

Unraveling the genetic and molecular pathogenesis of  
pediatric and young adult glioma

Xiaoyang Liu

Bachelor of Science

Department of Human Genetics

McGill University

Montreal, Quebec, Canada

December, 2012

A thesis submitted to McGill University in partial fulfillment of the  
requirements of the degree of Doctor of Philosophy

© Copyright Xiaoyang Liu, 2012.

## ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my supervisor Dr. Nada Jabado, who has provided me with continuous support and guidance throughout this PhD thesis work. Her great passion and acute intuition for research has always been inspiring and encouraging me as a student. Her patience in mentoring and devotion to student success has enabled my growth and maturation in the scientific field.

I also owe appreciation to the help from my supervisory committee members: Dr. Janusz Rak, Dr. Jacquetta Trasler, Dr. Peter Siegel and Dr. Giuseppina Ursini-Siegel. Many thanks go to our collaborators and their group members, and this is by no means the complete list: Dr. Jacek Majewski, Dr. Stefan Pfister, Dr. Andrey Korshunov, Dr. Steffen Albrecht, Dr. Marcel Kool, Dr. Gelareh Zadeh, Dr. Zhifeng Dong, Dr. Jeffrey Atkinson, Dr. Pierre Lepage, Dr. Abhijit Guha, Dr. Guido Reifenberger and Dr. Uri Tabori.

I am thankful for the help, collaboration and great time given by former and current members of the Jabado lab: Damien Faury, Caroline Sollier, Noha Gerges, Karine Jacob, Adam Fontebasso, Dong Anh Khuong

Quang, Margret Shrinian, Takrima Haque, Hani Halabi, Nisreen Ibrahim, Tenzin Gayden and Denise Bechet. I am also indebted to the administrative staff: Melanie Cotingco, Marlene Aardse, Danuta Rylski, Thomas Leslie and Ross MacKay.

At the end, I gave my special thanks to my dearest parents, for the more than 20 years of support and belief in me. I am lucky and thankful to be able to share my happy and difficult time with them, and to feel their deepest love.

# TABLE OF CONTENTS

TABLE OF CONTENTS .....	iv
LIST OF TABLES .....	vii
LIST OF FIGURES.....	viii
LIST OF ABBREVIATIONS .....	x
CONTRIBUTION OF THE AUTHORS.....	xix
SIGNIFICANCE OF THESIS .....	xxv
Chapter 1: INTRODUCTION.....	1
1.1    General Overview of Gliomas.....	2
1.1.1 Classification of gliomas .....	2
1.1.2 Epidemiology of gliomas .....	4
1.1.3 Clinical aspects of gliomas .....	5
1.1.4 Genetic alteration in gliomas .....	6
1.2    Adult and Pediatric GBM .....	12
1.2.1 Brain location of GBM in children and adults .....	13
1.2.2 Distinct genetic and molecular alterations between adult and pediatric glioblastoma .....	14
1.2.2.1 Adult GBM.....	14
1.2.2.1 (I). Molecular subgrouping.....	15
1.2.2.1 (II). Epigenetic regulation.....	16
1.2.2.1 (II-a) IDH1/2 mutations .....	16
1.2.2.1 (II-b). Downregulation of DNA methyltransferase .....	20
1.2.2.1 (II-c). MGMT promoter hypermethylation .....	21
1.2.2.1 (III). RTK/RAS/PI3K signalling alteration .....	22
1.2.2.1 (IV). P53 pathway alteration .....	24
1.2.2.1 (V). Rb pathway alteration.....	25
1.2.2.2 Pediatric GBM .....	26
1.2.2.2 (I). Molecular subclassification of pediatric GBM .....	27
1.2.2.2 (II). Distinct gene expression from adult GBM.....	28
1.2.2.2 (III). Different copy number alterations from adult GBM.....	30
1.2.2.2 (IV). Genetic mutations in pediatric GBM .....	31
1.2.3 Targeted therapy for glioblastoma.....	32
1.3. Y-box binding protein 1.....	34
1.3.2. YB-1 structure and expression .....	34
1.3.2. Function of YB-1.....	36
1.3.3. Regulation of YB-1 intracellular localization.....	39

1.3.4. Mouse model of YB-1 .....	41
1.3.5. YB-1 involvement in cancer .....	42
1.4. Introduction to whole exome sequencing .....	44
1.4.1. Overview of next-generation sequencing .....	44
1.4.2. Whole exome sequencing technology .....	45
1.4.3. Application of WES in cancer .....	46
1.5. Rationale and objectives of study .....	49
Figures .....	52
Chapter 2: Sub-cellular localization of Y-box Protein 1 regulates proliferation, migration and tumorigenicity in astrocytomas .....	61
2.1. ABSTRACT .....	62
2.2. INTRODUCTION .....	64
2.3. MATERIALS AND METHODS .....	68
2.4. RESULTS .....	73
2.5. DISCUSSION .....	79
2.6. AUTHOR CONTRIBUTIONS .....	83
2.7. FIGURES .....	84
2.8. TABLES .....	90
2.9. SUPPLEMENTARY FIGURES .....	91
CONNECTING TEXT CHAPTER 2 TO 3 .....	96
Chapter 3: Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma .....	97
3.1. ABSTRACT .....	99
3.2. INTRODUCTION .....	101
3.3. MATERIALS AND METHODS .....	102
3.4. RESULTS .....	107
3.5. SUMMARY .....	117
3.6. ACKNOWLEDGEMENTS .....	118
3.7. AUTHOR CONTRIBUTIONS .....	119
3.8. FIGURES .....	120
3.9. SUPPLEMENTARY FIGURES .....	129
3.10. SUPPLEMENTARY TABLES .....	139
3.11. CONNECTING TEXT CHAPTER 3 TO 4 .....	168
Chapter 4: Frequent <i>ATRX</i> mutations and loss of expression in adult diffuse astrocytic tumors carrying <i>IDH1/IDH2</i> and <i>TP53</i> mutations .....	169
4.1. ABSTRACT .....	170
4.2. INTRODUCTION .....	172

4.3. MATERIALS AND METHODS.....	175
4.4. RESULTS .....	179
4.5. DISCUSSION .....	185
4.6. ACKNOWLEDGEMENTS.....	191
4.7. CONFLICT OF INTEREST.....	191
4.8. AUTHOR CONTRIUTIONS .....	191
4.9. FIGURES.....	192
4.10. TABLES .....	198
4.11. SUPPLEMENTARY FIGURES.....	202
4.12. SUPPLEMENTARY TABLES.....	206
Chapter 5: DISCUSSION .....	211
5.1. YB-1: an oncoprotein or anti-oncogenic protein? .....	214
5.2. Whole exome/genome sequencing: current issues.....	218
5.3. Glioma pathogenesis: entering the epigenetic stage. ....	223
5.3.1. H3.3 .....	225
5.3.2. IDH .....	230
5.3.3. ATRX .....	234
5.4. Conclusion remarks and significance of research .....	237
LIST OF REFERENCES .....	239

## LIST OF TABLES

Table 2.1: Overexpression and nuclear localization of YB-1 found in pediatric GBM and pilocytic astrocytoma samples, using immunohistochemistry on tissue microarrays. ....	90
Table S3.1: Clinical information for 48 whole exome sequencing samples .....	139
Table S3.2: Summary of sequence analysis of pediatric GBMs.....	141
Table S3.3: Candidate somatic mutations in the 6 tumor samples with matched germline DNA.....	142
Table S3.4: Summary of somatic mutations in pediatric glioblastoma and 5 cancer types from Parsons et al. ....	146
Table S3.5: Mutations in selected genes H3F3A, ATRX, DAXX, IDH1, PDGFRA, EGFR, TP53.....	147
Table S3.6: Comparison of genes mutated in pediatric GBM and in each of the 4 described adult GBM molecular subgroups. ....	149
Table S3.7: Top 100 differentially expressed genes by standard deviation, used for unsupervised hierarchical clustering, ordered as presented in Figure 3.3b. ....	150
Table S3.8: Numbers of CNAs of each type identified in each tumor sample.....	154
Table S3.9: CNA regions identified in each tumor sample. ....	156
Table 4.1: ATRX mutations detected by Sanger sequencing in gliomas, other published cancer datasets and the ATRX syndrome. ....	198
Table 4.2: <i>ATRX</i> , <i>IDH1</i> , <i>IDH2</i> and <i>TP53</i> alterations and ALT detected in different types of adult gliomas. ....	200
Table S4.1a: Sample information of the 140 adult gliomas. ....	206
Table S4.1b: ATRX alterations in the 17 pediatric gliomas and 34 pilocytic astrocytomas analyzed in this study.....	209
Table S4.2: Number of samples with alternative lengthening of telomeres in different subtypes of adult gliomas. ....	210

## LIST OF FIGURES

Figure 1.1: Incidence rates of primary brain and CNS tumors for different age groups, from CBTRUS statistical report of data from 2005-2009. ....	52
Figure 1.2: Percentage composition of different histological types of primary brain and CNS gliomas (n=90828).....	53
Figure 1.3: Incidence rates of grade I, II, III and IV astrocytomas (Astro), primitive neuroectodermal tumors (PNETs), ependymoma (Epend) and craniopharyngioma (Cranio). Data are collected from CBTRUS from 1997 to 2001. ....	54
Figure 1.4: Genetic pathways leading to grade II and III gliomas, as well as primary and secondary glioblastoma. ....	55
Figure 1.5: Three core signalling pathways disrupted in adult glioblastoma through genetic alterations. ....	56
Figure 1.6: characteristic genetic alterations and gene expression of 4 molecular subtypes of adult glioblastoma.....	57
Figure 1.7: Structure of YB-1 protein. ....	58
Figure 1.8: Multiple functions of YB-1 in the cell.....	59
Figure 1.9: Procedure of whole exome sequencing. ....	60
Figure 2.1: Effect of YB-1 overexpression and silencing on cellular growth in GBM cell lines U87 and SF188.....	84
Figure 2.2: Subcellular localization of YB-1 after ectopic overexpression and shRNA silencing in GBM cell lines. ....	86
Figure 2.3: Nuclear expression of YB-1 caused by YB-1 silencing with shRNA led to increased EGFR; effect of YB-1 overexpression and silencing on cellular growth and migration in NHA cell line. ....	87
Figure 2.4: YB-1 overexpression decreases tumorigenicity of GBM cells in xenograft mouse model.....	89
Figure S2.1: Characterization of molecular alterations in U87, SF188 and NHA cell lines. ....	91
Figure S2.2: Migration and invasion assays for SF188 and U87 showed no significant difference among EV control, YB-1 knock down (shYB1) and overexpression (HAYB1).92	
Figure S2.3: Nuclear cytoplasmic extraction confirming that overexpressed YB-1 is mainly localized in the cytoplasm.....	93
Figure S2.4: Representative IHC pictures showing negative, cytoplasmic and nuclear expression of YB-1 (upper panels), as well as positive and negative EGFR expression (lower panels) in pediatric GBM tumor paraffin-embedded slides.....	94

Figure S2.5: Nuclear expression of YB-1 in pediatric GBM patient samples is significantly associated with overexpression of EGFR. ....	95
Figure 4.1: Most frequent mutations in paediatric GBM. ....	120
Figure 3.2: Mutations in <i>H3F3A</i> , <i>ATRX</i> and <i>DAXX</i> distinguish paediatric from adult GBM. ....	123
Figure 3.3: H3F3A mutation variants show distinct expression profiles and are associated with alternative lengthening of telomeres. ....	127
Figure S3.1: Immunohistochemical staining for ATRX showing concordance between sequencing data and protein expression in samples. ....	129
Figure S3.2a: Comparison of mutated genes between paediatric and adult GBM. ....	130
Figure S3.2b: Comparison of frequently mutated genes indicated that paediatric GBM is distinct from the previously identified molecular subgroups in adult GBM ....	132
Figure S3.3: Unsupervised clustering of the 100 genes most differentially regulated in 27 paediatric glioblastoma also analyzed by whole exome sequencing. ....	134
Figure S3.4: Immunohistochemical staining for ATRX & p53 and ALT FISH of patients with G34V or K27M-H3.3. ....	135
Figure S3.5: H3.3 Lysine 36 is methylated in a G34R mutant (PGBM14). ....	136
Figure S3.6: Single nucleotide polymorphism (SNP) array profiling reveals differences in copy number aberrations (CNAs) in ATRX, H3F3A and TP53 –mutated paediatric glioblastoma. ....	137
Figure 5.1: Schematic representations of the localization of missense mutations identified in ATRX in this study and other cancers. ....	192
Figure 4.2: ATRX alterations significantly overlap with IDH1 or IDH2, and TP53 mutations in adult gliomas. ....	194
Figure 4.3: IDH1 mutations are frequent in astrocytomas with decreased ATRX mRNA expression. ....	195
Figure 4.4: ATRX alterations are frequent in IDH1 (or IDH2) and TP53 mutant adult gliomas regardless of grade. ....	197
Figure S4.1: ATRX immunohistochemical staining of a tissue microarray, with examples of a negative and a positive core at high magnification. ....	202
Figure S4.2: GBM samples age distribution between wild type and mutants in ATRX (a) and IDH (b). ....	203
Figure S4.3: ATRX alteration is frequent in astrocytomas and oligoastrocytomas, but does not target oligodendrogliomas. ....	204
Figure S4.4: Examples of negative and positive alternative lengthening of telomeres in adult gliomas assessed by FISH. ....	205

## LIST OF ABBREVIATIONS

2-HG	D-2-hydroxyglutarate
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
A/P domain	alanine/proline-rich domain
aHGG	adult high grade glioma
ALT	alternative lengthening of telomeres
AML	acute myeloid leukemia
ATRX	$\alpha$ -thalassemial/mental-retardation-syndrome-X-linked
BAQ	base alignment quality
CBTRUS	the Central Brain Tumor Registry of the United States
CCDS	consensus coding sequence
ChIP-seq	chromatin immunoprecipitation-sequencing
CIHR	Canadian Institute for Healthy Research
CIMP	CpG island methylator phenotype
CNA	copy number aberration
CNS	central nervous system
CRS	cytoplasmic retention signal
CSD	cold shock domain
CTD	c-terminal domain

DAXX	death-domain associated protein
DIPG	diffuse intrinsic pontine glioma
EGFR	epidermal growth factor receptor
EGFRvIII	EGFR variant III
EMT	epithelial-to-mesenchymal transition
EV	empty vector
FBS	fetal bovine serum
FISH	fluorescence in situ hybridization
GBM	glioblastoma
G-CIMP	glioma CpG island methylator phenotype
GEP	gene expression profile
H3.3	histone 3 variant 3
Her-2	human epidermal growth factor receptor 2
HGDH	2-hydroxyglutarate dehydrogenase
HGG	high grade glioma
HRM	high-resolution melting
IB	immunoblotting
ICR	Institute for Cancer Research
IDH	isocitrate dehydrogenase
IF	immunofluorescence

IHC	immunohistochemistry
Indels	insertions/deletions
JHDM	Jumonji-C domain histone demethylase
LGG	low grade glioma
LOH	loss of heterozygosity
MAPK	mitogen-activated protein kinase
MBD	methyl-CpG-binding domain
Mdm2	murine double minute 2
MDR	multidrug resistance
<i>MGMT</i>	O <sup>6</sup> -methylguanine-DNA methyltransferase
MHC	major histocompatibility complex
mRNP	messenger ribonucleoproteins
NF1	neurofibromatosis type 1
Nf1	neurofibromin 1
NGS	next-generation sequencing
NHA	normal human astrocyte (HTert-immortalized)
NKFP	National Research and Development Fund
NLS	nuclear localization signal
NTD	n-terminal domain
OTKA	Hungarian Scientific Research Fund

PA	pilocytic astrocytoma
PanNET	pancreatic neuroendocrine tumor
pedGBM	pediatric GBM
pHGG	pediatric high grade glioma
PI3K	phosphoinositide-3-kinase
PIP2	phosphatidylinositol (4,5)-bisphosphate
PIP3	phosphatidylinositol (3,4,5)-triphosphate
PTEN	phosphatase and tensin homolog
Rb	retinoblastoma tumor suppressor protein
RSK	ribosomal S6 kinase
RTK	receptor tyrosine kinase
shYB1	YB-1 knock down using shRNA
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SSCP	single-strand conformation polymorphism
TCGA	the Cancer Genome Atlas Network
TET	TATA-binding protein-associated factor
TMA	tissue microarray
WES	whole exome sequencing
WGS	whole genome sequencing

WHO	World Health Organization
YB-1	Y box binding protein 1
$\alpha$ -KG	$\alpha$ -ketoglutarate

## ABSTRACT

Gliomas are the most common primary brain tumors in children and adults, consisting of astrocytomas, oligodendrogliomas, oligoastrocytomas and ependymomas. Glioblastoma (GBM, grade IV) is the most aggressive astrocytoma with poor clinical outcome. DNA copy number and gene expression signatures indicated differences between the molecular pathogenesis of pediatric and adult cases, and there is currently insufficient information about pediatric GBM to improve treatment. To understand the molecular pathogenesis and develop new therapeutic targets for pediatric and young adult GBM, we studied in this thesis work the molecular function of YB-1, previously identified to be overexpressed in pediatric GBM, in astrocytoma formation. We also performed a comprehensive mutational analysis using whole exome sequencing to further explore the genetic alterations important in pediatric GBM. Our results suggest that YB-1 modulates proliferation and migration, based on its sub-cellular localization, and argue for caution in targeting YB-1 for therapeutic intervention. We also discovered somatic mutations in the H3.3-ATRX-DAXX chromatin remodeling pathway in 44% of pediatric GBMs, and frequent ATRX alterations in adult diffuse astrocytic gliomas. These findings indicate a central role of epigenetic regulation perturbation

in driving pediatric and young adult gliomas, and provide novel targets for therapeutic development.

## ABRÉGÉ

Les gliomes sont les tumeurs cérébrales pédiatriques et adultes les plus fréquentes. Parmi eux, on distingue les astrocytomes, les oligodendrogliomes, les oligoastrocytomes et les épendymomes. Le glioblastome (GBM, grade IV) est l'astrocytome plus agressif avec un sombre pronostic clinique. Les données de profil d'expression de gènes et de variation du nombre de copies de l'ADN ont montré que les tumeurs pédiatriques se distinguent des tumeurs adultes. Cependant, l'état des connaissances actuel est insuffisant pour permettre un traitement plus efficace. Pour comprendre la pathogenèse moléculaire et de développer de nouvelles cibles thérapeutiques pour les GBM pédiatriques et du jeune adulte, nous avons étudié dans ce travail de thèse, la fonction moléculaire de YB-1, précédemment identifié pour être surexprimé dans les GBM pédiatriques, potentiellement impliqué dans la formation des astrocytomes. Nous avons également effectué un séquençage d'exome pour explorer les altérations génétiques majeures dans les GBM pédiatriques. Nos résultats suggèrent que YB-1 module la prolifération et la migration cellulaire, en fonction de sa localisation sub-cellulaire, et incitent à la prudence dans un éventuel ciblage thérapeutique de YB-1. Nous avons également découvert des mutations somatiques dans la voie

H3.3-ATRX-DAXX impliquée dans le remodelage de la chromatine dans 44% des GBM pédiatriques, ainsi que de fréquentes modifications d'ATRX dans les gliomes diffus de l'adulte. Ces résultats montrent le rôle centrale la régulation épigénétique dans la formation des gliomes pédiatriques et du jeune adulte, permettant de fournir de nouvelles cibles pour le développement thérapeutique.

## CONTRIBUTION OF THE AUTHORS

### Chapter 2: Sub-cellular localization of Y-box Protein 1 regulates

**proliferation, migration and tumorigenicity in astrocytomas.** Xiao-Yang

Liu\*, Noha Gerges\*, Damien Faury, Caroline Sollier, Steffen Albrecht,

Brian Meehan, Zhifeng Dong, Peter Siegel, Andrey Korshunov, Stefan

Pfister, Janusz Rak and Nada Jabado. Manuscript in preparation. \*These

authors contributed equally to the manuscript.

**Xiaoyang Liu** participated in the study design of the project, performed in vitro and in vivo functional experiments including immunofluorescence & confocal microscopy, nuclear cytoplasmic fractionation, proliferation & migration assays, intracranial & subcutaneous mice xenograft, as well as the data analysis. She also participated in the immunohistochemical staining of tissue microarrays TMA, analyzed the TMA results, and drafted the manuscript. **Noha Gerges** generated NHA cell lines with YB-1 overexpressed or silenced. **Damien Faury** participated in generating YB-1 overexpressed or silenced SF188 & U87 cell lines, performed nuclear cytoplasmic fractionation, proliferation & migration assays independently, as well as quantitative real-time PCR. **Caroline Sollier** generated some of

the SF188 & U87 cell clones and performed some of the immunofluorescence. **Dr. Steffen Albrecht** participated in the pathological review and analysis of TMA results. **Brian Meehan** helped in the design of mice experiments. **Dr. Zhifeng Dong** performed IHC of tissue microarray. **Dr. Peter Siegel** participated in the study design. **Dr. Andrey Korshunov** and **Dr. Stefan Pfister** provided the TMA samples. **Dr. Janusz Rak** participated in the study design. **Dr. Nada Jabado** conceived the study design, participated in data interpretation, provided leadership, managed the project and drafted the manuscript.

### **Chapter 3: Driver mutations in histone H3.3 and chromatin remodelling**

**genes in paediatric glioblastoma.** Jeremy Schwartzentruber\*, Andrey Korshunov\*, Xiao-Yang Liu\*, David TW Jones, Elke Pfaff, Karine Jacob, Dominik Sturm, Adam M Fontebasso, Dong-Anh Khuong Quang, Martje Tönjes, Volker Hovestadt, Steffen Albrecht, Marcel Kool, Andre Nantel, Carolin Konermann, Anders Lindroth, Natalie Jäger, Tobias Rausch, Marina Ryzhova, Jan O. Korbel, Thomas Hielscher, Peter Hauser, Miklos Garami, Almos Klekner, Laszlo Bognar, Martin Ebinger, Martin U. Schuhmann, Wolfram Scheurlen, Arnulf Pekrun, Michael C. Frühwald, Wolfgang Roggendorf, Christoph Kramm, Matthias Dürken, Jeffrey

Atkinson, Pierre Lepage, Alexandre Montpetit, Magdalena Zakrzewska, Krzystof Zakrzewski, Pawel P. Liberski, Zhifeng Dong, Peter Siegel, Andreas E. Kulozik, Marc Zapatka, Abhijit Guha, David Malkin, Jörg Felsberg, Guido Reifenberger, Andreas von Deimling, Koichi Ichimura, V. Peter Collins, Hendrik Witt, Till Milde, Olaf Witt, Cindy Zhang, Pedro Castelo-Branco, Peter Lichter, Damien Faury, Uri Tabori, Christoph Plass, Jacek Majewski\*\*, Stefan M. Pfister\*\*, Nada Jabado\*\*. *Nature* (482(7384):226-31,2012). \*These authors contributed equally to the manuscript. \*\*co-senior authors on this manuscript.

**Xiaoyang Liu** participated in the study design, IHC staining of tissue microarray, some WES & GEP data analysis, and drafting of the manuscript. She also extracted DNA & RNA, analyzed results from SNP array and tissue microarray IHC&FISH, and performed clinical data analysis and all the statistical analysis. **Jeremy Schwartzentruber** performed WES analysis and participated in drafting the manuscript. **Dr. Andrey Korshunov** provided TMA samples and performed FISH experiments. **Dr. David TW Jones** performed GEP analysis. **Dr. Elke Pfaff**, **Dr. Karine Jacob**, **Dr. Dominik Sturm**, **Adam M Fontebasso**, **Dr. Dong-Anh Khuong Quang**, **Dr. Martje Tönjes** and **Dr. Volker Hovestadt** participated in

the analysis of sequencing results, SNP data, or clinical data. **Dr. Steffen Albrecht** participated in the pathological review and TMA analysis. **Dr. Marcel Kool, Dr. Andre Nantel, Dr. Carolin Konermann, Dr. Anders Lindroth, Dr. Natalie Jäger, Dr. Tobias Rausch, and Dr. Jan O. Korbel** participated in the analysis of GEP data. **Dr. Marina Ryzhova, Dr. Thomas Hielscher, Dr. Peter Hauser, Dr. Miklos Garami, Dr. Almos Klekner, Dr. Laszlo Bogнар, Dr. Martin Ebinger, Dr. Martin U. Schuhmann, Dr. Wolfram Scheurlen, Dr. Arnulf Pekrun, Dr. Michael C. Frühwald, Dr. Wolfgang Roggendorf, Dr. Christoph Kramm, Dr. Matthias Dürken, Dr. Jeffrey Atkinson, Dr. Magdalena Zakrzewska, Dr. Krzysztof Zakrzewski, Dr. Pawel P. Liberski, Dr. Andreas E. Kulozik, Dr. Marc Zapatka, Dr. Abhijit Guha, Dr. David Malkin, Dr. Jörg Felsberg, Dr. Guido Reifenberger, Dr. Andreas von Deimling, Dr. Koichi Ichimura, Dr. V. Peter Collins, Dr. Hendrik Witt, Dr. Till Milde and Dr. Olaf Witt** provided clinical samples. **Dr. Pierre Lepage** performed Sanger sequencing. **Dr. Alexandre Montpetit** participated in the study design of sequencing platform. **Dr. Zhifeng Dong** and **Dr. Peter Siegel** performed IHC of TMA. **Cindy Zhang, Dr. Pedro Castelo-Branco, and Dr. Uri Tabori** performed Western Blot experiments of methylated histone. **Damien Faury** helped with the study design and DNA extraction. **Dr. Peter Lichter and Dr. Christoph Plass** participated in

the study design. **Dr. Jacek Majewski** participated in the study design, WES analysis and manuscript drafting. **Dr. Stefan M. Pfister** participated in the study design and provided clinical samples. **Dr. Nada Jabado** conceived the project, participated in the data analysis, provided leadership, managed the project and drafted the manuscript. All authors contributed to the final manuscript.

