biopsy samples was available as a validation cohort. A further five DIPG cases with matched peripheral blood samples were obtained at autopsy at the Hospital Sant Joan de Déu (Barcelona, Spain) and were sequenced after exome capture using Agilent SureSelect technology. All patient material was collected after informed consent and was subject to local research ethics committee approval. Cases included 23 girls and 29 boys (1:1.26 ratio). The median age of the patients was 6.6 years, and the median overall survival time was 11.6 months. A summary of the tumor cohort and clinicopathological information are provided in **Supplementary Table 2**.

Whole-genome and whole-exome sequencing. Exome capture was carried out on the four autopsy cases using the 50Mb Agilent SureSelect platform (Agilent Technologies), and samples underwent paired-end sequencing on an Illumina HiSeq 2000 with a 100-bp read length. Library preparation for the biopsy samples was carried out by the Illumina FastTrack service, and entire genomes underwent paired-end sequencing on an Illumina HiSeq 2000. The median coverage for the tumor genomes was 37–67× (matched normal genomes, 34–41×). Reads were mapped to the hg19 build of the human genome using BWA (Burrows-Wheeler Aligner), and PCR duplicates were removed with PicardTools 1.5.

Genome analysis. Somatic SNVs were called using the Illumina Genome Network (IGN) Cancer Normal pipeline version 1.0.2 and the Genome Analysis Toolkit v2.4-9. Structural variations were called using IGN and SVDetect. Variants were annotated using the Ensembl Variant Effect Predictor v71 tool incorporating SIFT and PolyPhen predictions, COSMIC v64 and dbSNP Build 137 annotations. Copy number was obtained by calculating \log_2 ratios of tumor/normal coverage binned into exons of known genes, smoothed using circular binary segmentation and processed using in-house scripts. LOH was calculated using APOLLOH. Schematics showing the locations of recurrent mutations were produced by the St. Jude Washington University Protein Paint tool. Statistical analysis was carried out using R3.0.1. Continuous variables were analyzed using Student's *t* tests. Count data were compared using Fisher's exact test.

Cell culture and drug sensitivity. Primary cultures were derived from DIPG case samples taken at either biopsy or autopsy at multiple centers, representing

both cases with mutant and wild-type *ACVR1* and with both *H3F3A* and *HIST1H3B* mutations encoding p.Lys27Met, in addition to cells from a pediatric glioblastoma specimen arising in the thalamus with an *H3F3A* mutation encoding p.Lys27Met. A summary of the cases from which these cells were derived is provided in **Supplementary Table 3**. Cells were grown under adherent stem cell conditions using flasks coated with laminin (Sigma) in neurobasal medium (Invitrogen) supplemented with B-27 (Invitrogen) and the growth factors epidermal growth factor (EGF), basal fibroblast growth factor (bFGF), platelet-derived growth factor (PDGF)-AA and PDGF-BB (all from Shenandoah Biotech). The ALK2 inhibitors LDN-193189 (Sigma) and dorso-morphin (Abcam) were tested for their effects on cell viability using a highly sensitive luminescent assay measuring cellular ATP levels (CellTiter-Glo, Promega). Drug was added in various concentrations, and cells were assayed in triplicate after 72 h. Statistical analysis was carried out using GraphPad Prism 6.0 (GraphPad Software).

Allelic expression of *ACVR1*. SU-DIPG-IV cells were subjected to full transcriptome sequencing as part of the DIPG Preclinical Consortium. Counts of reads aligned to the *ACVR1* coding region in NCBI_36 were analyzed for the ratio of mutant to wild-type sequence and visualized in Genome Browse (Golden Helix). RNA from the NCHP_DIPG011 primary tumor was reverse transcribed, PCR amplified and underwent Sanger sequencing to determine whether both mutant and wild-type alleles were expressed (**Supplementary Table 4**).

Overexpression of mutant ALK2. The *ACVR1* mutations encoding p.Arg206His and p.Gly328Glu were introduced into pcDNA3.1 by site-directed mutagenesis as previously described¹⁶, and constructs were transfected into primary QCTB-R059 and SU-DIPG-VI cells using Lipofectamine (Invitrogen), with protein collected after 24 h using standard procedures. Protein blotting was carried out with horseradish peroxidase (HRP)-conjugated antibody to Flag (A8592, Sigma; 1:1,000 dilution) and antibody to phosphorylated SMAD1/5/8 (9511, Cell Signaling Technology; 1:1,000 dilution) under standard conditions. Relative levels of phosphorylated SMAD1/5/8 were measured with ImageJ software (National Institute of Mental Health).

Statistical analysis. Statistical analysis was carried out using GraphPad Prism 6.0 and R 3.0.1. Comparisons of the numbers of coding SNVs and the mutation rates in biopsy and autopsy cases were performed by *t* test. For analysis of categorical association between cases with *ACVR1* mutations and mutations in *HIST1H3B* or *TP53*, sex and histology, Fisher's exact test was used. Differences in survival were analyzed by the Kaplan-Meier method, and significance was determined by the log-rank test. All tests were two-sided, and a *P* value of less than 0.05 was considered significant. A sum-of-squares *F* test was used to assess differences in dose response curves for *ACVR1*-mutant versus wild-type cells.