**Chapter 4: Frequent *ATRX* mutations and loss of expression in adult diffuse astrocytic tumors carrying *IDH1/IDH2* and *TP53* mutations.** Xiao-Yang Liu, Noha Gerges, Andrey Korshunov, Nesrin Sabha, Dong-Anh Khuong-Quang, Adam M Fontebasso, Adam Fleming, Djihad Hadjadj; Jeremy Schwartzenuber, Jacek Majewski, Zhifeng Dong, Peter Siegel, Steffen Albrecht, Sidney Croul, David TW Jones, Marcel Kool, Martje Tonjes, Guido Reifenberger, Damien Faury, Gelareh Zadeh, Stefan Pfister, Nada Jabado. *Acta Neuropathologica* (124(5):615-25, 2012).

**Xiaoyang Liu** participated in the study design, whole exome & Sanger sequencing analysis, and IHC of *ATRX* & *DAXX* antibody on TMA. She also performed DNA extraction, analysis of sequencing & IHC results of *ATRX*, *DAXX*, *IDH* & *TP53* for clinical samples, clinical data & statistical

analysis, and drafted the manuscript. **Noha Gerges** participated in the analysis of ATRX & DAXX IHC results on TMA. **Dr. Andrey Korshunov** provided clinical samples and performed ALT-FISH experiment. **Nesrin Sabha** provided clinical samples and performed some of the clinical data gathering. **Dr. Dong-Anh Khuong-Quang** and **Djihad Hadjadj** was involved in some of the clinical data analysis. **Adam M Fontebasso**, **Jeremy Schwartzentruber**, and **Dr. Jacek Majewski** performed WES analysis. **Dr. Adam Fleming** helped with the study design. **Dr. Zhifeng Dong** and **Dr. Peter Siegel** performed IHC staining of TMA with ATRX & DAXX antibodies. **Dr. Steffen Albrecht** performed pathological review and analysis of TMA IHC. **Dr. Sidney Croul** provided samples and performed IHC staining of p53 & IDH. **Dr. David TW Jones** and **Dr. Marcel Kool** performed the analysis of GEP results with datasets available in literature. **Dr. Martje Tonjes** performed TP53 and IDH sequencing. **Dr. Guido Reifenberger** provided samples and participated in the study design. **Damien Faury** provided help with study design and DNA extraction. **Dr. Gelareh Zadeh** and **Dr. Stefan Pfister** provided clinical samples and participated in the study design. **Dr. Nada Jabado** conceived the project, participated in the data analysis, provided leadership, managed the project and drafted the manuscript. All authors contributed to the final manuscript.

## SIGNIFICANCE OF THESIS

This thesis work is among the first efforts to study the function of YB-1 in astrocytoma genesis and to perform large-scale whole exome sequencing in pediatric glioblastomas. The results contribute significantly to the better understanding of genetic and molecular alterations in gliomas, improved subgrouping of glioblastoma, and the development of novel targeted therapies.

## Chapter 1: INTRODUCTION

This thesis focuses on the identification of genetic and molecular alterations important in the pathogenesis of pediatric and young adult gliomas. The Introduction will first give a general overview of gliomas in terms of classification, epidemiology, clinical aspects and genetic alterations. Then focus will be given to adult and pediatric glioblastoma (grade IV astrocytomas), especially the distinct molecular and genetic alterations we and others identified. The introduction will then move on to the YB-1 protein, an important target found to be upregulated in pediatric GBM. Finally, the whole exome sequencing technology as well as its applications in cancers will be discussed. This will lead to the rationale and objectives of the thesis work.

## 1.1 General Overview of Gliomas

### 1.1.1 Classification of gliomas

Gliomas, derived from glial cells, are the most common primary brain tumors in children and adults. Glial cells include astrocytes, oligodendrocytes and ependyma, which all come from neuroectoderm, as well as microglia derived from bone marrow. Being the principle cells for repair and scar formation in the brain, astrocytes have star-shaped and multipolar appearance, and act as metabolic buffers and nutrient suppliers to the neurons (Vinay Kumar 2005). Oligodendrocytes are responsible for wrapping the axons of neurons with myelin, whereas ependymal cells form the lining of the ventricular system. Microglia serve as a fixed macrophage system and the main active immune defense in the central nervous system (CNS) (Vinay Kumar 2005). Depending on the cell of origin, gliomas are classified histologically into astrocytic, oligodendroglial, oligoastrocytic and ependymal tumors (Jones, Perryman et al. 2012). For this thesis, the main focus is astrocytoma, oligodendroglioma and oligoastrocytoma, while ependymal tumors are less discussed.

According to the fourth edition of the World Health Organization (WHO) classification of tumors of the central nervous system (Louis 2007)

published in 2007, gliomas are further classified into four grades. Grade I and II are commonly referred to as low grade gliomas (LGG), while grade III and IV are known as high grade gliomas (HGG). Grade I, such as pilocytic astrocytoma (PA), refers to tumors with low proliferative potential and is likely to be cured by surgical resection alone. Grade II lesions consist of diffuse astrocytoma, pilomyxoid astrocytoma, pleomorphic astrocytoma, oligodendroglioma and oligoastrocytoma. They show low proliferation level, but are highly infiltrative, and often recur and progress to higher grade tumors. Tumors designated as grade III, including anaplastic astrocytoma, anaplastic oligodendroglioma and anaplastic oligoastrocytoma, have nuclear atypia and higher mitotic activity. Glioblastoma (GBM) are the major component of grade IV neoplasms and show beside the high proliferation index, high level of necrosis (pseudopalisading necrosis, a hallmark of GBM) and microvascular proliferation (Louis, Ohgaki et al. 2007). GBM are classified into two subtypes: primary GBM, which represent 90% of all GBM cases and indicate tumors which arise de novo and mainly affect patients older than 50 years; and secondary GBM, which progress from grade II or III astrocytoma and mainly affect younger patients (Louis 2007). In Children, GBM arise de novo and very rarely evolve from lower grade tumors. As

the main subject of study for this thesis, glioblastoma will be further discussed in section 1.2.

### **1.1.2 Epidemiology of gliomas**

Epidemiology data for gliomas in North America mainly comes from the Central Brain Tumor Registry of the United States (CBTRUS) (Dolecek, Propp et al. 2012). The incidence rate of gliomas is about 2-3 per 100,000 persons per year during childhood and increases with age in the adult setting, peaking around 75-84 years old with about 20 new cases per 100,000 person-years (Figure 1.1). All of the histological sub-types of gliomas are more prevalent in males (average annual age adjusted-incidence of 7.16 per 100,000 person) than in females (5.06 per 100,000 person), and more in white people than persons of black origin.

Astrocytoma is the most common type across age span, accounting for 76% of all gliomas, with the grade IV glioblastoma occupying 54% (Figure 1.2). The incidence rates of different grades of astrocytomas vary with age: grade I astrocytoma has the highest incidence in the 0-14 age group and decreases with age; whereas grade IV astrocytoma happens much more commonly in adult than in children less than 20 years old (Figure

1.3) (Kieran, Walker et al. 2010). Oligodendroglioma, oligoastrocytoma and ependymoma are relatively rare gliomas compared to astrocytoma, accounting for 3.3%, 6.2% and 6.7% of all gliomas, respectively.

### **1.1.3 Clinical aspects of gliomas**

The treatment and survival rate of gliomas vary largely depending on their grades. Grade I pilocytic astrocytoma has the best 5-year survival rate of 94.1% (Dolecek, Propp et al. 2012), and usually undergo surgery for maximum tumor removal (Ohgaki and Kleihues 2005). Starting from grade II gliomas, prognosis gets significantly worse due to the highly infiltrative nature of the tumors, with decreased survival rates in higher grades. The 5-year survival rates for grade II diffuse astrocytoma and grade II oligodendroglioma are 47.1% and 79.1%, respectively, both of which are generally treated with surgical intervention without radio- or chemotherapy (Ohgaki and Kleihues 2005; Dolecek, Propp et al. 2012). Grade III gliomas demonstrate a 5-year survival rate of 26% and 49%, for anaplastic astrocytoma and anaplastic oligodendroglioma, correspondingly. The most dismal 5-year survival is observed in GBM (4.7%) (Dolecek, Propp et al. 2012). The standard procedure for non-

brainstem high grade gliomas (HGG) includes surgical management for confirmation of diagnosis and attempt of maximal tumor resection, as well as adjuvant radiotherapy, although radiotherapy demonstrated more significant survival benefit in adult HGG (aHGG) compared to pediatric HGG (pHGG) (Laperriere, Zuraw et al. 2002). A large amount of clinical trials on single-agent and combination chemotherapy have been performed and are still going on for HGG. For example, an alkylating agent temozolomide, causing DNA damage and death of the tumor cells, has been used as the standard chemotherapy for newly diagnosed adult GBM, where it improved the overall progression free survival by a marginal 3-6 month, but failed to improve outcome of pHGG (Stupp, Mason et al. 2005; Stupp, Hegi et al. 2009; Cohen, Pollack et al. 2011).

#### **1.1.4 Genetic alteration in gliomas**

Grade I astrocytomas, pilocytic astrocytoma, are the most common brain tumors in children representing 23% of all pediatric brain tumors. They have limited proliferation index, and almost never progress to higher grade tumors. 15% of pilocytic astrocytomas occur in patients with inherited cancer predisposition syndrome neurofibromatosis type 1 (NF1), caused

by mutation of the *NF1* gene encoding the Ras inactivating protein neurofibromin 1 (Rodriguez, Giannini et al. 2008). The non-NF1-associated sporadic pilocytic astrocytomas show a limited number of genetic abnormalities and are in fact characterized by somatic genetic alterations affecting the mitogen-activated protein kinase (MAPK) pathway. These are mainly characterized by constitutive activation of *BRAF* due to point mutations (5-10%) or gene fusions (65%) caused by tandem duplication on chromosome 7q34 (Pfister, Janzarik et al. 2008; Jacob, Albrecht et al. 2009; Jones, Gronych et al. 2012), as well as KRAS mutations and RAF1 activating gene fusion at lower frequencies (Janzarik, Kratz et al. 2007; Forsheew, Tatevossian et al. 2009; Jones, Kocialkowski et al. 2009).

The genetic pathways leading to grade II, III and IV gliomas are summarized in Figure 1.4. *IDH1/2* mutations are prevalent in grade II glioma (>80% of cases). These mutations are detected with similar rate in grade III gliomas and secondary GBM as well, rendering them precursor mutations in gliomagenesis. The mutations invariably affect codon R132 of *IDH1* or R172 of *IDH2* (functionally analogous residue of each other), with R132H representing the most common change (Parsons, Jones et al.

2008; Yan, Parsons et al. 2009). *IDH1* encodes the enzyme isocitrate dehydrogenase, which catalyzes the production of  $\alpha$ -ketoglutarate ( $\alpha$ -KG) in the citric acid cycle, in the cytoplasm. IDH2 catalyzes the same reaction solely in the mitochondria (Kim and Liao 2012). Although the exact mechanism how *IDH1/2* mutations are involved in gliomagenesis is still being investigated, it is commonly believed that the mutant forms of IDH1/2 acquire neoenzymatic activity converting  $\alpha$ -KG to D-2-hydroxyglutarate (2-HG) (Dang, White et al. 2010). High levels of 2-HG, the neomorphic metabolite, could inhibit the activity of histone demethylases, therefore leading to altered epigenetic control and gene expression (Xu, Yang et al. 2011).

IDH1/2 mutations are strongly associated with *TP53* mutation or 1p/19q loss in grade II gliomas, depending on their histological subtypes. 60% of diffuse astrocytomas (grade II) harbor *TP53* mutations in addition to *IDH1/2* mutations, whereas about 70% of grade II oligodendrogliomas show loss of 1p/19q (Ohgaki and Kleihues 2011). In mixed oligoastrocytomas, half of the *IDH1/2* mutants also have *TP53* mutation, whereas the other half demonstrate loss of 1p/19q (Ohgaki and Kleihues 2011), with *TP53* mutations and 1p/19q deletions being mutually exclusive.

As a result, grade II gliomas are thought to develop from common precursor cells harboring *IDH1/2* mutations. Additional mutation in *TP53* favors the onset of astrocytoma phenotype; on the other hand, additional loss of 1p/19q triggers the formation of oligodendroglioma, with mixed oligoastrocytomas containing either of the two additional genetic hits (Figure 1.4).

The genetic alterations encompass a broad spectrum of different molecular pathways for high grade gliomas, which can occur either de novo or progress from precursor lower grade lesions (Louis 2007). Grade II gliomas can progress to higher grades: grade II astrocytoma to anaplastic astrocytoma or secondary GBM (grade IV), grade II oligodendroglioma to anaplastic oligodendroglioma (Louis 2007). In addition to *IDH1/2* mutation and 1p/19q loss, one third of the anaplastic oligodendrogliomas also harbor chromosome 9p deletion, which leads to homozygous loss of the tumor suppressor gene *CDKN2A*, suggesting its role in oligodendroglioma progression (Louis 2007). An intermediate stage of progression from low grade astrocytoma to GBM, anaplastic astrocytoma often harbors the previously mentioned *IDH1/2* and *TP53* mutations. They also show much higher frequency of LOH 10q (35-60%)

and 19q (46%) compared to diffuse astrocytoma, which are believed to be important for the transformation from low to high grade astrocytomas (Louis 2007).

Similar to anaplastic astrocytoma (grade III), GBM (grade IV) can also occur de novo (primary GBM) or from lower grade precursor lesions (secondary GBM), which demonstrate distinct genetic alterations between each other. Although primary and secondary GBM are histologically indistinguishable, they probably develop through different genetic pathways. *IDH1/2* mutations, *TP53* mutations, loss of heterozygosity (LOH) of 19q and LOH of 22q are more frequently observed in secondary GBM (80%, 65%, 54% and 82%, correspondingly) (Nakamura, Ishida et al. 2005; Ohgaki and Kleihues 2007; Ohgaki and Kleihues 2011); whereas LOH of 10p, amplification of epidermal growth factor receptor (EGFR) and *PTEN* mutations are more common in primary GBM (47%, 36% and 25% respectively) (Fujisawa, Reis et al. 2000; Ohgaki and Kleihues 2007) (Figure 1.4). Loss of tumor suppressor genes caused by LOH of 10q (more than 60% of cases), such as *PTEN*, is important for the progression of grade II or III astrocytoma into secondary GBM (Figure 1.4), and is observed in approximately 70% of primary GBM as well (Fujisawa, Kurrer

et al. 1999; Fujisawa, Reis et al. 2000) (Figure 4). Since GBM is the main object of study for this thesis, its molecular pathogenesis will be discussed in more details in the next section.

## 1.2 Adult and Pediatric GBM

GBM is a highly malignant type of tumor, with extremely poor clinical outcome, both in children and adults. Pediatric and adult GBM are indistinguishable in terms of their histology, both manifesting nuclear atypia, cellular pleomorphism, poor differentiation and high mitotic activity, as well as areas of necrosis and prominent vascularity (Louis 2007). Macroscopically, both tumors show poor delineation and high levels of invasion into neighboring brain, even into the contra lateral hemisphere (Louis 2007). However, accumulating evidences have pointed out that there are many differences between pediatric and adult GBM, such as their localization, genetic and molecular alterations, suggesting distinct pathways to their pathogenesis. Due to their similarity in histology and the scarcity of pediatric samples, current treatments for pediatric GBM have been mainly adapted from those for adults, and, similar to adult studies, show universal failure in all clinical trials of pediatric GBM. As a result, further understanding of the genetic and molecular alterations involved in pediatric GBM formation is mandated in order to develop more effective treatment.

### 1.2.1 Brain location of GBM in children and adults

In adult patients, GBM is the most common primary brain tumor and classified into two subtypes: primary and secondary (Louis 2007). The majority of adult GBM (>90%) occur *de novo*, without previous history of less-malignant lesions. They usually occur in older patient with a mean age of onset about 62 year old. Only about 5% of adult GBMs develop from progression of pre-existing lower grade astrocytomas (grade II diffuse astrocytoma or grade III anaplastic astrocytoma), with a mean progression time of 2-5 years and mean age of onset at 45 years (Ohgaki, Dessen et al. 2004; Ohgaki and Kleihues 2007). The most often location of adult glioblastoma is in the cortex of the cerebral hemispheres (supratentorial), whereas brain stem, cerebellum and spinal cord (infratentorial region) are relatively infrequent in this type of tumor (Louis 2007).

Unlike its adult counterpart, pediatric GBM is relatively rare, accounting for only about 3% of all pediatric CNS tumors (Das, Mehrotra et al. 2012). It occurs mainly *de novo*, and progression of lower grade astrocytoma to GBM is extremely rare in children. Infratentorial locations are more common for pediatric compared to adult GBM (Rineer, Schreiber et al.

2010). For example, brain stem localization is much more frequently found in pediatric (10%) than adult GBM (1%) (Jones, Perryman et al. 2012).

Other infratentorial tumors also take preference to pediatric patients: thalamic tumors account for about 13% of all pediatric GBM (Kramm, Butenhoff et al. 2011; Das, Mehrotra et al. 2012), and cerebellar tumors for about 5% (Das, Mehrotra et al. 2012).

### **1.2.2 Distinct genetic and molecular alterations between adult and pediatric glioblastoma**

The genetic pathways leading to the onset of pediatric and adult GBMs are quite different, in terms of the important genetic mutations, epigenetic regulations, copy number aberrations and gene expression profiles involved in their pathogenesis. Therefore, the details of their genetic and molecular alterations will be discussed separately.

#### **1.2.2.1 Adult GBM**

Genetic alterations found in adult GBM involve disruption of epigenetic regulation (such as *IDH1/2* mutations), activation of RTK/RAS/PI3K pathway to promote cell proliferation and survival, impaired p53 signalling to inhibit cell senescence and apoptosis, as well as inactivation of RB signalling to induce cell cycle progression (Figure 1.5). These aberrations and the molecular signatures allow for a better characterization and subgrouping of adult GBM, and will be discussed in this section according to their functional relevance.

#### 1.2.2.1 (I). Molecular subgrouping

In order to stratify different prognostic and treatment group of GBM patients, subgrouping based on genetic and molecular signatures is intensely studied. Efforts led by the Cancer Genome Atlas Network (TCGA) have shown great breakthroughs in the molecular classification of adult GBM, using integrated genomic approach (Verhaak, Hoadley et al. 2010). They identified four subtypes based on gene expression pattern in two hundred GBM samples: proneural, neural, classical and mesenchymal (Figure 1.6). These molecular subclasses improved previous high-grade glioma grouping into proneural, proliferative and mesenchymal subclasses (Phillips, Kharbanda et al. 2006).

Each of the four subtypes is characterized by different genetic and molecular alterations (Figure 1.6), using integrated analysis of genomic copy number data, genetic mutation data of 601 genes and gene expression signatures (Verhaak, Hoadley et al. 2010). The proneural class is characterized by *IDH1* point mutations and *PDGFRA* amplification, which are largely mutually exclusive events. *TP53* mutations also occur more frequently in this group compared to the others. Almost all the secondary GBMs have the gene expression signature of this class. The main feature for the classical subtype is the concurrent amplification of *EGFR* and homozygous deletion of *CDKN2A*, while absence of *TP53* mutations was observed. The genetic lesion characterizing the mesenchymal class is hemizygous deletion of *NF1*, leading to its decreased expression. The functions and pathway involvement of these mutations will be discussed into details in the following sections 1.2.2.1 (II), (III), (IV) and (V).

#### 1.2.2.1 (II). Epigenetic regulation

##### 1.2.2.1 (II-a) IDH1/2 mutations

IDH mutations were mostly observed in the proneural molecular subtype of adult GBM. It is the most prevalent alteration identified so far in secondary GBM (>80%), and extremely rare in primary (<5%). It has been used as a standard molecular marker to distinguish secondary GBM from primary (Ohgaki and Kleihues 2011). As discussed in section 1.1.4, mutant IDH1/2 exerts its function by creating onco-metabolites (2-HG) and inhibiting histone demethylases. These enzymes remove the methylation from histone, which is one of the key post-translational modifications of histone involved in epigenetic control (Sarkies and Sale 2012). Since the epigenetic marks such as methylation and acetylation of histone as well as methylation of DNA determine the accessibility and expression level of DNA in a certain chromatin region, they are tightly regulated (Sarkies and Sale 2012). In fact, disruption in epigenetic regulation was already shown to play an important role in the development of cancer (Dubuc, Mack et al. 2012; Sarkies and Sale 2012).

IDH mutations have been associated with a CpG island methylator phenotype (CIMP) in a subset of the proneural GBM (Noushmehr, Weisenberger et al. 2010). CIMP was first discovered in colorectal cancers (Toyota, Ahuja et al. 1999), and tumors in this subset

demonstrated concerted hypermethylation at a great number of loci within the promoter CpG islands. Promoter hypermethylation usually leads to transcriptional repression of related genes (Jones and Baylin 2007). In glioblastomas, a large proportion of the hypermethylated genes showed downregulation of their expression, with a few upregulated. Patients with this CIMP phenotype have younger age of tumor onset and better prognosis, reflecting the clinical features of IDH mutant tumors. This CIMP phenotype is also observed in lower grade gliomas (defined as glioma-CIMP, or G-CIMP) with IDH mutations (Noushmehr, Weisenberger et al. 2010; Turcan, Rohle et al. 2012), suggesting its strong association with this molecular alteration instead of histological grading.

*In vitro* experiments proved IDH mutation to be the direct cause of G-CIMP, using primary human astrocyte cell lines (Turcan, Rohle et al. 2012). Expression of mutant IDH in these cell lines produced oncometabolite 2-HG and progressively led to hypermethylation in a great number of genes, changing their expression. These genes overlap largely with IDH mutant tumor samples from patients.

The mechanism for the induction of G-CIMP by IDH mutation is believed to relate to the inhibition of histone and DNA demethylation by the neoenzymatic product 2-HG. One of the targets of 2-HG inhibition is the Jumonji-C domain histone demethylase family (JHDMS), such as JHDM1A, KDM2A and JMJD2A (Chowdhury, Yeoh et al. ; Xu, Yang et al. 2011), whose function depends on  $\alpha$ -KG.  $\alpha$ -KG is the product of wild type IDH enzymatic activity, and its binding to JHDMS are competitively inhibited by 2-HG. JHDMS can remove methyl group from methylated histone lysine residues, such as H3K4 (lysine 4 on histone 3), H3K9, H3K27, H3K79, *etc.* Methylation of histone lysine is an important component of epigenetic marks on histone tails and is involved in DNA accessibility and transcriptional regulation of a certain region. In fact, addition of 2-HG or mutant IDH1 has been shown to retain histone lysine methylation marks (Stupp, Mason et al. 2005; Xu, Yang et al. 2011). Since histone lysine methylation at critical positions like H3K4 and H3K27 is often coupled with changed chromatin accessibility and increased promoter CpG methylation (Meissner, Mikkelsen et al. 2008), its alterations probably connect with G-CIMP phenotype and abnormal gene expression. Another  $\alpha$ -KG dependent enzyme involved in demethylation is the TATA-binding protein-associated factor (TET) family of proteins, which

are 5-methylcytosine hydroxylases catalyzing oxidation of 5-methylcytosine (5mC) of methylated DNA into 5-hydroxymethylcytosine (5hmC) (Xu, Yang et al. 2011), a critical step in Tet-initiated DNA demethylation reactions (Wu and Zhang 2011). Tet binds preferentially to CpG islands and maintains the hypomethylated state of the region (Wu and Zhang 2011). In myeloid cancers, Tet impairing mutations leading to significantly decreased 5hmC levels are associated with CpG islands hypermethylation (Figueroa, Abdel-Wahab et al. 2010; Ko, Huang et al. 2010). Therefore, inhibition of Tet proteins activity in gliomas probably serves as a mediator for IDH mutations to establish the G-CIMP phenotype. The relationship between IDH mutations and G-CIMP provide an interesting example of connection between genetic mutations and epigenetic alterations.

#### 1.2.2.1 (II-b). Downregulation of DNA methyltransferase

Similar to other types of cancers, GBM is characterized (80%) by global DNA hypomethylation (decreased 5mC) with regional hypermethylation at promoter CpG islands (Dubuc, Mack et al. 2012). Decreased expression level of DNMT (de novo DNA methyltransferase responsible for DNA methylation) identified in GBM, potentially explains the global DNA hypomethylation (Cadieux, Ching et al. 2006; Fanelli, Caprodossi et al.

2008; Dubuc, Mack et al. 2012). The molecular function of this global DNA hypomethylation is not completely understood yet, but it is linked to increased genomic instability and high cellular proliferation (Nagarajan and Costello 2009), potentially contributing to tumor onset.

#### 1.2.2.1 (II-c). MGMT promoter hypermethylation

Hypermethylation of the *MGMT* (*O*<sup>6</sup>-methylguanine-DNA methyltransferase) promoter CpG island leading to decreased expression is observed in 75% of secondary and 36% of primary GBM tumors (Ohgaki and Kleihues 2007). *MGMT* is responsible for DNA cross-linking repair by removing alkylation at the *O*<sup>6</sup> position of guanine, and its promoter hypermethylation is a positive predictor for response to alkylating agent chemotherapy such as temozolomide (Hegi, Diserens et al. 2005). Under normal circumstances, *MGMT* prevent the cells from carcinogenesis caused by alkylating agents, which induce DNA damage and cell death. In the case of alkylating chemotherapy, the repair activity of *MGMT* would hinder the DNA-damaging effect and reduce the effectiveness of treatment. This is the reason why *MGMT* promoter hypermethylated patients respond better to alkylating agents.

### 1.2.2.1 (III). RTK/RAS/PI3K signalling alteration

Genetic alterations in the RTK/RAS/PI3K signalling pathway are mainly seen in primary GBM (88%) (Figure 1.5a) and are infrequent in secondary (Ohgaki and Kleihues 2007; TCGA 2008). RTKs (receptor tyrosine kinase) such as EGFR, Erbb2, PDGFRA and Met are located on the cell membrane. Upon growth factor stimulation, these RTKs can activate through signalling cascades their downstream players such as Ras, PI3K and Akt. Ras is a guanine nucleotide-binding GTPase tethered to the membrane. Upon exchange of their bound GDPs with GTPs, Ras can activate three downstream branches: PI3K, MAPK (mitogen-activated protein kinase) and cytoskeleton control pathways. They're involved in a variety of cellular functions, such as promoting proliferation, survival and motility (Weinberg 2007; TCGA 2008).

According to the comprehensive genomic characterization of 206 primary GBM tumors by TCGA using copy number, mutation, expression and DNA methylation data (TCGA 2008), EGFR activation via amplification and/or mutation is the most common RTK activated in primary GBM, accounting for 45% of all cases. EGFR amplification leads to its increased expression; while EGFR mutations often happen in addition to amplification and give

rise to constitutively active tyrosine kinase activity such as Variant III (EGFRvIII) with deletions in the extracellular domain (Ekstrand, Sugawa et al. 1992; Nishikawa, Ji et al. 1994; Batra, Castelino-Prabhu et al. 1995; Huang, Nagane et al. 1997; Lee, Vivanco et al. 2006). Mutation of *ERBB2*, amplification of *PDGFRA* and *MET* in primary adult GBM showed lower frequencies of 8%, 13% and 4%, respectively.

PI3K (phosphoinositide 3-kinase) is responsible for converting phosphatidylinositol (4,5)-bisphosphate (PIP<sub>2</sub>) into phosphatidylinositol (3,4,5)-triphosphate (PIP<sub>3</sub>). PIP<sub>3</sub> serves as a docking site for Akt to get activated at the cell membrane. PI3K mutations involve both *PIK3CA*, which encodes the catalytically active subunit p110 $\alpha$ , and *PIK3R1*, which encodes the inhibitory regulatory subunit p85 $\alpha$ . In GBM, mutations are observed in either subunit in a mutually exclusive manner (TCGA 2008) and are thought to disrupt the binding and inhibitory effect of p85 $\alpha$  on p110 $\alpha$ . Overall, 15% of the TCGA analyzed GBMs showed PI3K mutations.

PTEN (phosphatase and tensin homolog) is a phosphatase removing phosphate from PIP<sub>3</sub> and converting it to PIP<sub>2</sub>. Therefore, PTEN

functionally counteract the effect of PI3K and inhibit the PI3K pathway. It is considered a tumor suppressor, and its loss of function or expression is commonly seen in cancer (Weinberg 2007). Mutation or homozygous deletion of *PTEN* accounts for 36% of primary adult GBM, which leads to abrogation of its inhibitory effect on PI3K pathway.

Nf1 (neurofibromin 1) inhibits Ras activity by hydrolysing Ras-bound GTP; therefore it is a negative regulator of the RTK/RAS pathway. Inherited *NF1* mutations in family have been linked to neurofibromatosis type 1, characterized by outgrowth of tumors along the nerves and skin pigmentation. In GBM, mutation or homozygous deletion of *NF1* is detected in 18%, contributing to the increased RTK/RAS/PI3K pathway activity. *RAS* mutation and *AKT* amplification are rare events, each identified in about 2% of GBMs (TCGA 2008).

#### 1.2.2.1 (IV). P53 pathway alteration

The p53 signaling pathway is critical in inducing cell senescence and apoptosis, and is found to be altered in 87% of GBMs (TCGA 2008) (Figure 1.5b). P53 is one of the most commonly seen tumor suppressor genes in cancer. The expression of p53 is inhibited by Mdm2 (murine

double minute 2) through ubiquitin targeted degradation, and its activity is suppressed by Mdm4 via binding of the transcriptional activation domain (Weinberg 2007). Mdm2 is in turn negatively regulated by p14<sup>ARF</sup>, encoded by the *CDKN2A* gene. An alternative splicing variant produced from *CDKN2A* gene is p16<sup>INK4A</sup>, which is involved in the RB pathway. Oncogene activation alone will usually trigger cell senescence and apoptosis, if the p53 pathway function is intact. However, cancer cells employ many ways to escape this safeguard in order to achieve immortality. In the case of GBM, homozygous deletion or mutation of *CDKN2A* gene is found in about 50% of patients; Mdm2 and Mdm4 are amplified in 14% and 7% of tumors, correspondingly. *TP53* mutation is found in 35% of primary GBM and more frequently in 65% of secondary GBM (Ohgaki and Kleihues 2007; TCGA 2008; Ohgaki and Kleihues 2011).

#### 1.2.2.1 (V). Rb pathway alteration

Altered in 78% of primary adult GBM (Figure 1.5c), RB signalling regulates cell cycle progression. When cells are not dividing, Rb protein (the retinoblastoma tumor suppressor protein) binds to and inhibits E2F transcription factor, thus preventing cell transition from G1 to S phase.

Upon growth stimuli, Rb protein can be phosphorylated by the complex of cyclin dependent kinase CDK4/6 and cyclin D2 (encoded by the *CCND2* gene) and release E2F to promote G1/S transition. The formation of complex between cyclin D2 and CDK4/6 is due to the relief of its inhibition by the INK4 protein family, such as p16<sup>INK4A</sup> (encoded by *CDKN2A*), p15<sup>INK4B</sup> (encoded by *CDKN2B*) and p18<sup>INK4C</sup> (encoded by *CDKN2C*) (Weinberg 2007). In primary GBM, homozygous deletion or mutation of *CDKN2A* is found in 52% of cases, while homozygous deletions of *CDKN2B* and *CDKN2C* are found in 47% and 2%, respectively. 18% of the GBM tumors have amplification of *CDK4*, and 11% have homozygous deletion or mutation of the *RB1* gene.

#### 1.2.2.2 Pediatric GBM

Due to the scarcity of the pediatric GBM tissues, their molecular and genetic alterations were less well studied than adult GBM. Our group is among the first ones to focus on pediatric GBM specifically. We and others have demonstrated that great differences exist between pediatric and adult GBM in terms of their genetic aberrations and molecular signatures (Wong, Tsang et al. 2006; Faury, Nantel et al. 2007; Haque,

Faury et al. 2007; Bax, Mackay et al. 2010; Paugh, Qu et al. 2010; Qu, Jacob et al. 2010; Schiffman, Hodgson et al. 2010; Barrow, Adamowicz-Brice et al. 2011). However the molecular pathogenesis of pediatric GBM is not completely understood and further studies are warranted to identify therapeutic targets and improve treatment. Since pediatric GBM is much rarer than its adult counterpart, international multi-institutional collaboration is necessary to obtain relatively large number of samples to study.

#### 1.2.2.2 (I). Molecular subclassification of pediatric GBM

Molecular subclassification of pediatric GBMs has not been as clear as adult GBM. According to the gene expression profiles of 32 pediatric GBM samples, our group suggested in 2007 that at least two subsets of pediatric GBM exist: one associated with Ras/Akt pathway activation, the other one without (Faury, Nantel et al. 2007). The subset with Ras/Akt activation showed upregulation of expression in genes related to proliferation and neural stem-cell phenotype, similar to the proliferative and mesenchymal subgroups of adult GBM proposed by Phillips *et al* (Phillips, Kharbanda et al. 2006). These patients showed much worse overall survival compared to the group without Ras/Akt activation.

In 2010, using integrated molecular genetic profiling, Paugh and colleague showed three major subclasses (HC1, HC2 and HC3) of pediatric high grade glioma (Paugh, Qu et al. 2010). Cell cycle regulation genes are most significantly overexpressed in HC1; upregulation of genes involved in neuronal differentiation and extracellular matrix-receptor interactions are the key features for HC2 and HC3, respectively. The HC1, HC2 and HC3 groups largely recapitulate the subgroups of proliferative (Prolif), proneural (PN) and mesenchymal (Mes) subtypes in adult GBM (Phillips, Kharbanda et al. 2006). Amplification of PDGFRA and/or PDGFB was found in 88% of HC1 group, explaining its highly proliferative profile.

#### 1.2.2.2 (II). Distinct gene expression from adult GBM

Although the molecular subclassification of pediatric GBM reflects the adult GBM subgrouping, their gene expression signatures varied from each other. Our group was the first to show that pediatric and adult GBM cluster separately based on their gene expression profiles, and this holds true even when the corresponding subgroups between pediatric and adult GBMs are compared (Faury, Nantel et al. 2007; Haque, Faury et al. 2007). An independent cohort analyzed by Paugh and colleague also

demonstrated significant separation between pediatric and adult GBM in terms of gene expression profile, with minor overlap observed. These results suggest that pediatric and adult GBM are probably biologically distinct entities.

The differential gene expression between pediatric and adult GBM shed light into the understanding of pediatric GBM pathogenesis. One of the highly differentially expressed proteins between pediatric and adult GBM is YB-1 (Y box binding protein 1, encoded by the *YBX1* gene) (Faury, Nantel et al. 2007). YB-1 expression in pediatric GBM increased by 3.8-fold compared to normal brain, which is not observed in adult GBM. Overexpression of YB-1 has been observed in various types of cancers, and its nuclear localization seems to indicate poor prognosis. Given its pleiotropic functions important in oncogenesis, such as transcriptional activation of EGFR and inhibitory interaction with p53 (Kohno, Izumi et al. 2003; Lasham, Moloney et al. 2003; Homer, Knight et al. 2005; Wu, Lee et al. 2006; Evdokimova, Tognon et al. 2009), YB-1 is being extensively studied for its molecular mechanism in driving cancer. In order to improve pediatric GBM treatment, YB-1 is a good candidate to study as a therapeutic target. A more comprehensive introduction to currently known

YB-1 functions and its involvement in cancer will be provided in Section 1.3.

#### 1.2.2.2 (III). Different copy number alterations from adult GBM

Pediatric GBM in general harbor fewer and different copy number changes at the genomic level compared to adult (Wong, Tsang et al. 2006; Bax, Mackay et al. 2010; Paugh, Qu et al. 2010; Qu, Jacob et al. 2010; Barrow, Adamowicz-Brice et al. 2011; Jones, Perryman et al. 2012). Broad low-magnitude amplification of 1q is a chromosomal imbalance detected mainly in pediatric GBM (20-29%) and rare in adult. A lot of the commonly seen copy number aberrations in adult GBM such as 10q loss (70%) are much less frequent in pediatric (27%). In terms of focal copy number aberrations, genes involved in the three critical pathways (RTK/RAS/PI3K, p53 and RB pathways) for the genetic alterations in adult GBM are less affected in pediatric. For example, *EGFR* amplification is a hallmark of adult primary GBM, but is very rare in pediatric GBM. PTEN deletion is also less seen in pediatric setting. Focal homozygous deletion of *CDKN2A/CDKN2B* was found in about 20% of pediatric GBM, also much less common than adult. On the other hand, pediatric GBM exhibit

amplification of *PDGFRA* more often (~15%) than adult (~10%), another receptor tyrosine kinase involved in the RTK/RAS/PI3K pathway.

Prior to the work elaborated by this thesis, *PDGFRA* amplification had been the most common genetic amplification found in pediatric GBM, whereas all the other focally amplified genes such as *CCND2*, *PDGFB*, *MET*, *CMYC* and *MDM4* are less significant with a frequency of only 1%-4%.

#### 1.2.2.2 (IV). Genetic mutations in pediatric GBM

Unlike adult GBM, no large scale genetic mutation analysis had been performed before this thesis work. Due to its high prevalence in secondary GBM, *IDH1* status at codon R132 has been examined in pediatric GBM, and absence or extreme rarity of mutation was identified (Yan, Parsons et al. 2009; Paugh, Qu et al. 2010). EGFRvIII deletion mutation is also very rare, observed only in 1 out of 20 pediatric GBM samples (5%) (Bax, Gaspar et al. 2009). Oncogenic activating missense mutation of *BRAF* at codon V600E was seen in 2 samples of an 11 pediatric GBM cohort (18%), and occurred in association with homozygous deletion of *CDKN2A* (Schiffman, Hodgson et al. 2010). B-Raf is a member of the Raf kinase family, which is part of the MAPK pathway downstream of receptor

tyrosine kinase. BRAF activating mutation seems to be mutually exclusive from PDGFR amplification, and it probably grants the tumor cells a highly proliferative and malignant phenotype when the p14<sup>ARF</sup> and p16<sup>INK4A</sup> encoded by CDKN2A are lost. Overall, the lack of information regarding genetic mutations in pediatric GBM provides area of further research in order to identify the driving force of its tumorigenesis.

### 1.2.3 Targeted therapy for glioblastoma

Although aggressive radiotherapy combined with chemotherapy is the standard treatment for glioblastoma after maximal surgical removal, patient prognoses remain poor. Therefore, the advancement in the molecular and genetic pathogenesis of glioblastoma is being translated into development of novel targeted therapies. Since amplification of *EGFR* and *PDGFRA* is common in adult and pediatric GBM correspondingly, clinical trials inhibiting EGFR, PDGFR (Imatinib) and PI3K pathway (XL765 Sanofi-Aventis) are currently being carried out (Cage, Mueller et al. 2012). An inhibitor of BRAF<sup>V600E</sup> (PLX4032), which has demonstrated efficacy in melanoma patients, is also considered for clinical trial for pediatric GBM (Flaherty, Puzanov et al. 2010). Due to the highly vascular

nature of glioblastoma, anti-angiogenic drugs such as bevacizumab (targeting VEGF) and cediranib (targeting VEGFR) are also being tested in clinical trials, in combination with chemotherapy (Jones, Perryman et al. 2012). Decision on the application of a certain targeted therapy should depend on the genetic and molecular background of the patient, due to the heterogeneity of GBM.

### 1.3. Y-box binding protein 1

Analysis of genes differentially regulated in adult, pediatric GBM and normal brain from our group identified Y-box binding protein 1 (YB-1) as one of the most highly upregulated genes in children (Faury, Nantel et al. 2007). It is a member of the cold-shock domain (CSD) protein superfamily and is a transcriptional and translational regulator (Sakura, Maekawa et al. 1988). The function of YB-1 encompasses a broad spectrum of cellular processes including DNA repair, cell proliferation and drug resistance (Kohno, Izumi et al. 2003). Increased YB-1 expression has been detected in a wide range of human cancers (Eliseeva, Kim et al. 2011), but the exact mechanism how disruption of normal YB-1 level may contribute to oncogenesis remains unclear.

#### 1.3.2. YB-1 structure and expression

YB-1 protein was first identified as a protein binding to the Y box in the promoter of the major histocompatibility complex (MHC) class II (Didier, Schiffenbauer et al. 1988) gene and to the enhancer region of the EGFR gene (Sakura, Maekawa et al. 1988). It is encoded by the gene *YBX1* and consists of an alanine/proline-rich domain (A/P domain) at the N-terminus,

a cold shock domain (CSD) and a C-terminal domain (CTD) containing both a nuclear localization signal (NLS) and a cytoplasmic retention signal (CRS) (Figure 1.7). The CTD and CSD domains are responsible for YB-1 binding to nucleic acids, whereas all the three domains are involved in binding to its interacting proteins (Bouvet, Matsumoto et al. 1995; Matsumoto, Meric et al. 1996; Manival, Ghisolfi-Nieto et al. 2001; Eliseeva, Kim et al. 2011). YB-1 protein is highly conserved throughout evolution, with 96% identity to its mouse homologue.

The level of YB-1 expression varies with age and tissue type. It is highly expressed during embryogenesis; however, its level decreases after birth in brain, heart and skeletal muscles. High expression is maintained after birth in tissues with abundant proliferative activity, such as testicles, liver and spleen (Miwa, Higuchi et al. 2006). Under normal physiological condition the majority of YB-1 expression is in the cytoplasm, but it can also locate into the nucleus when cells encounter G1/S transition (Jurcrott, Bergmann et al. 2003) or stress (DNA damaging agents, UV irradiation, oxidative stress, hyperthermia, etc) (Koike, Uchiumi et al. 1997; Stein, Jurcrott et al. 2001; Das, Chattopadhyay et al. 2007). The expression of YB-1 is regulated at the transcriptional level: Twist and c-

Myc are two upstream transcriptional activators of *YBX1*, through binding to the E-box of its gene promoter (Uramoto, Izumi et al. 2002; Shiota, Izumi et al. 2008).

### 1.3.2. Function of YB-1

YB-1 functions in a wide variety of cellular processes, depending on its localization (Figure 1.8). In the cytoplasm, YB-1 regulates mRNA translation and stability. At high level of YB-1, it binds to the cap structure of mRNPs in the cytoplasm and displaces the eIF4F translation initiation factor complex via its c-terminal domain and cold shock domain. The ability of YB-1 to bind to the 5' cap structure of mRNP is eliminated by the phosphorylation of YB-1 by Akt at amino acid Ser102 when PI3K pathway is activated (Evdokimova, Ruzanov et al. 2006). In this way, unphosphorylated YB-1 inhibits translation initiation and thus decreasing global protein synthesis (Evdokimova, Ruzanov et al. 2001; Bader, Felts et al. 2003; Nekrasov, Ivshina et al. 2003). However, minimal amount of YB-1 is necessary to initiate translation, possibly due to its ability to unwind the mRNA secondary structure (Evdokimova, Kovrigina et al. 1998; Pisarev, Skabkin et al. 2002). High expression of YB-1 in the cytoplasm can also bind along the length of mRNA and form multimer

complexes of 20 YB-1 monomers, rendering the mRNA inaccessible to translation initiation factors and exonucleases (Skabkin, Kiselyova et al. 2004). Therefore, high level of YB-1 in the cytoplasm inhibits translation and protects the mRNAs from degradation.

In the nucleus, YB-1 performs its function in transcriptional regulation, DNA repair and pre-mRNA splicing. It is a transcriptional activator for the genes *EGFR*, *MET*, *PIK3CA*, *PDGFB*, *MDR1* (multidrug resistance) and Cyclin A & B1, and a transcriptional repressor for p53 and p21 (cyclin dependent kinase inhibitor 1A) (Ohga, Koike et al. 1996; Okamoto, Izumi et al. 2000; Stein, Jurchott et al. 2001; Stenina, Shaneyfelt et al. 2001; Jurchott, Bergmann et al. 2003; Lasham, Moloney et al. 2003; Stratford, Habibi et al. 2007; Astanehe, Finkbeiner et al. 2009; Finkbeiner, Astanehe et al. 2009; Sengupta, Mantha et al. 2011). Therefore, YB-1 promotes cell growth, survival and proliferation by upregulating the RTK/PI3K pathway and cell cycle progression while inhibiting apoptosis, at the transcriptional level. All of these are key processes disrupted in glioblastoma via genetic and molecular alterations, which suggests YB-1 as an important factor in the tumorigenesis of glioblastoma. The upregulation of *MDR1* transcription

by YB-1 helps explain resistance of tumor cells to chemotherapeutic agents.

The DNA repair activity of YB-1 involves two mechanisms. First, YB-1 demonstrates intrinsic exonuclease and endonuclease activity and can bind to cisplatin (cross-linking agent in chemotherapy) damaged or mismatched DNA (Ise, Nagatani et al. 1999; Izumi, Imamura et al. 2001; Gaudreault, Guay et al. 2004; Guay, Evoy et al. 2008). Second, YB-1 can interact with other proteins involved in DNA repair and regulate their activity, such as PCNA (Ise, Nagatani et al. 1999), p53 (Okamoto, Izumi et al. 2000), DNA ligase III $\alpha$  (Das, Chattopadhyay et al. 2007), MSH2(Gaudreault, Guay et al. 2004). In mouse embryonic stem cells with one allele of YB-1 knocked out, increased sensitivity to cisplatin was observed (Shibahara, Uchiumi et al. 2004). Therefore, YB-1 plays an important role in maintaining DNA integrity, thus contributing to the resistance to chemotherapy and radiotherapy.

The functions of YB-1 in the nucleus and cytoplasm are summarized in Figure 1.8. Overall, through its activity in translational and transcriptional

control, as well as DNA repair, YB-1 regulates many important cellular processes such as proliferation, drug resistance and stress response.

### 1.3.3. Regulation of YB-1 intracellular localization

YB-1 is mainly expressed in the cytoplasm, bound to mRNPs. When cells encounter stress response (UV irradiation, anticancer drugs, oxidative stress), G1/S transition or growth factor stimuli, YB-1 can translocate into the nucleus (Koike, Uchiumi et al. 1997; Stein, Jurchott et al. 2001; Jurchott, Bergmann et al. 2003; Das, Chattopadhyay et al. 2007) and regulate transcription of pro-growth genes (EGFR, MET, PIK3CA, PDGFB), cell cycle control genes (cyclin A & B1, p53, p21) and multidrug resistance gene (MDR1), as well as performing its DNA repair function (Figure 1.8). It is believed that localization of YB-1 into the nucleus requires functional p53 (Zhang, Homer et al. 2003; Homer, Knight et al. 2005). However, it is still debatable whether YB-1 nuclear translocation requires its phosphorylation. Some groups showed Akt mediated YB-1 phosphorylation at amino acid Ser102 (S102) led to the nuclear localization of YB-1 and induced breast cancer cell growth (Sutherland, Kucab et al. 2005) and ovarian cancer cell malignant characteristics (Basaki, Hosoi et al. 2007). However, other groups have

demonstrated that nuclear localization of YB-1 does not depend on its phosphorylation at Ser102 by Akt (Evdokimova, Ruzanov et al. 2006; Bader and Vogt 2008).

Proteolytic cleavage of YB-1 by the 20S proteasome can also trigger the nuclear localization of YB-1 (Stenina, Shaneyfelt et al. 2001; Sorokin, Selyutina et al. 2005). When cells encounter DNA-damage stress, the C-terminal fragment of 105 amino acids containing the cytoplasmic retention signal (CRS) is cleaved off by the 20S proteasome. The N-terminal truncated cleavage product has only the nucleus localization signal (NLS) without CRS and is found to accumulate in the nuclei (Sorokin, Selyutina et al. 2005). Similar process happens when endothelial cells are stimulated with thrombin: the full-length YB-1 is cleaved into a truncated form containing amino acids 1 to 219 without the CRS, which can go into the nucleus and activate the transcription of *PDGFB* (Stenina, Shaneyfelt et al. 2001).

The exact mechanism of YB-1 intracellular localization regulation is still not completely clear, and further research is needed for a better understanding.

#### 1.3.4. Mouse model of YB-1

Up to now, two groups have studied YB-1 function during embryogenesis using YB-1 knocked out mice (Lu, Books et al. 2005; Uchiumi, Fotovati et al. 2006), both showing embryonic lethality and developmental abnormality in YB-1<sup>-/-</sup> mice (after day 13.5 or 10.5). Normal mouse embryos expressed YB-1 ubiquitously throughout the whole body, with especially high level in the brain. YB-1 knocked out mice demonstrated severe growth retardation, multi-organ hypoplasia, exencephaly and neural tube closure defect, suggesting its critical role in embryonic development, especially brain development. One group showed reduced response to oxidative, DNA-damaging and oncogene-induced stress in YB-1<sup>-/-</sup> mouse embryonic fibroblasts (Lu, Books et al. 2005); while the other group showed reduced cell growth and proliferation, as well as increased contact inhibition in these cells (Uchiumi, Fotovati et al. 2006). These results proved the important role of YB-1 *in vivo*, in compliance with its basic cellular functions described previously.

### 1.3.5. YB-1 involvement in cancer

Overexpression of YB-1 has been discovered in many types of cancers, such as breast, ovarian, skin, lung, prostate, liver, hematopoietic cancers, *etc* (Eliseeva, Kim et al. 2011). Although the association is not always identified, high level of nuclear YB-1 expression seems to be linked to a more malignant, aggressive and drug resistant phenotype of the tumors, as well as adverse patient prognoses (Eliseeva, Kim et al. 2011). A great number of data suggest that YB-1 increases cell proliferation and protect them from apoptosis (Eliseeva, Kim et al. 2011). Overexpression of YB-1 in the mammary gland of transgenic mice was able to induce breast carcinomas (Bergmann, Royer-Pokora et al. 2005). However, recent evidence has pointed out that YB-1 involvement in cancer may have a double face. In 2009, Evdokimova and colleagues discovered that increased YB-1 level actually led to reduced proliferation rates in breast cancer cell lines (Evdokimova, Tognon et al. 2009). Meanwhile, these cells underwent epithelial-to-mesenchymal transition (EMT), a process essential for epithelial cancers to become metastatic. They showed higher motility and invasiveness, as well as increased expression of mesenchymal markers such as N-cadherin and Vimentin and loss of epithelial marker E-cadherin. The decreased cell proliferation is due to the

inhibitory activity of YB-1 on cap-dependent translation of mRNAs, many of which are pro-growth transcripts, such as cyclin B1, D1, and D3. On the other hand, YB-1 promotes cap-independent IRES mediated translation of Snail1 and Twist, which are known EMT inducers. Therefore, the translational effect of cytoplasmic YB-1 serves to promote cell invasive phenotype while limiting their proliferative ability (Evdokimova, Tognon et al. 2009).

In summary, the role of YB-1 in the development of cancer is complicated due to its important functions in multiple cellular processes. YB-1 located in the nucleus turns on cell proliferation through transcriptional regulation of genes involved in RTK/PI3K signalling, apoptosis and cell cycle control, whereas cytoplasmic YB-1 inhibits cell growth and promotes cell motility via translational regulation. Further study is warranted in order to better understand the involvement of YB-1 in cancer, specifically in GBM in our case.

## **1.4. Introduction to whole exome sequencing**

Whole exome sequencing (WES), a type of target-enriched next-generation sequencing (NGS), is an important method used in this thesis research. Its application in cancer research has recently been spreading rapidly and widely. Many genetic variations believed to have causative relationship with diseases have been discovered by this method. Since WES is relatively new, the principle and applications of WES will be discussed in this section.

### **1.4.1. Overview of next-generation sequencing**

Since the invention of Sanger sequencing in 1977, researchers have been relying on it to identify mutations in genes and even whole genomes (Sanger, Air et al. 1977; Sanger, Nicklen et al. 1977). Starting from 2005, the scientific community has been witnessing an evolutionary transition from the conventional Sanger sequencing into the next-generation sequencing (NGS) technology, which allows for high-throughput massively parallel DNA sequencing at the genomic level. Shortly after the NGS technology was made commercially available, more than 100 research publications were made within two years between 2006 and 2008, employing this method (Schuster 2008). Up to now, it has become a

widely used tool to understand genetic variants involved in monogenic Mendelian disorder, complex trait disorders and even cancers (Majewski, Schwartzentruber et al. 2011), thanks to its fast speed and affordable cost.

#### **1.4.2. Whole exome sequencing technology**

Whole exome sequencing (WES) is a type of NGS that targets the protein coding sequences, i.e. the exons. Instead of sequencing the whole genome in the case of whole genome sequencing (WGS), WES captures the complete exomic sequences, which constitutes 1% of the human genome (about 30 mega bases in length (Ng, Turner et al. 2009)). The protein coding sequences have been estimated to account for about 85% of the mutations causing diseases (Choi, Scholl et al. 2009). Given the same cost, data size and storage capacity, more coverage in sequence depth can be achieved by WES compared to WGS (Majewski, Schwartzentruber et al. 2011). Exomic sequence enrichment capture kits are readily available from commercial companies, which facilitate the WES procedure.

The process of WES contains the preparation of genomic library, coding sequence enrichment, sequencing of DNA and data analysis. The general

steps involved WES are outlined in Figure 1.9 (Haas, Katus et al. 2011; Ku, Cooper et al. 2012). During genomic library preparation, genomic DNA is fragmented and ligated with adaptors at both ends. The exome sequence enrichment is performed using hybridization of fragmented DNA with oligonucleotide probes in commercial kits, either in an array-based platform or in solution. Hybridized DNA fragments are then eluted, PCR amplified and sequenced. Analysis of the WES results includes alignment of reads with reference sequence, variant calling and filtering for important candidates, which require a great amount of bioinformatics tools.

#### 1.4.3. Application of WES in cancer

Important mutations have been identified using WES in different types of cancers. Somatic mutations in DNMT3A, which encodes a DNA methyltransferase, were discovered in 20.5% of acute myeloid leukemia (AML) (Yan, Xu et al. 2011). It was associated with poor patient prognosis and older age of disease onset. In familial pancreatic cancer, *PALB2* was identified as a susceptibility gene using unbiased sequencing of whole exomes. Three different germ line protein-truncating mutations of *PALB2* were considered to play a critical role in this disease (Jones, Hruban et al. 2009). A study on diffuse large B-cell lymphoma showed recurrent somatic

mutations affecting *EZH2* (a histone 3 methyltransferase) in 21.7% of cases (Morin, Johnson et al. 2010). All of these findings demonstrate that WES method is highly effective in better understanding of the pathogenesis and providing potential biomarkers for cancers.

Although there is great advantage to apply WES in the study of cancer genetics, attention is still needed on several aspects. First of all, the NGS technology tends to have more base-calling errors compared to Sanger sequencing; therefore mutation findings from NGS platform are usually validated using the conventional Sanger sequencing (Koboldt, Ding et al. 2010; Majewski, Schwartzentruber et al. 2011). Second, the surgically removed tumor samples used for DNA extraction should contain mainly tumor cells, with minimal contamination from peripheral tissues, in order to increase the chance of detecting tumor related mutations. Third, in order to identify somatic mutations important for cancer cell development, we need to compare tumor DNA sequence with its paired normal DNA, such as the blood DNA from the same patient. Finally, WES data provide the first line of discovery for important mutations, and it is often necessary to carry out functional assays for the findings, which will better characterize the mutations and prove its role in oncogenesis.

In summary, WES provides us with an excellent tool to understand the pathogenesis of cancer. Given its efficiency in identifying tumor related mutations in other types of cancers, we believe it will be very helpful to characterize genetic alterations in pediatric glioblastoma.

## 1.5. Rationale and objectives of study

With the development of integrated genomic analysis, the understanding of molecular and genetic pathogenesis of gliomas has been greatly advanced. *IDH1/2* mutations were shown to be the most prevalent genetic lesion in grade II and grade III adult gliomas, as well as secondary adult glioblastomas (grade IV). In addition, the molecular pathways (RTK/RAS/PI3K, p53, RB pathways) important in the development of adult glioblastoma have also been characterized, with major factors in these pathways frequently disrupted. Integrated information on copy number, genetic mutations and expression profile at the genomic level even enabled molecular subgrouping of adult GBM, which helps in prediction of clinical outcome and stratification of treatment.

However, insufficient information is known for pediatric glioblastoma for improvement of therapeutic plan. Our group and others previously demonstrated that pediatric and adult GBM harbor distinct genetic and molecular alterations, including differences in copy number aberrations, gene expression profiles and genetic mutations. This is probably the reason why treatments adapted from adult clinical trial universally failed in pediatric GBM. Clinical outcome of pediatric GBM patients are still quite

poor, and there is an urgent need to improve treatment and identify novel therapeutic targets.

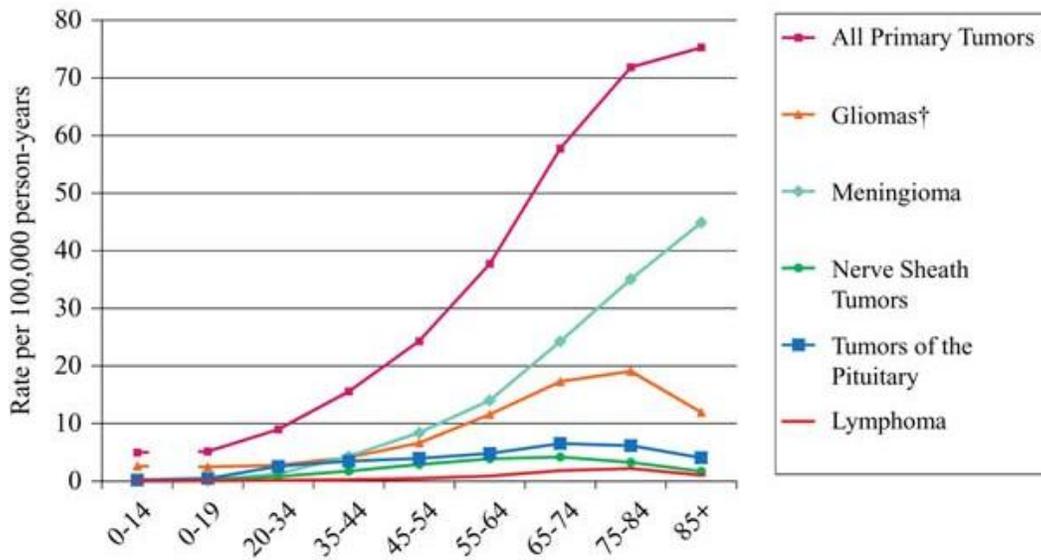
**The first objective of this study is to understand the role of YB-1 in astrocytoma formation and progression.** We previously identified YB-1 protein to be upregulated in 75% of pediatric GBM samples. This protein is overexpressed and extensively involved in a wide variety of cancers, and it regulates the three major pathways disrupted in adult glioblastoma genesis (RTK/PI3K, p53 pathway and cell cycle control). The role of YB-1 in other types of cancers such as breast cancer seems to be complicated and depends on its sub-cellular localization. Therefore further research is needed to better characterize the molecular mechanism how YB-1 is involved in glioblastoma genesis, before targeting it for therapy. To address this, we used primary GBM cell lines and primary immortalized human astrocytes to investigate the sub-cellular localization and function of YB-1 following stable knock-down or ectopic expression of this factor. We also assessed the tumorigenicity of YB-1 overexpressing cells in immune-compromised mice. Finally, we analyzed the expression levels and correlation of YB-1 and EGFR in a number of patient samples, in order to validate our *in vitro* findings.

**The second objective of this thesis is to explore for novel genetic alterations important in pediatric GBM.** Although large scale DNA copy number abnormalities detection (by single nucleotide polymorphism array, or SNP array) and gene expression profiling have been performed on pediatric GBM, not much has been done at the genetic mutation level. With the advancement of next-generation sequencing, the whole exome of tumor cells can be sequenced by WES, which allows us to identify all the somatic mutations in a given pediatric GBM sample. In this study, we used a comprehensive approach integrating WES data, gene expression profiles and copy number alterations, in an effort to identify the possible driver mutations of pediatric GBM.

## Figures

**Figure 1.1: Incidence rates of primary brain and CNS tumors for different age groups, from CBTRUS statistical report of data from 2005-2009.**

The incidence rate of gliomas (orange line) is low (about 2-3 per 100,000 persons per year) during childhood and increases with age in the adult setting, peaking around 75-84 years old with about 20 new cases per 100,000 person-years. The incidence rate of gliomas drops after 85 year old. Adapted from Dolecek TA et al., Neuro-Oncology, 2012.



**Figure 1.2: Percentage composition of different histological types of primary brain and CNS gliomas (n=90828).**

Astrocytoma is the most common type of primary brain and CNS gliomas across age span, accounting for 76% of all gliomas, with the grade IV glioblastoma occupying 54%. Adapted from Dolecek TA et al., Neuro-Oncology, 2012.

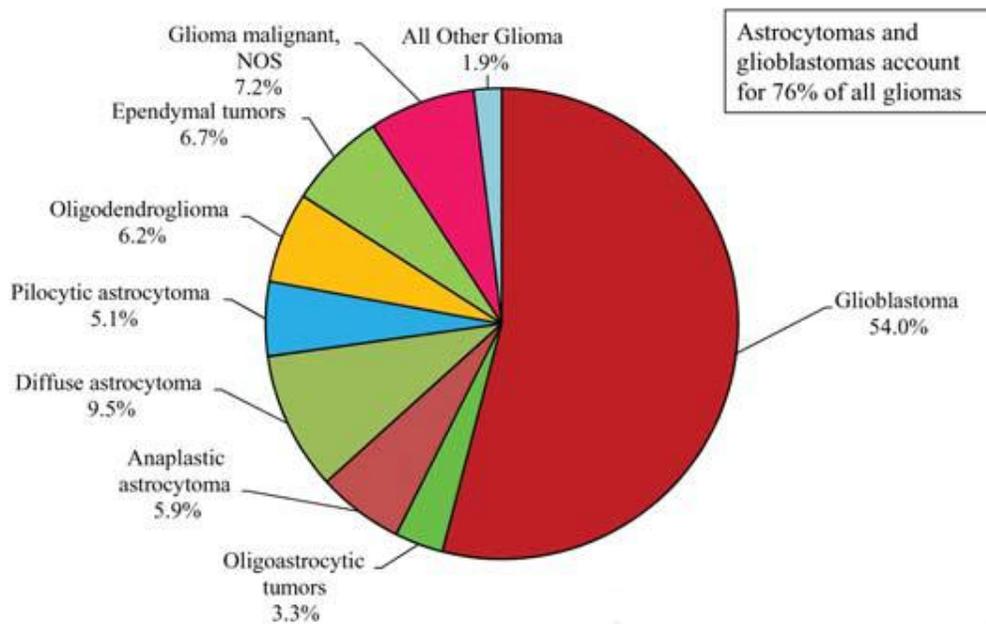
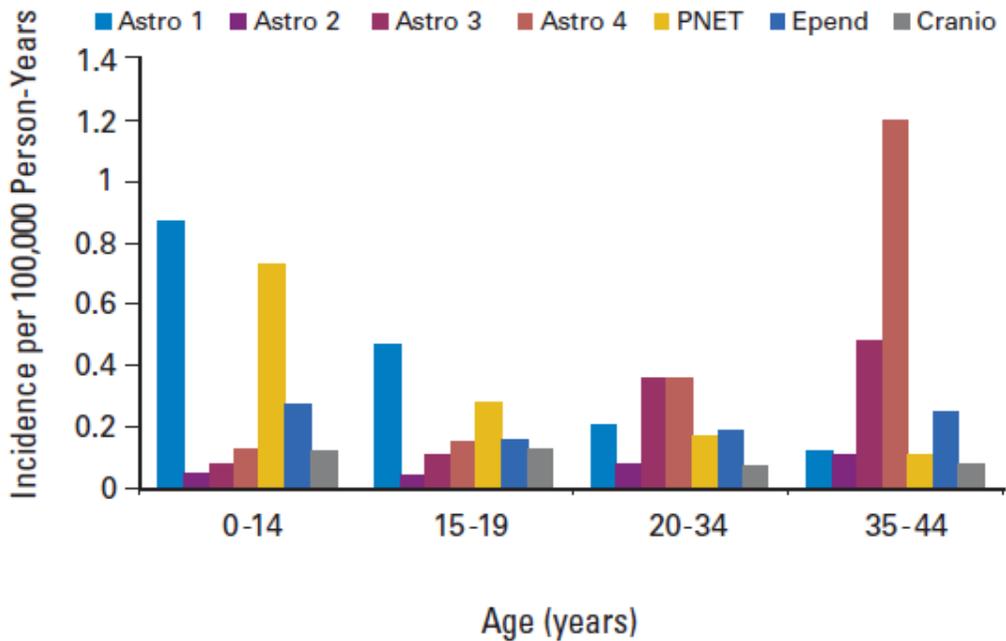


Figure 1.3: Incidence rates of grade I, II, III and IV astrocytomas (Astro), primitive neuroectodermal tumors (PNETs), ependymoma (Epend) and craniopharyngioma (Cranio). Data are collected from CBTRUS from 1997 to 2001.

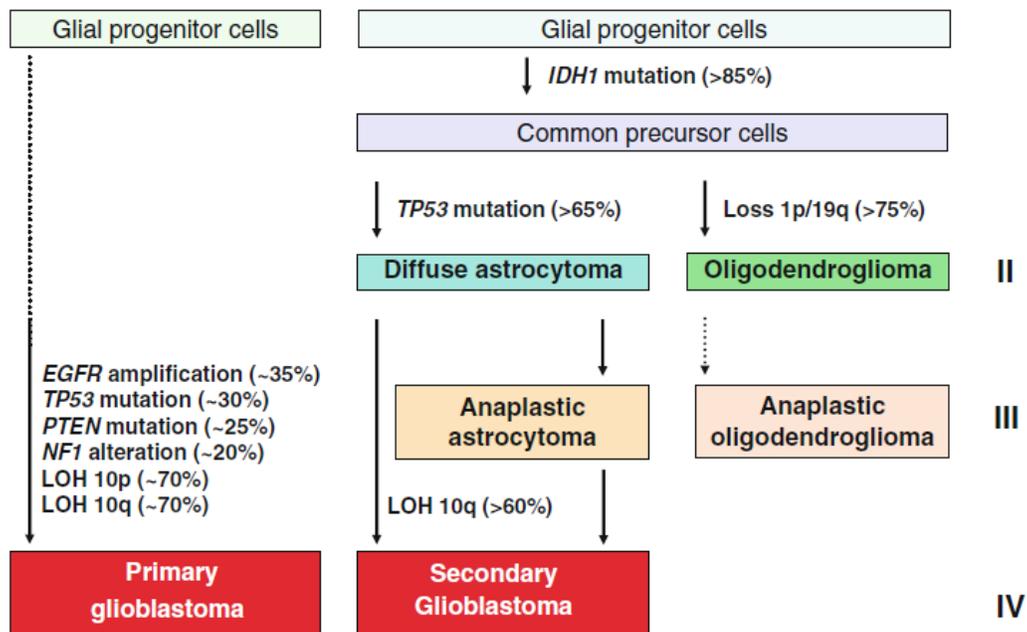
The incidence rates of different grades of astrocytomas vary with age: grade I astrocytoma has the highest incidence in the 0-14 age group and decreases with age; whereas grade II, III and IV astrocytomas are rare in children, but happen much more commonly with increased age in adult.

Adapted from Kieran MW et al., JCO, 2010.



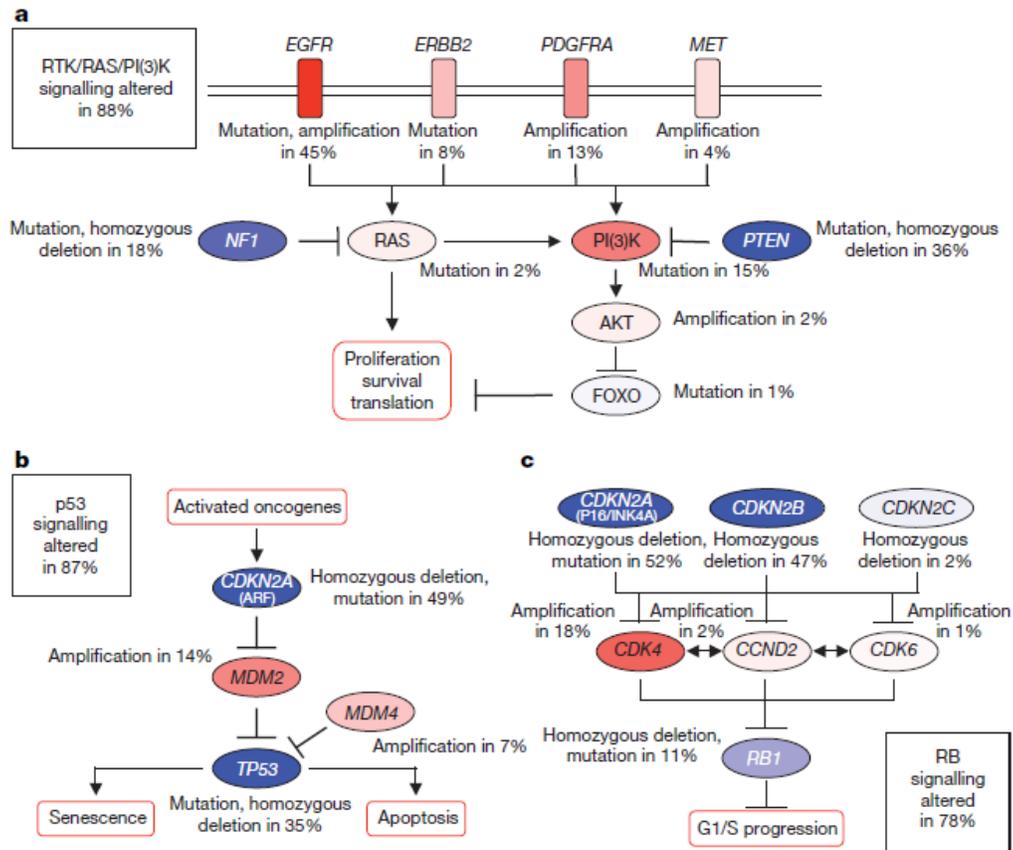
**Figure 1.4: Genetic pathways leading to grade II and III gliomas, as well as primary and secondary glioblastoma.**

*IDH* mutation is considered a precursor lesion in gliomagenesis. Additional *TP53* mutation leads to diffuse astrocytoma (grade II) formation, whereas loss of 1p/19q favors the onset of oligodendroglioma. Secondary GBM is formed through a progressive pathway, with history of previous identified lower grade lesions. Primary glioblastoma occurs de novo and harbors distinct genetic lesions from secondary, such as *EGFR* amplification, *PTEN* mutation, *NF1* alteration and LOH (loss of heterozygosity) of chromosome 10p. Adapted from Ohgaki H & Kleihues P, Brain Tumor Pathol, 2011.



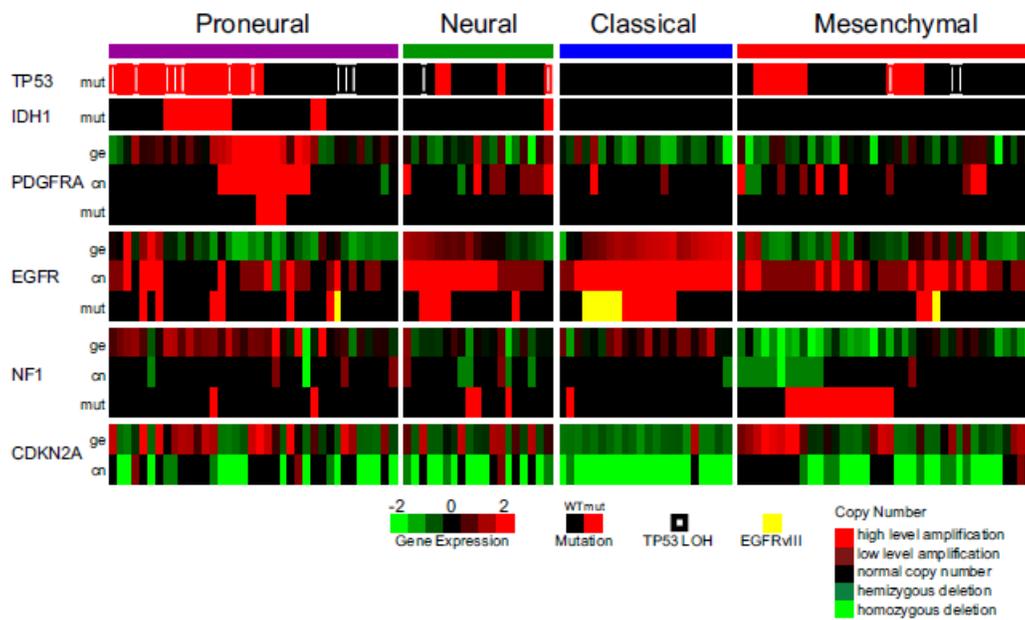
**Figure 1.5: Three core signalling pathways disrupted in adult glioblastoma through genetic alterations.**

Frequent genetic mutations and/or copy number alterations were found in a). RTK/RAS/PI3K pathway. b). p53 signalling. c). RB pathway for cell cycle control. Red color stands for activating genetic lesions; blue indicates inactivating events. The higher the frequency, the deeper the color. Adapted from The Cancer Genome Atlas Research Network, Nature, 2008.



**Figure 1.6: characteristic genetic alterations and gene expression of 4 molecular subtypes of adult glioblastoma.**

Data for mutations (mut), gene expression (ge) and copy number (cn) were visualized for important genes characterizing each subtype of adult glioblastoma. Adapted from Verhaak RG, Cancer cell, 2010.



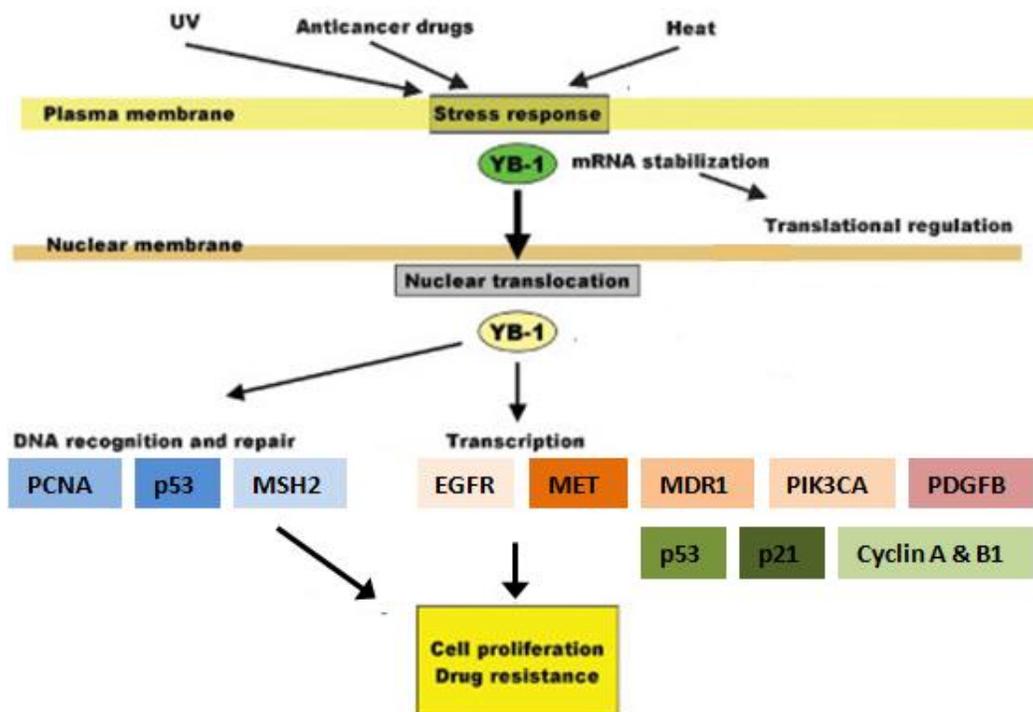
**Figure 1.7: Structure of YB-1 protein.**

YB-1 consists of an N-terminal domain (NTD), cold shock domain (CSD) and a C-terminal domain (CTD). It has both nuclear localization signal (NLS) and a cytoplasmic retention signal (CRS) located in the CTD.



**Figure 1.8: Multiple functions of YB-1 in the cell.**

YB-1 protein acts as an mRNA stabilizer and translational regulator by binding to the mRNPs in the cytoplasm. When cells encounter UV irradiation, DNA-damaging reagents or hyperthermia, YB-1 will translocate into the nucleus. The function of nuclear YB-1 includes transcriptional regulation of genes such as EGFR, MET, MDR1, PIK3CA, PDGFB, p53, p21 and cyclin A & B1. YB-1 in the nucleus can also help in DNA repair either by itself or through interaction with other DNA repair enzymes such as PCNA, p53 and MSH2. The nuclear functions of YB-1 explain its role in promoting proliferation and drug resistance of the cell. Adapted from Kohno K et al., Bioassays, 2003.



**Figure 1.9: Procedure of whole exome sequencing.**

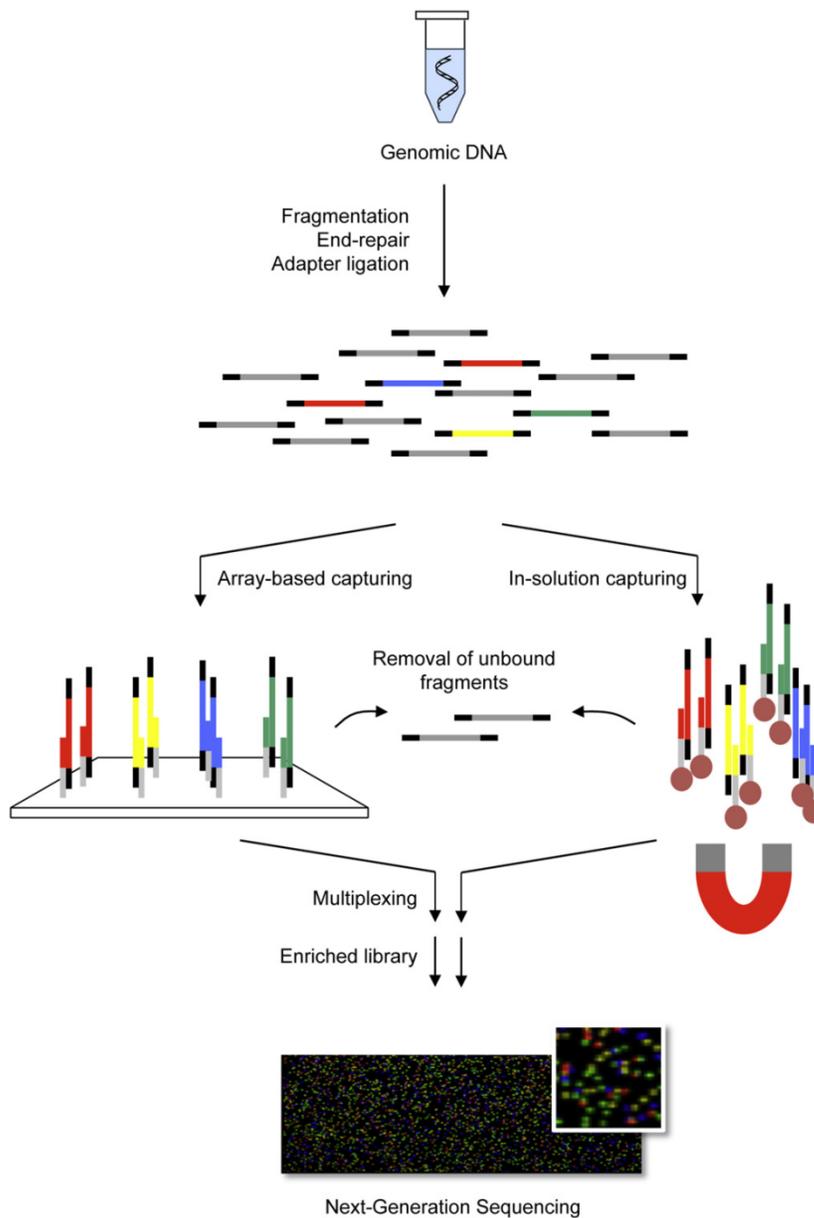
Genomic DNA is first fragmented and ligated with adaptors at both ends.

Then, the DNA fragments are hybridized on to the exome sequence

enrichment probes from the exome capture kits, either array-based or in-

solution. After, the bound DNAs are eluted and amplified for sequencing.

Adapted from Haas J et al., Mol Cell Probes, 2011.



Chapter 2: Sub-cellular localization of Y-box Protein 1 regulates proliferation, migration and tumorigenicity in astrocytomas

Xiao-Yang Liu\*, Noha Gerges\*, Damien Faury, Caroline Sollier, Steffen Albrecht, Brian Meehan, Zhifeng Dong, Peter Siegel, Andrey Korshunov, Stefan Pfister, Janusz Rak and Nada Jabado

\*These authors contributed equally to the manuscript.

Manuscript in preparation.

## 2.1. ABSTRACT

Y-Box-Protein-1 (YB-1) is a DNA/RNA-binding protein mandated for embryonic development and implicated in cancer progression/aggressiveness. We previously established elevated YB-1 levels in pediatric Glioblastoma (GBM), a high-grade brain tumor, which possibly contributes to oncogenesis. We investigated herein the effects of stable knock-down or overexpression of YB-1 in a pediatric GBM (SF188), an adult GBM (U87) and normal human astrocytes cell lines (NHA immortalized with H-Tert). In the two GBM cell lines, YB-1 silencing using shRNA reduced YB-1 level by 90%, and surprisingly increased proliferation and EGFR level. Silencing of YB-1 in these cells caused residual YB-1 to locate in the nucleus, indicative of the role of this sub-cellular localization in cell survival. Overexpressed YB-1 was predominantly cytoplasmic in the GBM and NHA cell lines, and decreased cell proliferation and increased migration. Orthotopic injection of YB-1 overexpressed U87 cells into immune-compromised mice showed decreased tumorigenicity. Immunohistochemistry (IHC) on tissue microarrays showed strong nuclear YB-1 expression in 64% of the 107 pediatric GBM samples, but only in 17% of the 63 grade I pilocytic astrocytomas, suggesting a role of YB-1 in a more invasive tumor

phenotype. We also confirmed the association of nuclear YB-1 overexpression with EGFR overexpression in the 107 pediatric GBM samples. Our results suggest that YB-1 modulates proliferation and migration, based on its sub-cellular localization. While ascertaining the role of nuclear YB-1 in driving cell growth and the association of nuclear YB-1 with EGFR overexpression and more aggressive tumor phenotype, our data argue for caution in targeting YB-1 for therapeutic intervention.

## 2.2. INTRODUCTION

Y-box protein 1 (YB-1) was identified as a transcription factor that binds to the Y-box of *EGFR* family promoters (Sakura, Maekawa et al. 1988).

Increased YB-1 expression has been detected in a wide range of human cancers including breast, ovarian, colorectal and lung cancers, *etc*, and linked to tumor progression and unfavorable outcome, especially when YB-1 is localized in the nucleus (Eliseeva, Kim et al. 2011). However, the exact mechanisms by which YB-1 mediates oncogenesis remain unclear and controversial. This ubiquitously expressed DNA/RNA binding protein belongs to the cold shock domain family of proteins, which is highly conserved with a 98% identity between human and mouse proteins (Evdokimova and Ovchinnikov 1999; Kohno, Izumi et al. 2003). YB-1 is broadly expressed throughout fetal development (Grant and Deeley 1993; Ito, Tsutsumi et al. 1994) where it plays an important role as targeted knock-out of YB-1 causes late embryonic death with neural tube closure defects, exencephaly and hypoplasia of other major organs (Lu, Books et al. 2005). After birth, YB-1 expression level correlates with the cell proliferation state and is turned down in quiescent cells, while high levels are detected *in vivo* in actively proliferating adult tissues (Shibao, Takano et al. 1999) and regenerating tissue following damage (Ito, Tsutsumi et al.

1994; Ito, Yoshida et al. 2003). In the brain, a high level of YB-1 is expressed prior to the first week after birth, thereafter YB-1 expression substantially declines and is low in the adult brain (Funakoshi, Kobayashi et al. 2003; Lu, Books et al. 2005; Uchiumi, Fotovati et al. 2006). YB-1 is also induced in response to mitogenic stimuli (Sabath, Podolin et al. 1990; Ito, Tsutsumi et al. 1994; Ito, Yoshida et al. 2003).

YB-1 has pleiotropic functions depending on its subcellular localization. In mammalian cells, it is a major structural component of messenger ribonucleoproteins (mRNPs) in the cytoplasm where it acts as translational regulator repressing many mRNPs encoding stress- and growth-related proteins when the YB-1/mRNA ratio is high (Evdokimova, Ovchinnikov et al. 2006; Evdokimova, Ruzanov et al. 2006; Eliseeva, Kim et al. 2011). Meanwhile, cytoplasmic YB-1 can promote translation of cap-independent mRNAs, such as Snail1 and other transcription factors implicated in epithelial mesenchymal transition (EMT) (Evdokimova, Tognon et al. 2009). In the nucleus, YB-1 functions as a transcriptional activator for genes involved in cell proliferation and cell cycle control, such as the *EGFR*, *MET*, *PIK3CA* and cyclin A/B1 genes, as well as multidrug resistance gene *MDR1* (Ohga, Koike et al. 1996; Ohga, Uchiumi et al.

1998; Stein, Jurchott et al. 2001; Jurchott, Bergmann et al. 2003; Stratford, Habibi et al. 2007; Astanehe, Finkbeiner et al. 2009; Finkbeiner, Astanehe et al. 2009; Sengupta, Mantha et al. 2011). Nuclear YB-1 can also promote DNA repair and pre-mRNA splicing (Eliseeva, Kim et al. 2011), as well as interact with p53 to inhibit p53-induced cell death (Homer, Knight et al. 2005; Kim, Choi et al. 2008).

The regulation of YB-1 subcellular localization is complicated and remains unclear. Under normal physiological condition, YB-1 is mainly expressed in the cytoplasm in complex with mRNA; it is translocated into the nucleus when cell encounters G1/S cell cycle transition, growth factor/cytokine stimuli or stress stimuli (Eliseeva, Kim et al. 2011). P53 has been shown to aid the transport of YB-1 from cytoplasm into the nucleus (Zhang, Homer et al. 2003). Inconsistent results were shown for YB-1 phosphorylation by Akt at Ser102: it induces YB-1 translocation into the nucleus in breast and ovarian cancer cell lines (Sutherland, Kucab et al. 2005; Basaki, Hosoi et al. 2007), but not in mouse (Evdokimova, Ruzanov et al. 2006) or chicken embryonic fibroblast cell lines (Bader and Vogt 2008). The p90 ribosomal S6 kinase (RSK) has been shown to also phosphorylate YB-1 at Ser102 and to activate EGFR transcription, but no

direct link was shown between RSK-mediated YB-1 phosphorylation with its nuclear localization (Stratford, Fry et al. 2008). Recently,  $\Delta Np63\alpha$ , the predominant p63 protein isoform in squamous epithelia was shown to promote nuclear accumulation of YB-1 (Di Costanzo, Troiano et al. 2012), with the exact mechanism remaining somewhat unclear. Further studies are necessary in order to completely decipher the mechanism of YB-1 shuttling between cytoplasm and nucleus.

Pediatric GBM (pedGBM) is a highly aggressive cancer with a dismal three year survival of less than 20% of affected children (Pollack 1999; Broniscer 2006). Previously we identified increased YB-1 expression in 75% of pedGBM samples (Faury, Nantel et al. 2007; Haque, Faury et al. 2007) and were the first to associate YB-1 with brain tumors. Important genes/pathways altered during gliomagenesis including EGFR, PI3K/AKT pathway and p53 disruption are also tightly involved in YB-1 regulation and function. This warrants our studies herein on the role of YB-1 in astrocytomas formation/progression. To this issue, we used primary GBM cell lines derived from pediatric and adult GBM samples as well as primary H-Tert-immortalized astrocytes, to investigate the sub-cellular localization and function of YB-1 following stable knock-down or ectopic expression of

this factor. We also performed orthotopic xenograft of YB-1 overexpressing GBM cells into immune-compromised mice to assess tumorigenicity. Finally, YB-1 and EGFR expression levels and their correlation in pediatric GBM were investigated using patient sample tissue microarrays. Based on our data, we suggest that YB-1 regulates tumor formation and progression in astrocytomas depending on its sub-cellular localization. Its high nuclear expression in pediatric GBM is associated with EGFR overexpression and promotes cell proliferation, whereas cytoplasmic YB-1 increases cell migration while limits cell proliferation and tumorigenicity. Therefore, we conclude that direct targeting of YB-1 will prove challenging and will need to be combined with at least anti-proliferative therapies.

## 2.3. MATERIALS AND METHODS

### *Cell cultures*

SF188 (UCSF Brain Tumor Research Center Tissue Bank USA), U87 (ATCC) and HTert-Immortalized normal human astrocytes (NHA, kind gift from Dr. C Hawkins, HSC, Canada) were cultured in EMEM or DMEM (Wisent) supplemented with 10% fetal bovine serum (FBS). Growth factor

stimulations were performed following overnight serum-starvation. Cells were stimulated with epidermal growth factor (50ng/ml, Sigma, USA).

*Immunoblotting (IB), Immunofluorescence (IF) and Nuclear/ Cytoplasmic Fractionation*

IB was performed using protocols previously described (Faury, Nantel et al. 2007) for PTEN, pAKT and AKT antibodies. Similar procedure were used for antibodies against YB-1 (AVES), HA-tag (Covance), p53 (Santa Cruz), EGFR (Santa Cruz), CREB (Cell Signaling) and  $\beta$ -actin (Cell Signaling), with primary antibody dilutions of 1:20000, 1:500, 1:500, 1:200, 1:1000 and 1:1000, respectively. Nuclear/cytoplasmic fractionation was performed as described elsewhere (Jabado, Le Deist et al. 1994). IF was performed on paraformaldehyde-fixed cells, with primary antibodies HA (Covance 1:500) and YB-1 (Abcam 1:500) incubation overnight at 4°C. Slides were then incubated with secondary antibodies anti-rabbit-fluoro488 and anti-mouse-Cyr3 for 30 minutes at room temperature. Mounting media containing nuclear staining DAPI was applied at the end. IF images were acquired using Zeiss Axioplan 2 fluorescence microscope or Nikon confocal Eclipse TE2000-E microscope.

*Stable ectopic expression or silencing of YB-1*

Cells were counted and plated in 6-well plates and transfected using Lipofectamine™2000 and opti-MEM® I Reduced-Serum Medium (Invitrogen) the next day as previously described (Evdokimova, Ruzanov et al. 2006). N-terminal-HA-tagged YB-1 plasmid was a kind gift of Dr. V. Evdokimova (UBC, Canada). The plasmid used for silencing YB-1 was generated from pSUPER.neo+gfp (OligoEngine), to express Short Hairpin RNA targeting YB-1. Positive clones were selected and tested by IF and IB.

***RNA extraction, reverse-transcription, and Quantitative real-time PCR***

Total RNA was isolated using the Aurum Total RNA Mini Kit from Bio-Rad and quantified with ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, Delaware, USA). 50ng of extracted RNA was reverse-transcribed using the iScript synthesis kit (BioRad). Then 2µL of cDNA was mixed with Sofast Evagreen SuperMix (BioRad) following manufacturer's instructions. Cycling conditions were 5 sec at 95°C - 20 sec at 58°C for 40 cycles on the Roche LightCycler 480 (Roche). The primers used were 5'-CAAGAAGGTCATCGCAACGAAGGT-3'(YB1-forward), 5'-CAGGACCTGTAACATTTGCTGCCT-3' (YB1-reverse); 5'-AAATACAGCTTTGGTGCCACCTGC-3' (EGFR-forward), 5'-

AGGCCCTTCGCACTTCTTACACTT-3' (EGFR-reverse); 5'-GGCACCCAGCACAATGAAGATCAA-3' (Actin-forward) and 5'-TAGAAGCATTGCGGTGGACGATGGA-3' (Actin-reverse). YB1 and EGFR expression levels were calculated using the deltaCt method using Actin as a reference gene and a normal brain control as a calibrator.

### *Cell growth in monolayer and soft agar, migration assays*

Cell proliferation in monolayer was assessed by <sup>3</sup>H-thymidine incorporation assay (Coward, Wada et al. 1998), adapted for NHA and GBM cell lines. The radioactive signal representing DNA synthesis was read by the Liquid Scintillation Counter machine (Perkin Elmer). 3-D anchorage independent cell growth was assessed using soft agar assay: 1×10<sup>4</sup> cells in DMEM containing 10% FBS and 0.3% agarose were placed on top of a layer of 0.5% agarose. The number of colonies formed was counted under the microscope after 3 weeks. For migration assays, 100 µl of cells were added on the top chamber of 8.0 µm pores Transwell system (Corning). The bottom chambers were filled with either 600 µl of FBS-free (control) or 10% FBS (assay) media. After 24 hours of migration, bottom and top chamber were fixed with formaldehyde and stained with crystal violet. Non migrating cells were removed by scrubbing the top chamber

with cotton swab, and migrated cells at the bottom of the chamber were taken pictures with microscope. Signals representing the number of cells that have migrated through the chamber were analyzed by Scion Image. Each clone was plated in triplicate and each experiment was reproduced at least 3 times.

#### *Orthotopic xenograft mouse model*

U87 GBM cells were engineered to express luciferase and injected intracranially (150000 cells per mouse in 2  $\mu$ l media containing 10% FBS) into the right hemisphere of NOD/SCID mice (Harlan Laboratories, NOD.CB17). Mice were then monitored for bioluminescent signals representing tumor cell growth after luciferin injection once a week, using the IVIS Spectrum system (Perkin Elmer). Mice were sacrificed when 10% loss of weight was observed, and Kaplan-Meier Curve was plotted for survival.

#### *Tissue microarray (TMA) and Immunohistochemistry (IHC)*

We have previously described the features and preparation of the TMAs, as well as procedures and analysis of IHC (Liu, Gerges et al. 2012; Schwartzentruber, Korshunov et al. 2012). Two TMAs were used: one

containing 107 pediatric GBM tumor cores, and the other containing 63 pilocytic astrocytomas. Primary antibodies used for IHC include YB-1 (AVES, 1:2000 dilution) and EGFR (Santa Cruz, 1:500 dilution).

## 2.4. RESULTS

*Characterization of the cell lines used in this study.* Based on their frequent alteration in GBM and the relevance of these molecules/pathways in YB-1 sub-cellular localization, we investigated in the three cell lines used in this study PTEN expression, activation of the AKT pathway, p53 expression/mutational status, and EGFR levels. U87 and SF188 GBM cell lines have constitutively active AKT pathway as ascertained by absence of PTEN and increased phosphorylation of AKT in serum starved cells. P53 was mutated in SF188. HTert-immortalized NHA harbored none of these abnormalities (Figure S2.1).

*Stable YB-1 knock-down increases cellular proliferation while its ectopic expression doesn't give growth advantage in GBM cell lines.* We stably knocked-down YB-1 expression in 1 adult (U87) and 1 pediatric (SF188) GBM cell lines using an shRNA plasmid (shYB-1) and assessed its effects

on cell proliferation and migration, two hallmarks of GBM. We confirmed YB-1 protein levels to be reduced by more than 80%, in stable shYB-1 clones generated from both cell lines (Figure 2.1A) and investigated cell growth in monolayer and soft agar. Surprisingly, decreasing YB-1 levels induced in both cell lines increased growth rate using both assays (Figure 2.1B-C). We then investigated the effects of stable ectopic expression of wild type HA-tagged-YB-1 in these 2 GBM cell lines, and no growth advantage was observed in monolayer cultures (Figure 2.1D).

As invasion is one of the hallmarks of astrocytomas, we then assessed whether loss or gain of YB-1 affected the migration of these cell lines, and no significant influence was observed on cell migration (Figure S2.2).

Our findings indicate that, in GBM cell lines, decreasing YB-1 level enhances cell proliferation whereas ectopic expression of this factor doesn't give cell growth advantage.

*Silencing YB-1 increases the levels of nuclear YB-1 in GBM cell lines, while ectopic YB-1 is mainly expressed in the cytoplasm.* Our group is the one of the first efforts to establish astrocytoma cell lines with this degree of

stable knock-down of YB-1. Despite repeated attempts (about 47 clones screened per cell line), we still observed residual YB-1 expression between 10-20% in all stable knock-down clones generated and could not achieve complete silencing of the protein. This finding is concordant with reports using stable knock-down of other factors and that of groups using transient knock-down of YB-1 in other types of cell lines, which also achieve in their studies this level of inhibition for this factor (Mertens, Steinmann et al. 2002; En-Nia, Yilmaz et al. 2005; Stratford, Habibi et al. 2007; Chatterjee, Rancso et al. 2008). As YB-1 is predominantly cytoplasmic in physiological conditions in all cells (Kohno, Izumi et al. 2003; Evdokimova, Ovchinnikov et al. 2006), and nuclear YB-1 is potentially associated with increased tumor aggressiveness in cancers, we investigated the sub-cellular localization of this factor in YB-1 stable knock-down clones. To this issue, we used immunofluorescence and confocal imaging, and in parallel sub-cellular fractionation and immunoblotting. We showed in both GBM cell lines, that residual YB-1 following stable silencing is predominantly nuclear. Indeed, YB-1 was greatly reduced to absence in cytoplasmic extracts while their respective nuclear extracts were enriched in YB-1 compared to EV transfectant-cells (Figure 2.2A). This was further confirmed using confocal imaging: YB-1

(green staining) localized mainly in the nucleus (dapi, blue staining) in all shYB-1 cells as shown by the overlay (light blue) in the representative z-stack (Figure 2.2B).

In SF188 and U87 cells with ectopic expression of HA-tagged YB-1, predominant cytoplasmic localization was observed (Figure 2.2C, Figure S2.3). Despite constitutive activation of the AKT pathway in all these cell lines limited amounts of endogenous or HA-tagged YB-1 were observed in the nucleus. Also, there was no increase in the level of nuclear YB-1 in EV and HAYB-1 transfectants following EGFR (Figure S2.3) or insulin-growth-factor receptor stimulation (data not shown).

These results indicate that targeting YB-1 leads predominantly to the loss of the cytoplasmic component and induces the residual protein to localize to the nucleus possibly allowing the cell to survive and promoting cell growth; whereas ectopically overexpressed YB-1 was mainly localized in the cytoplasm.

***Nuclear YB-1 increases EGFR levels.*** Nuclear YB-1 increases the expression of several genes including EGFR where YB-1 binding sites

have been characterized in the promoter (Stratford, Habibi et al. 2007). We investigated whether increased nuclear YB-1 correlated with increased EGFR levels in shYB-1 clones compared to empty-vector transfectants. Quantitative real-time PCR analysis of mRNA extracted from shYB-1 U87 and SF188 cell lines showed increased EGFR levels compared to empty-vector transfectant cells (Figure 2.3A), translated into increased EGFR protein level in SF188 cells (Figure 2.3B).

***Ectopic expression of YB-1 in HTert-immortalized astrocytes reduces cell proliferation, increases cell migration.*** GBM cell lines harbor several molecular alterations that may affect YB-1 function and sub-cellular localization. Except for increased telomerase activity, HTert-immortalized human astrocytes (NHA) retain all of the physiological properties of normal astrocytes and have been used to model them (Kamnasaran, Qian et al. 2007). In NHA, endogenous YB-1 was predominantly expressed in the cytoplasm as did HA-tagged-YB-1 following stable overexpression (Figure 2.3C). YB-1 overexpression induced decreasing cell proliferation and increasing cell migration (Figure 2.3D-E). Stable silencing of YB-1 using shRNA in NHA cells led to decreased YB-1 levels both in the nucleus and the cytoplasm (Figure 2.3C); slightly decreased proliferation and migration

were observed in these cells (Figure 2.3D-E). These findings further confirm that nuclear YB-1 promotes cell proliferation while cytoplasmic YB-1 increases cell migration and limits cell growth.

***YB-1 overexpression decreases tumorigenicity and improves survival of GBM cells in orthotopic mice injection.*** Luciferase expressing U87 cells were transfected with empty-vector or HA-tagged-YB-1 plasmids to assess the effect of YB-1 overexpression *in vivo*. Cells were injected into the right hemisphere of immuno-compromised mice, and tumor onset was monitored by luciferase luminescent activity (Figure 2.4A). Delayed tumorigenic ability and increased mice survival was observed in U87 cells overexpressing YB-1 (Figure 2.4B-C). These results further validate the role of cytoplasmic YB-1 in limiting cell growth instead of promoting it, as shown by the *in vitro* experiments.

***YB-1 nuclear overexpression is associated with higher grade and EGFR overexpression in patient tumor samples.*** IHC staining of YB-1 and EGFR was performed on tissue microarrays (TMA) containing 107 pediatric GBM and 63 grade I pilocytic astrocytomas. Representative pictures of IHC positive and negative staining for YB-1 and EGFR are included (Figure

S2.4). YB-1 overexpression and its nuclear localization are present in significantly more samples of pediatric GBM (90% and 64% respectively) compared to pilocytic astrocytoma ( $p < 0.0001$ , Table 2.1). The nuclear overexpression of YB-1 in pediatric GBM patient samples is associated with the overexpression of EGFR (Figure S2.5), possibly through the transcriptional activation of EGFR gene by nuclear YB-1. However, we didn't identify any indication on progression free survival and overall survival based on YB-1 overexpression or nuclear localization.

## 2.5. DISCUSSION

Our data indicate that silencing of YB-1 expression in GBM cell lines leads to increased cellular growth, possibly through its nuclear function of activating transcription of pro-growth genes, such as EGFR. Stable YB-1 silencing in NHA cells caused decreased cell proliferation possibly due to a decreased YB-1 expression in the nucleus. Overexpressed YB-1 is mainly located in the cytoplasm to limit cellular growth and tumorigenicity while promote cell migratory ability. Despite constitutive activation of the AKT pathway in our GBM cell lines, a predominant cytoplasmic localization of endogenous and ectopic YB-1 was observed, and nuclear

shuttling of YB-1 was mainly independent of an active AKT pathway (SF188) or wild type p53 (U87), suggesting that sub-cellular regulation of YB-1 is governed by additional factors other than AKT and p53 in astrocytomas.

The role of YB-1 as an oncogene has been challenged as some groups identify it as anti-oncogenic (Bader, Felts et al. 2003; Bader and Vogt 2008) whereas others, traditionally in the cancer field, identify it as an oncogene promoting cellular growth and transformation (Kohno, Izumi et al. 2003; Sinnberg, Sauer et al. 2012). We could not generate stable transfectant-clones with complete silencing of YB-1, and interestingly in all the YB-1-silenced GBM clones we generated, residual YB-1 was predominantly present in the nucleus, indicative of the importance of this sub-cellular localization in cell survival. In keeping with the potential need of residual YB-1 expression for cells to survive, mouse embryonic fibroblast generated from YB-1 knock-out mice had a very limited number of cell division and senesced in the absence of this factor (Lu, Books et al. 2005). Importantly, incomplete silencing of YB-1, even though we achieved a drastic decrease of 80-90% of the protein levels, led to increased cell growth, and EGFR levels in the GBM cell lines. This may be

due to the dual effect of alleviating the repression on pro-mRNAs implicated in cell proliferation/growth following silencing of cytoplasmic YB-1, and the direct effects of residual nuclear YB-1 on transcription of genes involved in cell growth including EGFR. Up to 20% of cellular mRNAs are associated with YB-1 and, therefore, may be targets of YB-1-mediated regulation at the level of translation or stability (Evdokimova, Ruzanov et al. 2001; Evdokimova, Ovchinnikov et al. 2006; Evdokimova, Ruzanov et al. 2006). Thus, alleviating repression mediated by cytoplasmic YB-1 could prove detrimental and increase cell growth as evidenced in our cell lines following YB-1 silencing.

Ectopic expression of YB-1 in the cytoplasm increased cell migration in NHA cells, which are the most physiological model for astrocytes. It has been shown that, in non-invasive breast epithelial cells with a constitutively active Ras pathway, enforced expression of YB-1 induced an epithelial-mesenchymal transition (EMT) accompanied by enhanced metastatic potential and reduced proliferation rates (Evdokimova, Tognon et al. 2009), similar to our findings in NHA. In that report, YB-1 directly activated cap-independent translation of mRNAs encoding Snail1 and other transcription factors implicated in downregulation of epithelial and

growth-related genes, and activation of mesenchymal genes. Invasion is a hallmark of astrocytomas of all grades and this feature may be accentuated by aberrant YB-1 expression. It is interesting to note that in our study, in NHA, there was no need for an active Ras pathway for this migratory phenotype. This may be due to the fact that astrocytes are of mesenchymal origin and more easily prone to changes affecting their migration potential. We didn't observe significant changes when altering YB-1 expression in either SF188 or U87 cells, possibly due to the fact that they're GBM cell lines and already have high migratory ability even before any manipulation of the YB-1 level.

In summary, dual sub-cellular localization of YB-1 may be associated with dual function. We identify nuclear localization of YB-1 to be mainly independent of AKT activation in GBM, and indicate that it is presumably this portion which is needed for cell survival. Cytoplasmic YB-1 promotes cell migration, while limiting cell growth and tumorigenicity. We suggest that targeting YB-1 can be contemplated in astrocytomas as silencing the cytoplasmic component of this protein would prove beneficial in decreasing cell migration. However, it calls for caution and, at the very least, has to be envisaged in association with anti-proliferative therapies.

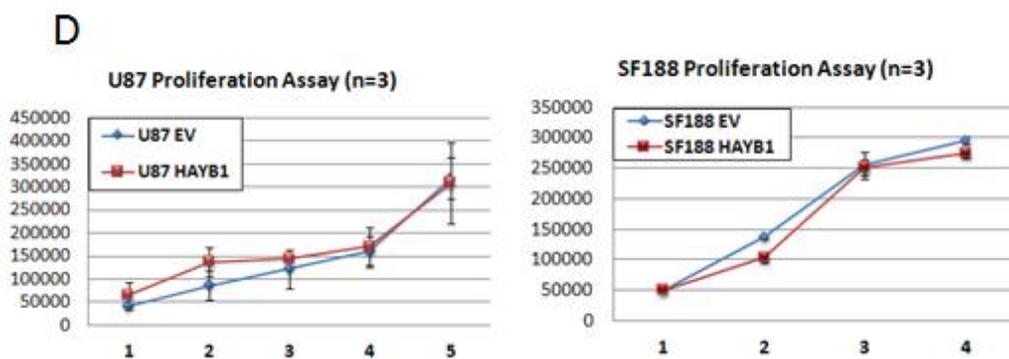
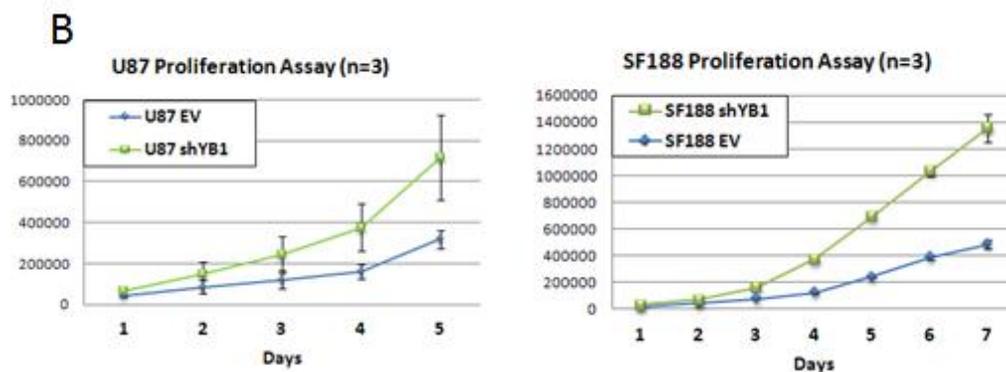
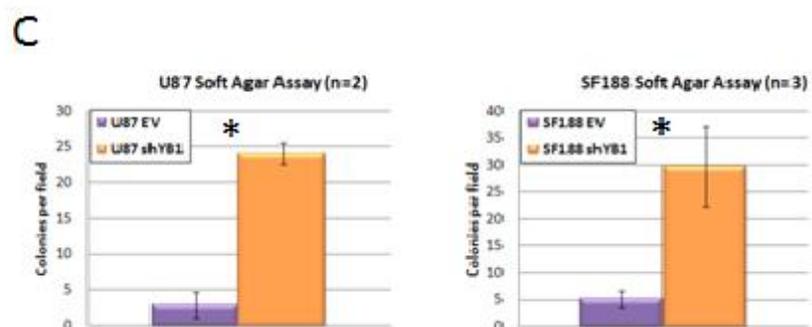
## 2.6. AUTHOR CONTRIBUTIONS

XY Liu, N Gerges, Damien Faury, C Sollier, S Albrecht, B Meehan, Z Dong and P Siegel performed experimental work. XY Liu, D Faury, C Sollier and N Jabado performed data analyses and produced text and figures. A Korshunov and S Pfister collected patient material and data. J Rak and N Jabado provided leadership for the project.

## 2.7. FIGURES

**Figure 2.1: Effect of YB-1 overexpression and silencing on cellular growth in GBM cell lines U87 and SF188.**

A). Ectopic expression of HA-tagged-YB-1 and knock down of YB-1 by shRNA silencing, shown by Western Blot. B). 2-D monolayer proliferation assay of empty vector (EV) and shYB1 plasmid transfected U87 and SF188 cells. C). Soft agar assay of EV and shYB1 transfected U87 and SF188 cells. D). 2-D monolayer proliferation assay of EV and HAYB1 plasmid transfected U87 and SF188 cells.



**Figure 2.2: Subcellular localization of YB-1 after ectopic overexpression and shRNA silencing in GBM cell lines.**

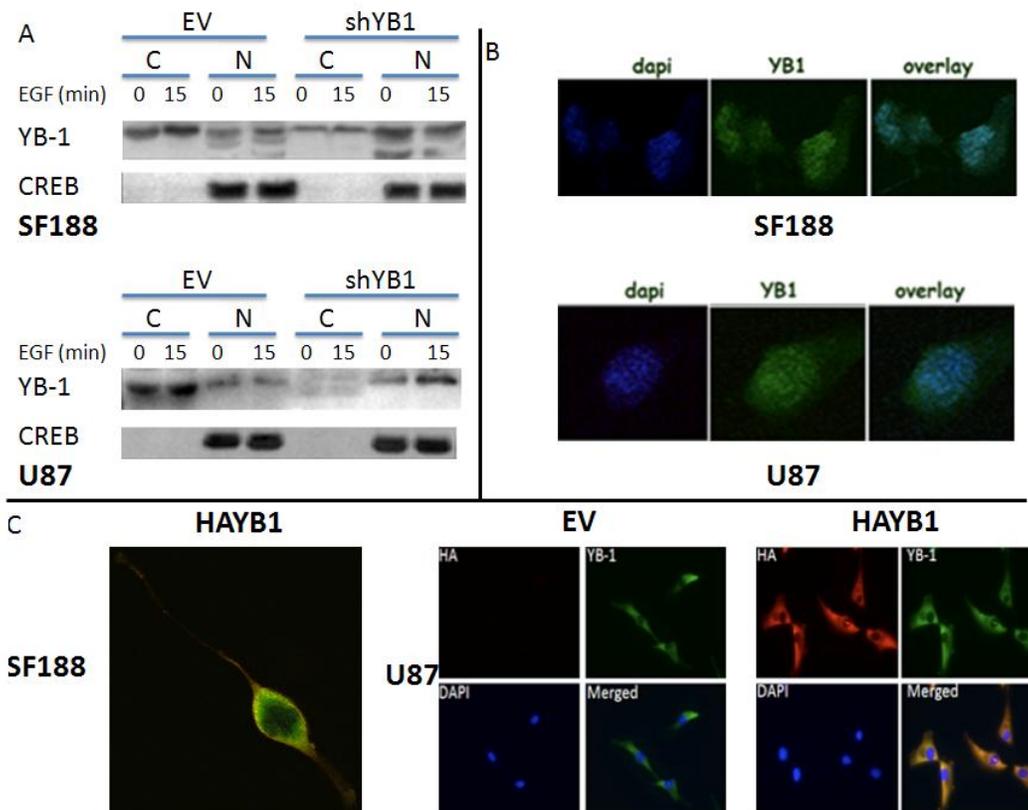
A). Nuclear cytoplasmic extraction showing decreased cytoplasmic YB-1 and increased nuclear YB-1 expression after silencing. B).

Immunofluorescence (IF) confocal image showing nuclear YB-1

expression after silencing. C). Endogenous and ectopic YB-1 are mainly

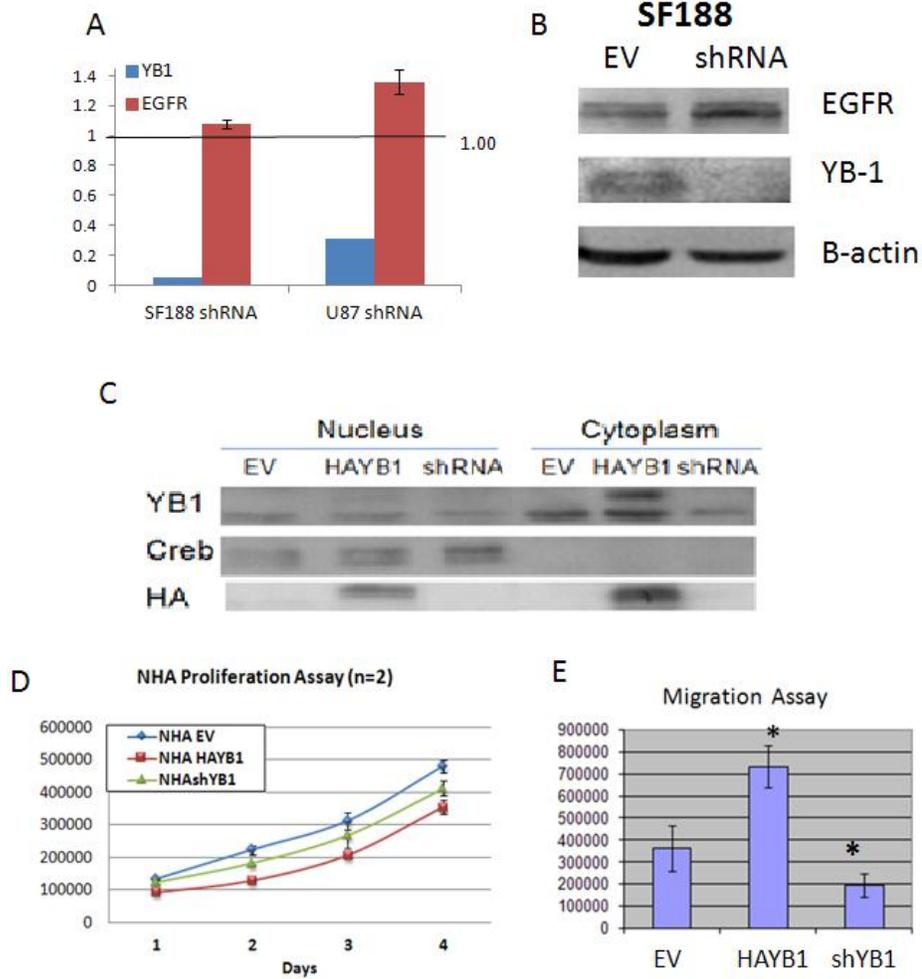
expressed in the cytoplasm of SF188 and U87 cells, shown in the HAYB1

transfected SF188 and U87 cells by IF imaging.



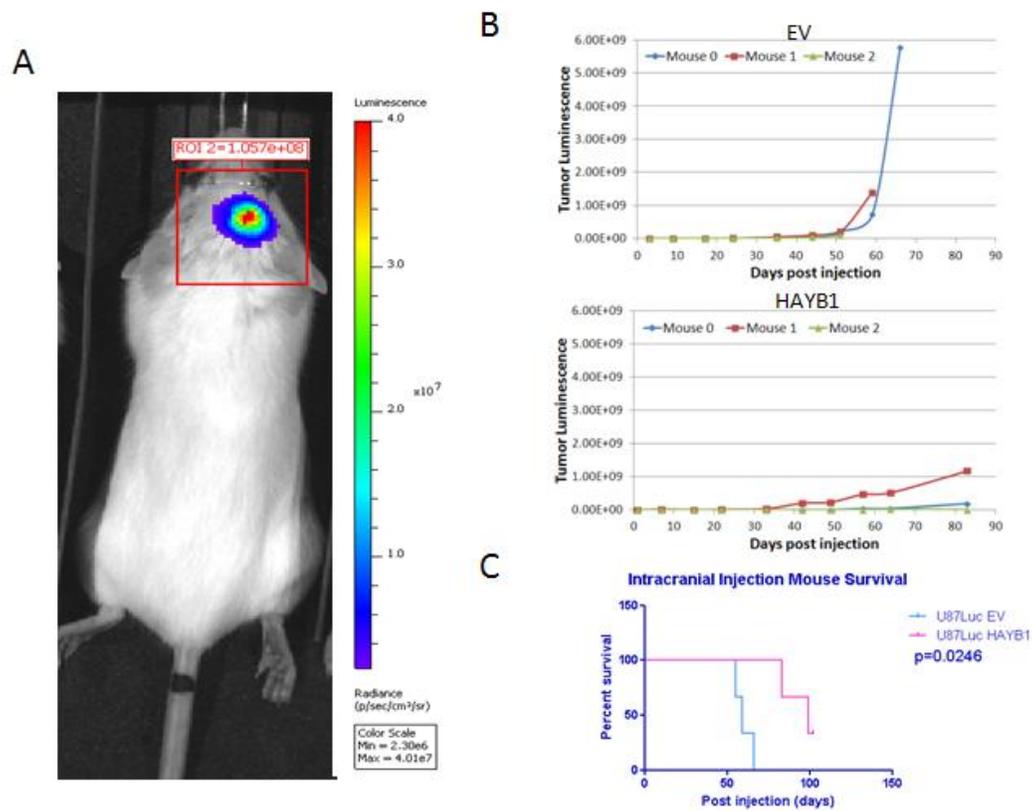
**Figure 2.3: Nuclear expression of YB-1 caused by YB-1 silencing with shRNA led to increased EGFR; effect of YB-1 overexpression and silencing on cellular growth and migration in NHA cell line.**

A). Quantitative real time PCR showed increased EGFR mRNA level after YB-1 silencing. The values were plotted as the ratio of shRNA to EV in SF188 and U87 cell lines. B). In SF188 cells, silencing of YB-1 increased the expression of EGFR at the protein level. C). Nuclear cytoplasmic extraction showing localization of YB-1 expression in YB-1 overexpressed and knocked-down NHA cells. B). Monolayer proliferation assay of NHA cells transfected with EV, HAYB1 and shYB1 plasmids. C). Migration assay of NHA cells transfected with EV, HAYB1 and shYB1 plasmids.



**Figure 2.4: YB-1 overexpression decreases tumorigenicity of GBM cells in xenograft mouse model.**

A). Luciferin luminescence photo to monitor the growth of intracranial tumor in the mouse after tumor cells injection. B). U87 cells overexpressing YB-1 showed decreased tumor growth intracranially in immune-compromised mice. C). Mice with intracranial injection of YB-1 overexpressing U87 cells survived better than those injected with control U87 cells.



## 2.8. TABLES

Table 2.1: Overexpression and nuclear localization of YB-1 found in pediatric GBM and pilocytic astrocytoma samples, using immunohistochemistry on tissue microarrays.

	YB-1 Overexpressed	YB-1 Nuclear
Pediatric GBM	96/107 (90%)	68/107 (64%)
Pilocytic Astrocytoma	19/63 (30%)	11/63 (17%)
p-value	<0.0001	<0.0001

## 2.9. SUPPLEMENTARY FIGURES

Figure S2.1: Characterization of molecular alterations in U87, SF188 and NHA cell lines.

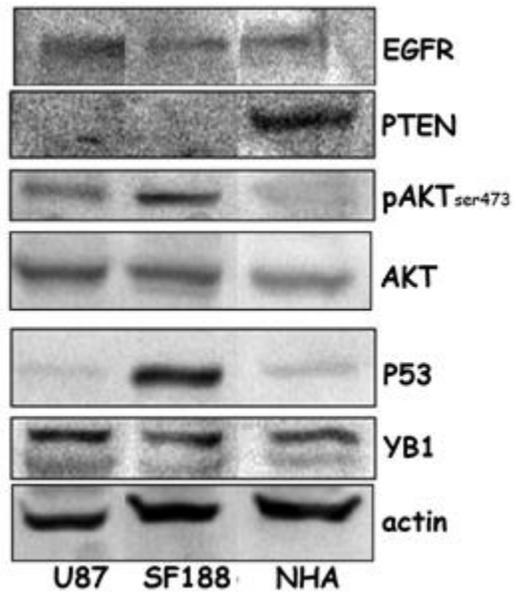


Figure S2.2: Migration and invasion assays for SF188 and U87 showed no significant difference among EV control, YB-1 knock down (shYB1) and overexpression (HAYB1).

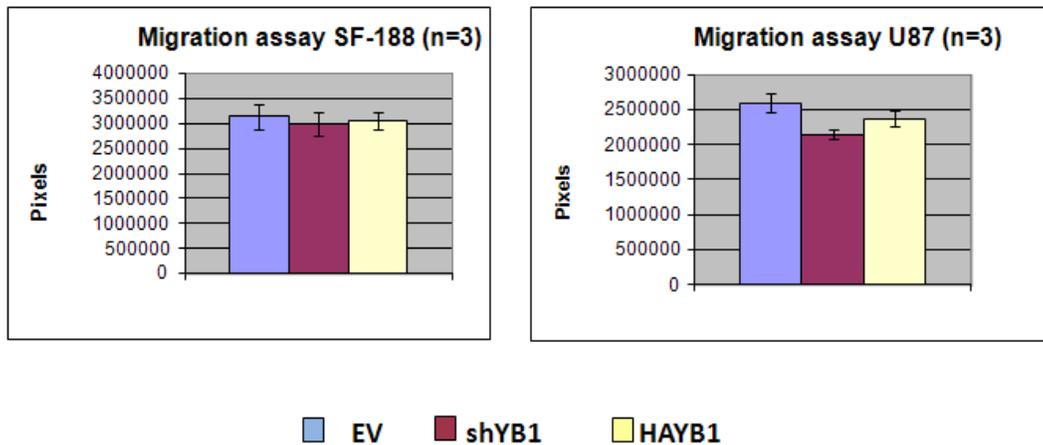


Figure S2.3: Nuclear cytoplasmic extraction confirming that overexpressed YB-1 is mainly localized in the cytoplasm.

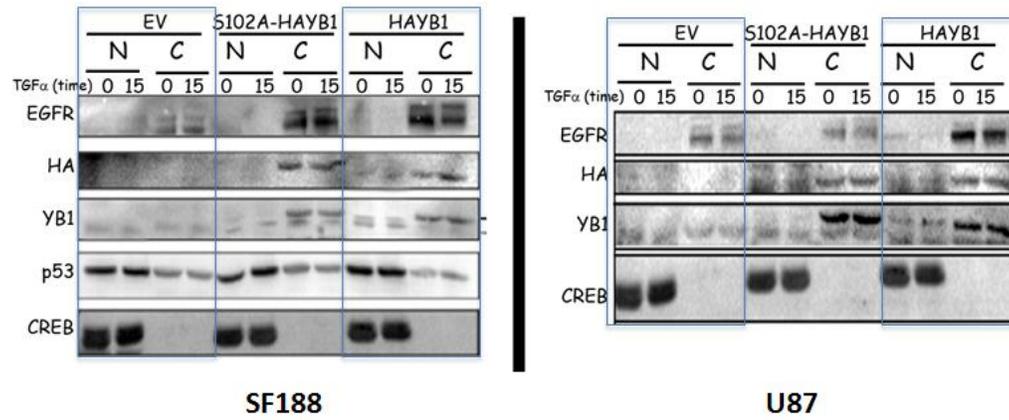


Figure S2.4: Representative IHC pictures showing negative, cytoplasmic and nuclear expression of YB-1 (upper panels), as well as positive and negative EGFR expression (lower panels) in pediatric GBM tumor paraffin-embedded slides.

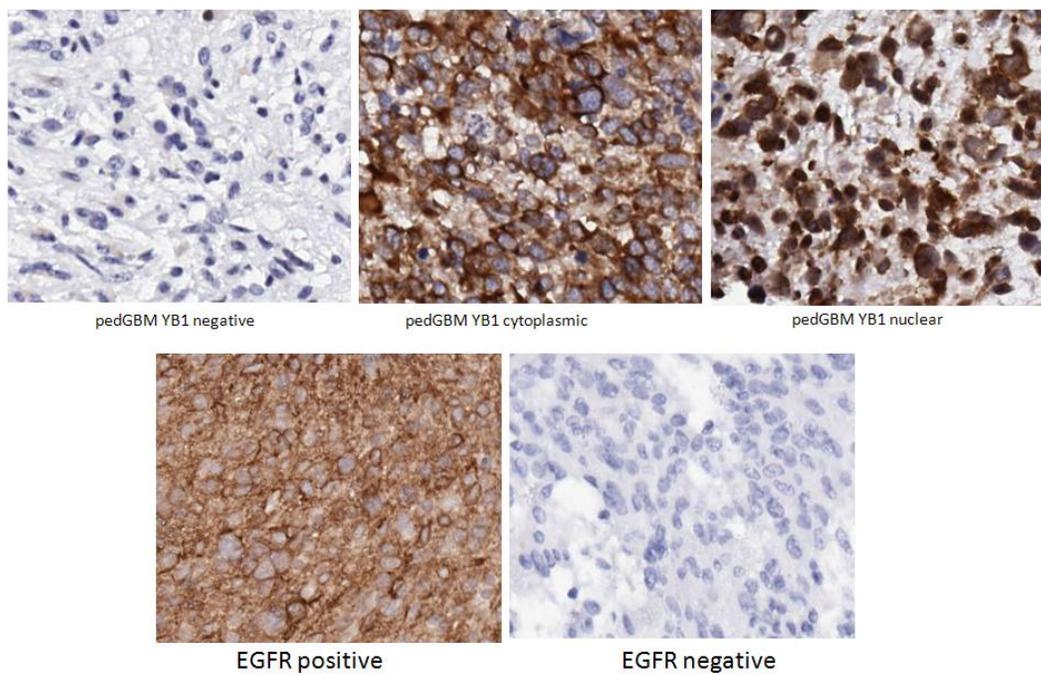
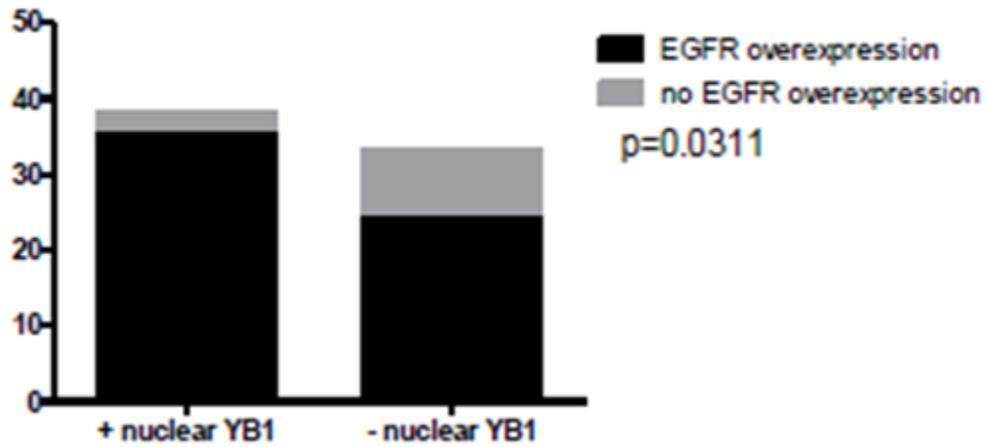


Figure S2.5: Nuclear expression of YB-1 in pediatric GBM patient samples is significantly associated with overexpression of EGFR.



## CONNECTING TEXT CHAPTER 2 TO 3

We previously identified YB-1 to be upregulated in 75% of pediatric GBM. According to its extensive involvement in other types of cancers and its regulatory role in major pathways disturbed in adult GBM, we investigated here in Chapter 2 its molecular function in astrocytoma genesis. We showed the dual function of YB-1 based on its intracellular localization and suggested caution when targeting YB-1 for therapeutic purpose. In order to improve disease management of pediatric GBM, novel targets need to be identified. There is still not much known about the alterations at the genetic mutation level in pediatric GBM. Identifying mutations at the genomic scale in pediatric GBM will help in understanding the driving force of the tumorigenesis, thus providing further insight into possible therapeutic targets. Therefore, in Chapter 3, we employed next-generation sequencing to perform large scale analysis of genetic mutations found in pediatric GBM, with the hope to better understand its pathogenesis and improve treatment.

Chapter 3: Driver mutations in histone H3.3 and chromatin  
remodelling genes in paediatric glioblastoma

Jeremy Schwartzentruber\*, Andrey Korshunov\*, Xiao-Yang Liu\*, David  
TW Jones, Elke Pfaff, Karine Jacob, Dominik Sturm, Adam M Fontebasso,  
Dong-Anh Khuong Quang, Martje Tönjes, Volker Hovestadt, Steffen  
Albrecht, Marcel Kool, Andre Nantel, Carolin Konermann, Anders Lindroth,  
Natalie Jäger, Tobias Rausch, Marina Ryzhova, Jan O. Korbel, Thomas  
Hielscher, Peter Hauser, Miklos Garami, Almos Klekner, Laszlo Bogнар,  
Martin Ebinger, Martin U. Schuhmann, Wolfram Scheurlen, Arnulf Pekrun,  
Michael C. Frühwald, Wolfgang Roggendorf, Christoph Kramm, Matthias  
Dürken, Jeffrey Atkinson, Pierre Lepage, Alexandre Montpetit, Magdalena  
Zakrzewska, Krzysztof Zakrzewski, Pawel P. Liberski, Zhifeng Dong, Peter  
Siegel, Andreas E. Kulozik, Marc Zapatka, Abhijit Guha, David Malkin,  
Jörg Felsberg, Guido Reifenberger, Andreas von Deimling, Koichi  
Ichimura, V. Peter Collins, Hendrik Witt, Till Milde, Olaf Witt, Cindy Zhang,  
Pedro Castelo-Branco, Peter Lichter, Damien Faury, Uri Tabori, Christoph  
Plass, Jacek Majewski\*\*, Stefan M. Pfister\*\*, Nada Jabado\*\*.

\*These authors contributed equally to the manuscript.

\*\*co-senior authors on this manuscript

*Nature* (482(7384):226-31,2012)

Permission was granted to reproduce the manuscript in Chapter 3.

### 3.1. ABSTRACT

Glioblastoma (GBM) is a lethal brain tumor in adults and children. However, DNA copy number and gene expression signatures indicate differences between adult and paediatric cases (Faury, Nantel et al. 2007; Haque, Faury et al. 2007; Paugh, Qu et al. 2010; Qu, Jacob et al. 2010). To explore the genetic events underlying this distinction, we sequenced the exomes of 48 paediatric GBM samples. Somatic mutations in the H3.3-ATR<sub>X</sub>-DAXX chromatin remodeling pathway were identified in 44% of tumors (21/48). Recurrent mutations in *H3F3A*, which encodes the replication-independent histone 3 variant H3.3, were observed in 31% of tumors, and led to amino acid substitutions at two critical positions within the histone tail (K27M, G34R/G34V) involved in key regulatory post-translational modifications. Mutations in *ATR<sub>X</sub>* ( $\alpha$ -thalassemia/mental-retardation-syndrome-X-linked) (Villard, Gecz et al. 1996) and *DAXX* (death-domain associated protein), encoding two subunits of a chromatin remodelling complex required for H3.3 incorporation at pericentric heterochromatin and telomeres (Lewis, Elsaesser et al. 2010; Dhayalan, Tamas et al. 2011), were identified in 31% of samples overall, and in 100% of tumors harbouring a G34R or G34V H3.3 mutation. Somatic *TP53* mutations were identified in 54% of all cases, and in 86% of

samples with *H3F3A* and/or *ATRX* mutations. Screening of a large cohort of gliomas of various grades and histologies (n=784) showed *H3F3A* mutations to be specific to GBM and highly prevalent in children and young adults. Furthermore, the presence of *H3F3A/ATRX-DAXX/TP53* mutations was strongly associated with alternative lengthening of telomeres and specific gene expression profiles. This is, to our knowledge, the first report to highlight recurrent mutations in a regulatory histone in humans, and our data suggest that defects of the chromatin architecture underlie paediatric and young adult GBM pathogenesis.

### 3.2. INTRODUCTION

Brain tumors are currently the leading cause of cancer-related mortality and morbidity in children. Glioblastoma multiforme (GBM) is a highly aggressive brain tumor and the first cancer to be comprehensively profiled by The Cancer Genome Atlas (TCGA) consortium. While GBM is less common in the paediatric setting than in adults, affected children show dismal outcomes similar to adult patients and the vast majority will die within a few years of diagnosis despite aggressive therapeutic approaches. Tumors arise *de novo* (primary GBM) and are morphologically indistinguishable from their adult counterparts. A number of comprehensive studies have identified transcriptome-based subgroups and indicator mutations in adult GBM, and have thus enabled its molecular sub-classification (Phillips, Kharbanda et al. 2006; Parsons, Jones et al. 2008; TCGA 2008; Noushmehr, Weisenberger et al. 2010). In contrast, although we and others have demonstrated the presence of distinct molecular subsets of childhood GBM and described different genetic alterations compared to adult cases, the paediatric disease remains understudied (Faury, Nantel et al. 2007; Haque, Faury et al. 2007; Bax, Mackay et al. 2010; Paugh, Qu et al. 2010; Qu, Jacob et al. 2010). There is currently insufficient information to improve disease management, and

because conventional treatments universally fail, there is a crucial need to identify relevant targets for the design of new therapeutic agents.

### 3.3. MATERIALS AND METHODS

*Samples Characteristics and Pathological Review.* All samples were obtained with informed consent after approval of the Institutional Review Board of the respective hospitals they were treated in and were independently reviewed by senior pediatric neuropathologists (SA, AK) according to the WHO guidelines. Forty-eight pediatric grade IV astrocytomas (glioblastoma GBM) patients between the age of 1 and 20 years were included in the study. Clinical characteristics of patients are summarized in Table S3.1. Samples were taken at the time of the first surgery, prior to further treatment as needed. Tissues were obtained from the London/Ontario Tumor Bank, the Pediatric Cooperative Health Tissue Network, the Montreal Children's Hospital and from collaborators in Hungary and Germany. Seven hundred and eighty-five glioma samples from all grades and histological diagnoses across the entire age range in this study were obtained from collaborators across Europe and North America.

*Alignment and variant calling for whole exome sequencing.* We followed standard manufacturer protocols to perform target capture with the Illumina TruSeq exome enrichment kit and sequencing of 100 bp paired end reads on Illumina HiSeq. We generated approximately 10 Gb of sequence for each subject such that >90% of the coding bases of the exome defined by the consensus coding sequence (CCDS) project were covered by at least 10 reads. We removed adaptor sequences and quality trimmed reads using the Fastx toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) and then used a custom script to ensure that only read pairs with both mates present were subsequently used. Reads were aligned to hg19 with BWA (Li and Durbin 2009), and duplicate reads were marked using Picard (<http://picard.sourceforge.net/>) and excluded from downstream analyses. Single nucleotide variants (SNVs) and short insertions and deletions (indels) were called using samtools (<http://samtools.sourceforge.net/>) pileup and varFilter (Li, Handsaker et al. 2009) with the base alignment quality (BAQ) adjustment disabled, and were then quality filtered to require at least 20% of reads supporting the variant call. Variants were annotated using both Annovar (Wang, Li et al. 2010) and custom scripts to identify whether they affected

protein coding sequence, and whether they had previously been seen in dbSNP131, the 1000 genomes pilot release dataset (October 2011), or in approximately 160 exomes previously sequenced at our center.

***Somatic mutation identification for whole exome sequencing.*** A variant called in a tumor was considered to be a candidate somatic mutation if the matched normal sample had at least 10 reads covering this position and had zero variant reads, and the variant was not reported in dbSNP131 or the 1000 genomes data set (October 2011). For the resulting 117 candidate somatic mutations, we manually examined the alignment of each to check for sequencing artifacts and alignment errors. Fifteen variants were easily identified as sequence-specific error artifacts commonly seen shortly downstream of GGC sequences on Illumina sequencers (Nakamura, Oshima et al. 2011). Once genes of interest were identified (*H3F3A*, *ATRX*, *DAXX*, *TP53*, *NF1*) we examined positions in these genes in the 34 tumor samples where less than 20% of the reads supported the variant. This identified only two additional variants, both in sample PGBM19 where there were low read counts for frameshift insertions in both *ATRX* (6/32 reads) and *DAXX* (8/47 reads).

*Immunohistochemistry and immunoblotting.* Formalin-fixed, paraffin-embedded sections of pediatric GBM and TMA (4µm) were immunohistochemically stained for ATRX and DAXX proteins. Unstained sections were subjected to antigen retrieval in 10mM citrate buffer (pH6.0) for 10 minutes at sub-boiling temperatures. Individual slides were incubated overnight at 4°C with rabbit anti-ATRX (1:750 dilution, Sigma, Cat. #: HPA001906) or rabbit anti-DAXX (1:100 dilution, Sigma, Cat. #: HPA008736) antibodies. Following incubation with the primary antibody, secondary biotin-conjugated donkey anti-rabbit antibodies (Jackson) were applied for 30 minutes. After washing with PBS, slides were developed with diaminobenzidine (Dako, Mississauga, ON, Canada) as the chromogen. All slides were counterstained using Harris haematoxylin. The criterion for positive staining was described previously by Heaphy *et al* (Heaphy, de Wilde et al. 2011). IHC staining on TMA was scored by three individuals independently, including a pathologist. To test the level of mono-, di- and tri-methylated H3 at position K36, cell lysates from tumor cells were analysed by Western Blot. Antibodies against H3K36me3 (Abcam, Cat. #: ab9050), H3K36me2 (Abcam, Cat. #: ab9049), H3K36me1 (Abcam, Cat. #: ab9048) and H3.3 (Abcam, Cat. #: ab97968) were used, with conditions suggested by the manufacturer.

***Gene Expression Profiling.*** Total RNA from frozen samples was hybridized to the Affymetrix-HG-U133 plus 2.0 gene chips (Affymetrix). Array quality assurance was determined using  $\beta$ -actin and GAPDH 3'/5' ratio, as recommended by the manufacturer.

***Genome-wide SNP Array.*** DNA from 31 of the 48 pediatric GBM tumors analyzed by whole exome sequencing was hybridized to Illumina Human Omni 2.5M Single Nucleotide Polymorphism (SNP) arrays, according to the manufacturer's protocol. Copy Number Alterations were analyzed using Illumina GenomeStudio Data Analysis Software (Illumina) as previously described (Peiffer, Le et al. 2006). Statistical analysis of Fisher's exact test was performed using GraphPad Prism software.

***Telomere specific fluorescence in situ hybridization (FISH).*** Telomere specific FISH was done using a standard formalin-fixed paraffin embedded FISH protocol (as described in Heaphy et al (Heaphy, de Wilde et al. 2011)), using an FITC peptide nucleic acid telomere probe from Dako.

### 3.4. RESULTS

To decipher the molecular pathogenesis of paediatric GBM, we undertook a comprehensive mutation analysis in protein-coding genes by performing whole-exome sequencing (WES) on 48 well-characterized paediatric GBMs, including 6 patients for whom we had matched non-tumor (germline) DNA. Samples from the tumor core containing more than 90% neoplastic tissue were collected from patients aged between 3 and 20 years (Table S3.1). Coding regions of the genome were enriched by capture with the Illumina TruSeq kit and sequenced with 100-base-pairs paired-end reads on an Illumina HiSeq 2000 platform. The median coverage of each base in the targeted regions was 61-fold, and 91% of the bases were represented by at least 10 reads (Table S3.2). We identified 87 somatic mutations in 80 genes among the 6 tumors for which we had matched constitutive DNA. The mutation count per tumor ranged from 3 to 31, with a mean of 15 (Table S3.3). This is much lower than the rate observed using Sanger sequencing in other solid tumors including adult GBM (Parsons, Jones et al. 2008), but somewhat higher than in another paediatric brain tumor, medulloblastoma (Parsons, Li et al. 2011)

(Table S3.4). Relevant mutations (as defined below) were validated by Sanger sequencing.

Initially, we focused on the distribution of somatic, non-silent protein-coding mutations in the six tumors with matched germline DNA. Four samples had recurrent heterozygous mutations in *H3F3A*, which encodes the replication-independent histone variant H3.3. Both mutations were single-nucleotide variants (SNVs), in two samples changing lysine 27 to methionine (K27M), and in two samples changing glycine 34 to arginine (G34R) (Figure 3.1a and Table S3.3). These mutations are particularly interesting because histone genes are highly conserved throughout eukaryotes (Figure 3.1b), and to our knowledge no human disorders have specifically been associated with mutations in histones, including H3.3. Both mutations are at or very near positions in the amino-terminal tail of the protein that undergo important post-translational modifications associated with either transcriptional repression (K27) or activation (K36) (Figure 3.1b). All four samples had additional mutations in *ATRX*, which encodes a member of a transcription/chromatin remodelling complex required for the incorporation of H3.3 at pericentric heterochromatin and at telomeres, as well as at several transcription factor binding sites(De La

Fuente, Viveiros et al. 2004; Goldberg, Banaszynski et al. 2010; Lewis, Elsaesser et al. 2010; Wong, McGhie et al. 2010; Iwase, Xiang et al. 2011). We extended our WES analysis to 42 additional tumor samples and focused on *ATRX* and *H3F3A*, as well as *DAXX* (because the gene product heterodimerizes with ATRX and participates in H3.3 recruitment to DNA (Lewis, Elsaesser et al. 2010; Dhayalan, Tamas et al. 2011)). A total of 15 samples had heterozygous H3.3 mutations (9 K27M, 5 G34R, 1 G34V) and 14 samples had a mutation in *ATRX*, including frameshift insertions/deletions (6 samples), gains of a stop codon (4 samples), and missense SNVs (4 samples). Nearly all of the *ATRX* mutations occurred either within the carboxy-terminal helicase domain or led to truncation of the protein upstream of this domain (Figure 3.1c). Mutations were accompanied by an absence of detectable ATRX protein by immunohistochemistry in samples for which paraffin material was available (Figure S3.1). Two samples had heterozygous *DAXX* mutations, simultaneously with an *ATRX* mutation in one sample (Figure 3.1a and Table S3.3). Overall, 21 of 48 samples (44%) had a mutation in at least one of these three genes. Notably, we also identified *TP53* mutations in 26 samples (25 somatic, 1 germline in PGBM26), which overlapped significantly with samples that had *ATRX*, *DAXX* and/or *H3F3A* mutations

(18/21 cases, 86%, Figure 3.1d;  $P=1.1\times 10^{-4}$ , permutation test). A list of all mutations discovered by WES in selected genes associated with GBM is given in Table S3.5.

*H3F3A*, *ATRX* or *DAXX* were not part of the 600 genes sequenced by The Cancer Genome Atlas (TCGA) glioblastoma project (TCGA 2008; Verhaak, Hoadley et al. 2010), and no *H3F3A* mutations were identified in 22 adult GBM samples sequenced previously (Parsons, Jones et al. 2008). To investigate whether *H3F3A* mutations are specific to GBM and/or paediatric disease, we sequenced this gene in 784 glioma samples from all grades and histological diagnoses across the entire age range (Figure 3.2a). H3.3 mutations were highly specific to GBM and were much more prevalent in the paediatric setting (32/90, 36%), although they also occurred rarely in young adults with GBM (11/318, 3%) (Figure 3.2b). K27M-H3.3 mutations occurred mainly in younger patients (median age 11 years, range 5-29) and thalamic GBM (Table S3.1), whereas G34R- or G34V-H3.3 mutations occurred in older patients (median age 20 years, range 9-42) and in tumors of the cerebral hemispheres (Figure 3.2b). Further comparison of our data set with adult GBM databases (Parsons, Jones et al. 2008; TCGA 2008; Noushmehr, Weisenberger et al. 2010;

Verhaak, Hoadley et al. 2010; Jiao, Shi et al. 2011), indicated limited overlap in frequently mutated genes between paediatric GBM and any of the four previously described adult GBM subtypes (Verhaak, Hoadley et al. 2010) (Figure 3.2c, Figure S3.2 and Table S3.6).

Somatic mutations in *ATRX* and *DAXX* have recently been reported in a large proportion (43%) of pancreatic neuroendocrine tumors (PanNETs), a rare form of pancreatic cancer with a 10-year overall survival of ~40%, and no reported association with *TP53* or *H3F3A* mutations (Jiao, Shi et al. 2011). A follow-up study found *ATRX* mutations in a series of cancers, including GBM, where *ATRX* (but not *DAXX*) mutations were identified in 3/21 paediatric GBMs (14%) and 8/122 adult GBMs (7%) (Heaphy, de Wilde et al. 2011). To evaluate further the prevalence of *ATRX* and *DAXX* mutations in paediatric GBM, we performed immunostaining for these proteins on a well-characterized tissue microarray (TMA) with samples from 124 paediatric GBM patients. Lack of immuno-positivity for *ATRX* was seen in 35% of cases (40/113 scored, 22 females and 18 males) and for *DAXX* in 6% (7/124 scored) (Figure 3.2d and Figure S3.1). Overall, 37% of samples had lost nuclear expression of either factor, corroborating our WES findings. Strikingly, *ATRX-DAXX* mutations (as assessed by

direct sequencing or loss of protein expression) were found in 100% of G34-H3.3 mutant cases in the larger cohort of GBMs (13/13) where sufficient material was available ( $P = 1.4 \times 10^{-8}$ , permutation test). The overlap of *ATRX* mutations with K27M-H3.3-mutated samples was not significant in either the exome data set (3/9 samples,  $P = 0.58$ ) or the full set of GBM screened (5/13,  $P = 0.40$ ) (Figure 3.2e).

The histone code – post-translational modifications of specific histone residues – regulates virtually all processes that act on or depend on DNA, including replication and repair, regulation of gene expression, and maintenance of centromeres and telomeres (Chi, Allis et al. 2010). Accordingly, although recurrent histone mutations have not previously been reported in cancer, mutations in genes affecting histone post-translational modifications are increasingly described (Fullgrabe, Kavanagh et al. 2011). H3.3 is a universal, replication-independent histone predominantly incorporated into transcription sites and telomeric regions, and associated with active and open chromatin (Talbert and Henikoff 2010). This role is conserved in the single histone H3 present in yeast, indicating its importance throughout evolution. It functions as a neutral replacement histone, but also participates in the epigenetic

transmission of active chromatin states and is associated with chromatin assembly factors in large-scale replication-independent chromatin remodelling events (Talbert and Henikoff 2010).

The non-random recurrence of the exact same mutation in different tumors, and the absence of truncating mutations, indicate that *H3F3A* mutations are most probably gain-of-function events. Lysine 27 is a critical residue of histone 3 and its variants, and methylation at this position (H3K27me), which may be mimicked by the terminal CH<sub>3</sub> of methionine substituted at this residue, is commonly associated with transcriptional repression (Bernstein, Mikkelsen et al. 2006). In contrast, H3K36 methylation of acetylation typically promotes gene transcription (Edmunds, Mahadevan et al. 2008; Kolasinska-Zwierz, Down et al. 2009). Thus, although their morphological phenotype is very similar (K27M and G34R/V mutant tumors are histologically indistinguishable), the two H3.3 variants are expected to act through a different set of genes. This indeed seems to be the case when looking at expression profiles of GBMs harbouring these two mutations. Unsupervised hierarchical clustering of gene expression from 27 of the WES cohort samples for which sufficient RNA was available revealed a clear separation in the expression of K27M versus G34R/V

mutant samples (Figure S3.3). Further analysis of just those samples harbouring an *H3F3A* mutation additionally showed a clear distinction in the expression pattern of these two variants (Figure 3.3a and Table S3.7). Among these differentially expressed genes were several linked to brain development that showed a clear mutation-specific expression pattern when comparing both between K27 and G34 mutants and with H3.3 wild-type GBMs, including *DLX2*, *SFRP2*, *FZD7* and *MYT1* (Figure 3.3b). We also identified increased levels of H3K36 trimethylation in cells carrying the G34V-H3.3 mutation in one sample for which we had available material (PGBM14) compared to other cells, potentially supporting this hypothesis (Figure S3.5).

ATRX loss, frequently observed in this study, has recently been shown to be associated with alternative lengthening of telomeres (ALT) in PanNETs and GBMs (Heaphy, de Wilde et al. 2011). We performed telomere-specific fluorescence *in situ* hybridization (FISH) on the samples with K27M or G34R/V mutations identified by WES for which we had slides available (Figure S3.4) and on the paediatric GBM TMA (Figure 3.3c). These experiments showed that ALT is strongly correlated with ATRX loss (37/47 samples with ALT showed ATRX loss,  $P < 0.001$ ). However, some

samples with nuclear ATRX staining still showed ALT, indicating that additional defects may also account for elongated telomeres in GBM. The presence of ALT was best explained by the simultaneous presence of *ATRX/H3F3A/TP53* mutations ( $P = 0.0002$ , Fisher's exact test). Tumors without *ATRX/H3F3A/TP53* mutations almost invariably showed shorter telomeres than are observed with ALT, as seen in telomerase-positive gliomas (Hakin-Smith, Jellinek et al. 2003).

Genetic stability was also assessed through evaluating DNA copy number aberrations (CNAs) in 31 of the 48 tumors using Illumina SNP arrays containing ~2.5 million oligonucleotides (Table S3.1, Table S3.8, Table S3.9). Loss of heterozygosity (whole chromosome changes, broad and focal heterozygous deletions, Table S3.9) was common in paediatric GBM samples, as we have previously reported (Qu, Jacob et al. 2010), and the focal gains and losses we identified in our study showed a high degree of overlap with other published paediatric data sets (Paugh, Qu et al. 2010). The number of CNAs per tumor was higher in samples with *H3F3A/ATRX-DAXX/TP53* mutations (Figure S3.6).

Recurrent point mutation in *IDH1* (mainly R132H) are gain of function mutations commonly identified in secondary GBM and the lower-grade tumors from which they develop (86-98% of these astrocytomas), and typically occur in younger adults (Parsons, Jones et al. 2008; Yan, Parsons et al. 2009). Strikingly, *IDH1* and *H3F3A* mutations were mutually exclusive in our sequencing cohort ( $P = 1.6 \times 10^{-4}$ ). Neomorphic enzyme activity resulting from *IDH1* mutation leads to the production of high quantities of the onco-metabolite 2-hydroxyglutarate (2-HG) (Dang, White et al. 2010). Increased 2-HG inhibits histone demethylases, specifically inducing increased methylation of both H3K27 and H3K36 (Cervera, Bayley et al. 2009; Dang, White et al. 2010), the two residues affected directly (K27) or indirectly (K36) by the mutations in *H3F3A* uncovered in this study. Furthermore, overlap of *H3F3A* and *TP53* mutations in children with GBM (all of the G34R/V and 82% of K27M mutations also harbor *TP53* mutations) mirrors the large overlap of *IDH1* mutations with *TP53* mutation in the proneural adult GBM subgroup (Verhaak, Hoadley et al. 2010). Thus, mutations which directly (*H3F3A*), or indirectly (*IDH1*) affect the methylation of H3.3 K27 or H3.3 K36, in combination with *TP53* mutations, characterize the pathogenesis of pediatric and young adult GBM.

### 3.5. SUMMARY

Our data indicate a central role of H3.3/ATRX-DAXX perturbation in pediatric GBM. Mutant H3.3 recruitment would occur across the genome and induce abnormal patterns of chromatin remodeling to yield distinct gene expression profiles for the K27 and G34 mutations. Additional loss of ATRX may act to reduce H3.3 incorporation at a subset of genes important in oncogenesis, preventing mutant H3.3 from altering their transcription. ATRX loss will also impair H3.3 loading at telomeres and disrupt their heterochromatic state, facilitating alternative lengthening of telomeres (ALT). Our findings provide an intriguing example of the interplay of genetic and epigenetic events in driving cancer, indicate a new mechanism through which these epigenetic alterations are brought about (mutation of key residues in a regulatory histone), and provide a rationale for targeting the chromatin remodeling machinery in this deadly pediatric cancer.

### 3.6. ACKNOWLEDGEMENTS

The authors are indebted to J. Rak, N. Sonenberg and C. Polychronakos for critical reading of this manuscript. D. M. Pearson, A. Wittmann, L. Sieber and L. Senf are acknowledged for technical assistance. This work was supported by the Cole Foundation, and was funded in part by Genome Canada and the Canadian Institute for Health Research (CIHR) with co-funding from Genome BC, Genome Quebec, CIHR-ICR (Institute for Cancer Research) and C17, through the Genome Canada/CIHR joint ATID Competition (project title: The Canadian Paediatric Cancer Genome Consortium: Translating next generation sequencing technologies into improved therapies for high-risk childhood cancer (NJ)), the Hungarian Scientific Research Fund(OTKA) ContractNo. T-04639, the National Research and Development Fund (NKFP) Contract No. 1A/002/2004 (P.H., M.G., L.B.), the PedBrain project contributing to the International Cancer Genome Consortium funded by the German Cancer Aid (109252) and the CNS tumor tissue bank within the priority program on tumor tissue banking of the German Cancer Aid (108456), the BMBF, the Samantha Dickson Brain Tumor Trust, a grant from the National Cancer Center Heidelberg (“Paediatric Brain Tumor Preclinical Testing”), and a guest scientist stipend to M.R. from the German Cancer

Research Center. X.-Y.L. and A.M.F. are the recipients of studentship awards from CIHR. K.J. is the recipient of a studentship from the Foundation of Stars. S.M.P. is the recipient of the Sybille Assmus Award for Neurooncology in 2009 and N.J. is the recipient of a Chercheur Boursier Award from Fonds de Recherche en Santé du Québec.

### 3.7. AUTHOR CONTRIBUTIONS

A.K., X.-Y.L., D.T.W.J., E.P., D.-A.K.Q., S.A., M.T., Z.D., P.S., N. Jager, H.W., T.M., C.Z., U.T., P.C.-B., D.F., P.L. and A.M. performed experimental work. X.-Y.L., J.S., J.M., D.T.W.J., M.K., D.M., V.H., K.J., A.M.F., D.S., A.N., C.K., A.L., A.K., T.R., J.M., J.O.K., T.H., P.L., C.P., S.M.P. and N. Jabado performed data analyses and produced the text and figures. M.R., M.E., M.U.S., W.S., A.P. and M.C.F., W.R., C.K., M.D., J.A., P.H., M.G., L.B., P.P.L., M. Zakrzewska, K.Z., A.E.K., M. Zapatka, A.G., A.K., G.R., J.F., A.v.D., K.I., V.P.C. and O.W. collected data and provided patient materials. J.M., S.M.P. and N. Jabado provided leadership for the project. All authors contributed to the final manuscript.

### 3.8. FIGURES

#### Figure 3.1: Most frequent mutations in paediatric GBM.

a, Most frequent somatic mutations in 48 paediatric glioblastoma tumors.

Mutations identified in genes listed in this table were confirmed by Sanger sequencing, and were not present in dbSNP nor in the 1000 Genomes data set (October 2011), except for the *TP53* SNP at R273, which is

associated with cancer. Detailed description of the mutations in affected

samples is provided in Supplementary Table 5. b, Three recurrent non-synonymous single nucleotide variants (SNVs) were observed in *H3F3A*.

The K27M, G34R and G34V mutations are shown in the context of the common post-translational modifications of the H3.3 N-terminal tail, which regulates the histone code. H3.3 has 136 amino acids, and is highly conserved across species from mammals to plants, including the residues subject to mutation in paediatric GBM (see multiple alignment of amino

acids 11 to 60). c, Schematic of the mutations observed in *ATRX* in the

48 WES samples. d, Schematic of the overlap between mutations

affecting *ATRX-DAXX*, *H3F3A* and *TP53*. Eight samples had all three mutations.

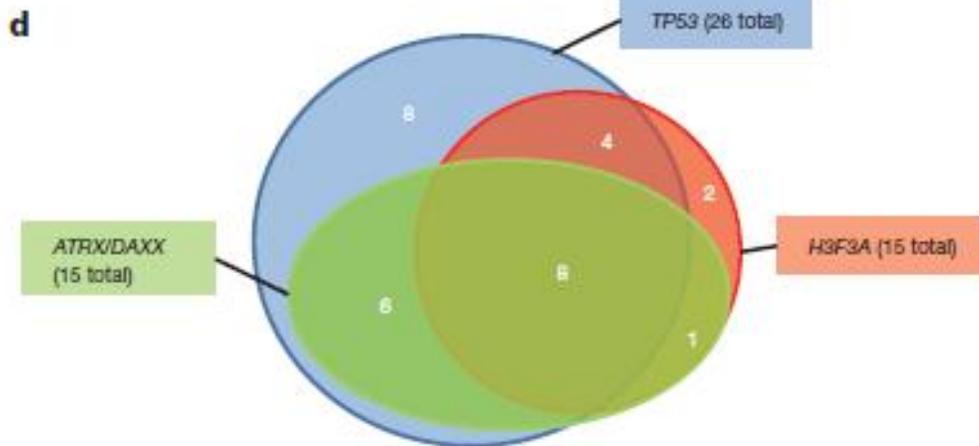
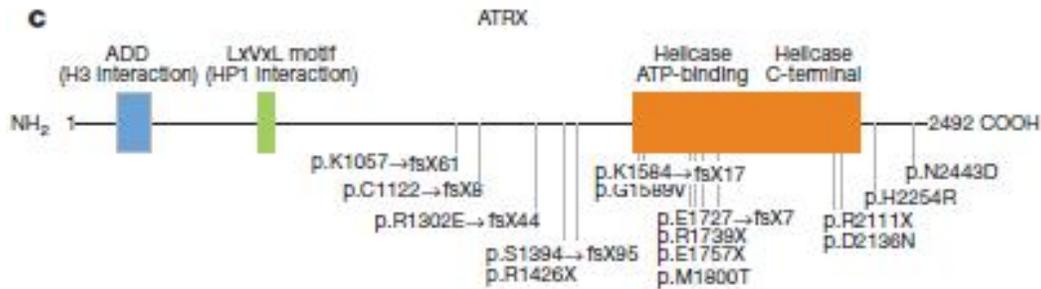
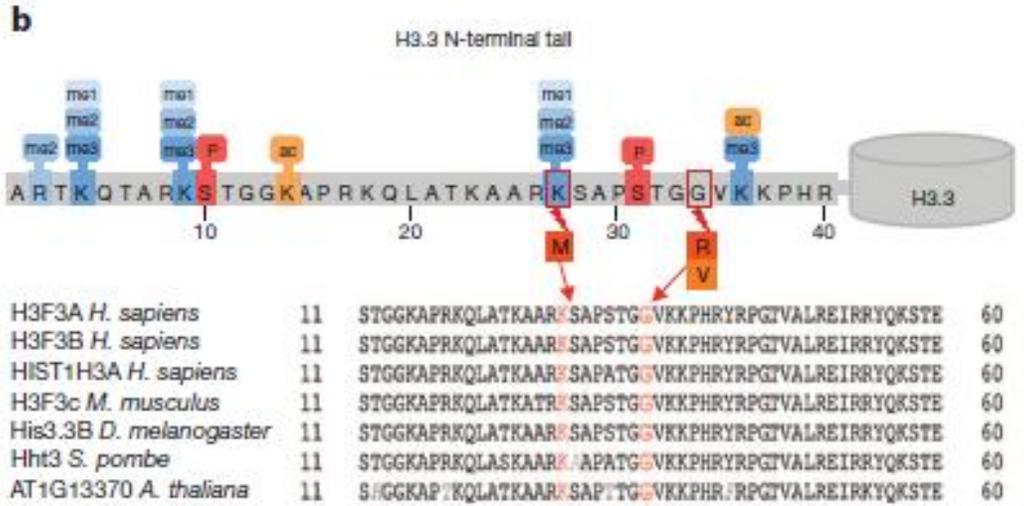
**a**

Sample ID	H3F3A	ATRX/DAXX	TP53	IDH1	NF1	PDGFRA
PGBM1	K27M	C1122fs	P152fs;R306X	WT	WT	WT
PGBM2	K27M	WT	R213X*	WT	WT	K385I
PGBM3	K27M	WT	N131del	WT	WT	WT
PGBM4	K27M	K1057fs*	G262fs*	WT	WT	WT
PGBM5	K27M	WT	WT	WT	Y2264fs	WT
PGBM6	K27M	M1800T*	WT	WT	F1247fs;V2230del	WT
PGBM8	K27M	WT	R273C	WT	WT	WT
PGBM9	K27M	WT	R273P*	WT	WT	WT
PGBM10	K27M	WT	WT	WT	T990fs	WT
PGBM11	G34R	S1394fs*	Y163C*	WT	WT	WT
PGBM12	G34R	E1727fs	R342X;R175H	WT	WT	Y849D*
PGBM13	G34R	R1739X*	T256fs*	WT	WT	WT
PGBM14	G34R	E1757X*	R273C;R248Q	WT	WT	WT
PGBM15	G34R	H2254R*	S51delins	WT	WT	WT
PGBM16	G34V	R2111X*	R342X*	WT	WT	K385M
PGBM17	WT	G1589V*	Y220C	R132H	WT	WT
PGBM18	WT	R1426X*	R273C;R196X	R132H	C622X;L1489fs	WT
PGBM19	WT	K1584fs†	R267Q;T230I	WT	R440X	WT
PGBM20	WT	N2443D*	R248W*	WT	WT	WT
PGBM21	WT	R238X (DAXX)	R267W;P152L	WT	R1947X	WT
PGBM22	WT	R1302fs;K1584fs‡	R337C;R175H	WT	R2616X;R461Xs‡	WT
PGBM23	WT	WT	I254S	R132H	WT	WT
PGBM24	WT	WT	R196X*	WT	WT	WT
PGBM25	WT	WT	R342X*	WT	G1526fs*	WT
PGBM26	WT	WT	R175H*	WT	Y2264fs*	WT
PGBM27	WT	WT	I251L*	WT	WT	WT
PGBM28	WT	WT	R273H	WT	T676fs	WT
PGBM29	WT	WT	V10G	R132H	WT	WT
PGBM30	WT	WT	G245S*	WT	WT	WT
PGBM31	WT	WT	WT	WT	WT	WT
PGBM32	WT	WT	WT	WT	T1627S	WT
PGBM33	WT	WT	WT	WT	Splicing	WT
PGBM34	WT	WT	WT	WT	WT	D842_I853delinsV
PGBM35	WT	WT	WT	WT	WT	WT
---	---	---	---	---	---	---
PGBM49	WT	WT	WT	WT	WT	WT

\*Homozygous mutations.

†Sample PGBM19 additionally has a DAXX mutation C629Sfs, whereas PGBM21 has no ATRX mutation but has the DAXX mutation shown.

‡Sample PGBM22 has a third ATRX mutation, p.D2136N, and a third NF1 mutation, p.A887T.



**Figure 3.2: Mutations in *H3F3A*, *ATRX* and *DAXX* distinguish paediatric from adult GBM.**

a, H3F3A mutations in a set of 784 gliomas from all ages and grades.

*H3F3A* mutations are exclusive to high-grade tumors and the vast majority occur in glioblastoma (GBM) and in the paediatric setting. A, diffuse

astrocytoma grade II; AA, anaplastic astrocytoma; AO, anaplastic

oligodendroglioma; AOA, anaplastic oligoastrocytoma; O,

oligodendroglioma; OA, oligoastrocytoma; PA, pilocytic astrocytoma. b,

H3.3 mutations are specific to pediatric and young adult glioblastoma

(GBM). K27M-H3.3 mutations occur mainly in younger patients (median

age 11 years) and G34R/V-H3.3 mutations occur in older children and

young adults (median age 20 years). No H3.3 mutations were identified in

older patients with GBM. c, Comparison of the most frequently mutated

genes in paediatric and adult GBM shows that *H3F3A*, *ATRX* and *DAXX*

mutations are largely specific to paediatric disease. Except for similarities

in the mutation rate for *TP53* and *PDGFRA* with the previously identified

proneural adult GBM subgroup, the rate and type of genes mutated were

distinct between paediatric and adult GBM whatever the molecular

subgroup (Verhaak, Hoadley et al. 2010) (Figure S3.2). Data for adult

GBM regarding other genes included in the table was compiled from ref

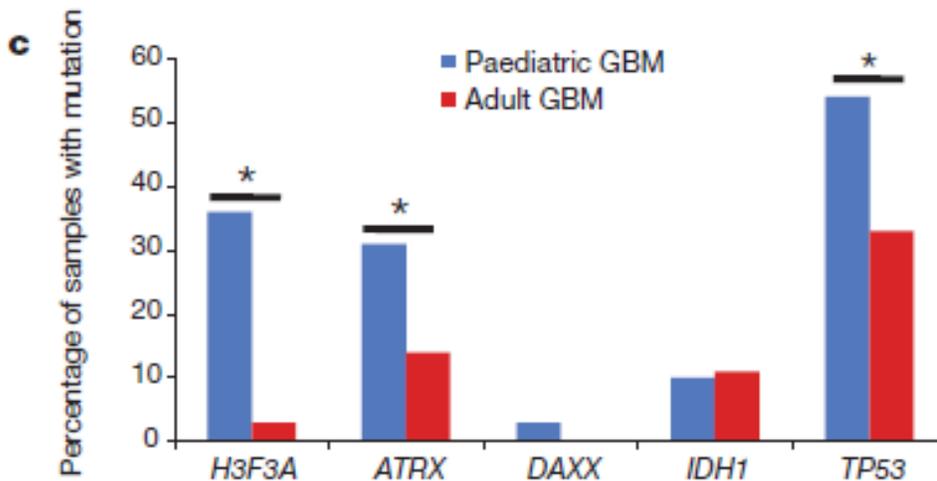
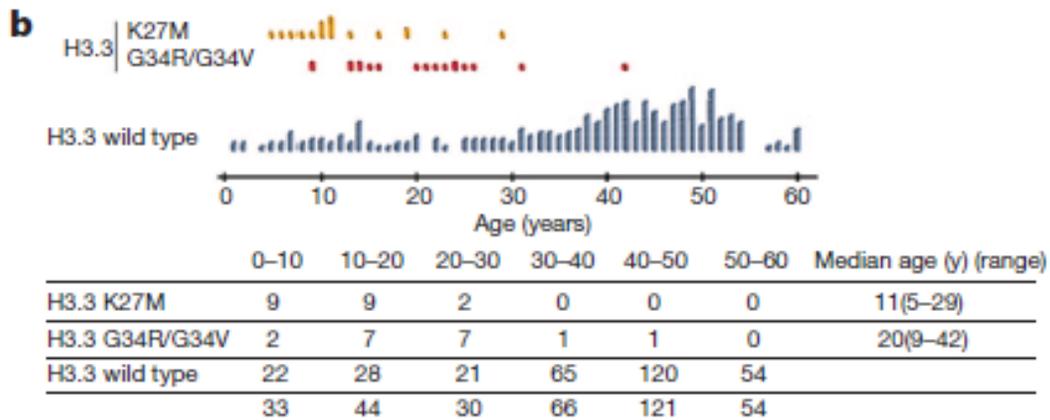
(Parsons, Jones et al. 2008) and (Verhaak, Hoadley et al. 2010). d. ATRX and DAXX immunohistochemical staining of a paediatric GBM tissue microarray (TMA) comprising 124 samples. View of the TMA slide and an example of a negative and of a positive core at high magnification to show specific nuclear staining (or lack thereof) for DAXX and ATRX. No gender bias for ATRX loss was observed. Overall survival and progression-free survival were similar in patients with and without loss of ATRX and/or DAXX (data not shown). e, Differential association of K27M and G34R/V H3F3A mutations with ATRX mutations. G34R/V-H3.3 mutations were always associated with ATRX mutations (two-sided Fisher's exact test,  $P = 0.0016$ ), whereas a non-significant overlap was observed for K27M.

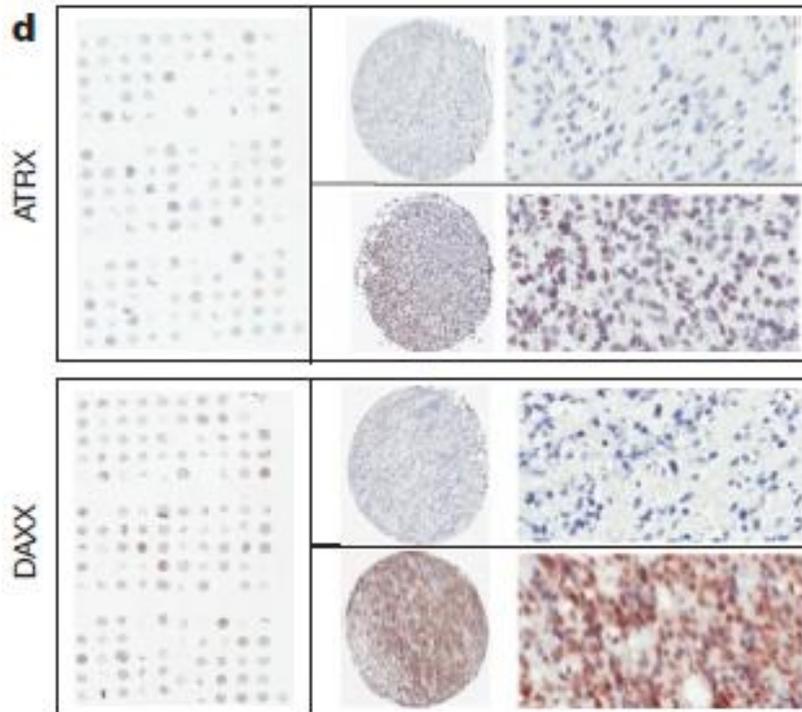
**a**

Grade		Diagnosis	Total no. of cases	No. H3F3A mut.	P-value
High grade	Grade IV	Adult GBM	318	11 (3.4%)	<0.0001*
		Paediatric GBM	90	32 (35.6%)	
	Grade III	Adult AA	108	0 (0%)	
		Paediatric AA	11	2 (18.2%)	
		Adult AO	34	0 (0%)	
		Paediatric AO	5	0 (0%)	
Grade II	Adult AOA	37	0 (0%)		
	Paediatric AOA	2	0 (0%)		
Low grade	Grade II	Adult A	57	0 (0%)	
		Paediatric A	10	0 (0%)	
		Adult O	41	0 (0%)	
	Grade I	Paediatric O	2	0 (0%)	
		Adult OA	23	0 (0%)	
		Paediatric OA	5	0 (0%)	
Grade I	Adult PA	7	0 (0%)		
	Paediatric PA	34	0 (0%)		
		Total GBM	408	43 (10.5%)	<0.0001†
		Total non-GBM	376	2 (0.5%)	

\*Fisher's two-tailed exact test between paediatric and adult groups.

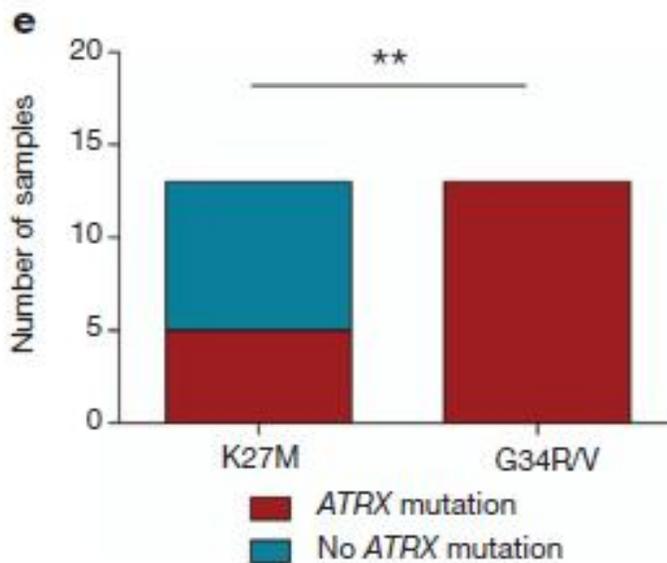
†Fisher's two-tailed exact test between GBM group (including paediatric and adult) and lower grade astrocytomas (including grade I, II and III).





	No. scored	No. negative	% negative
ATRX	113	40	35%
DAXX	124	7	6%
ATRX and/or DAXX	110	41*	37%

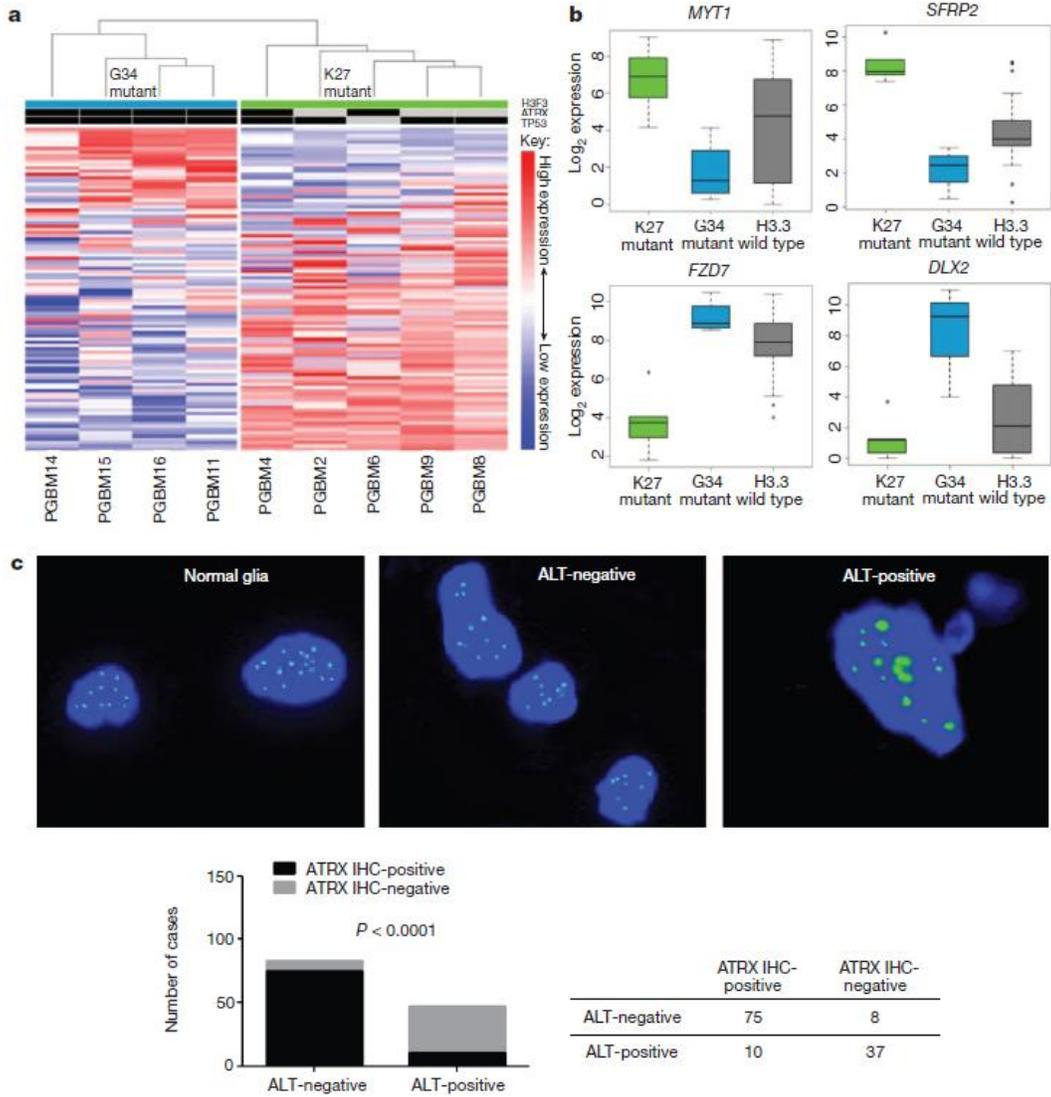
\* Six samples have negative staining for both ATRX and DAXX.



**Figure 3.3: H3F3A mutation variants show distinct expression profiles and are associated with alternative lengthening of telomeres.**

a, Unsupervised hierarchical clustering of differentially expressed genes in 27 of the 48 GBM samples analysed by whole-exome sequencing shows that samples with K27M and G34R/V H3.3 have specific gene expression profiles. Clustering was based on the top 100 genes by standard deviation from autosomal genes detected as present in 10% of samples (see also Figure S3.3). b, Genes involved in development and differentiation show H3.3 mutation-specific expression patterns. Expression levels of developmental-related genes including *DLX2*, *SFRP2*, *FZD7* and *MYT1* are distinct among H3.3-K27 mutant and H3.3-G34 mutants following gene expression profiling (see also Table S3.7). c, Alternate lengthening of telomere (ALT) is associated with the presence of mutant H3F3A/ATRX in a tissue microarray (TMA) comprising 124 paediatric GBM samples. We assessed ALT using telomere-specific FISH (shown here and in Figure S3.4) on the paediatric TMA we investigated for ATRX expression (Figure 3.2d) and using telomere-specific Southern blotting of high molecular weight genomic DNA (data not shown). Fisher's exact test was used to identify any association relationship. Representative images

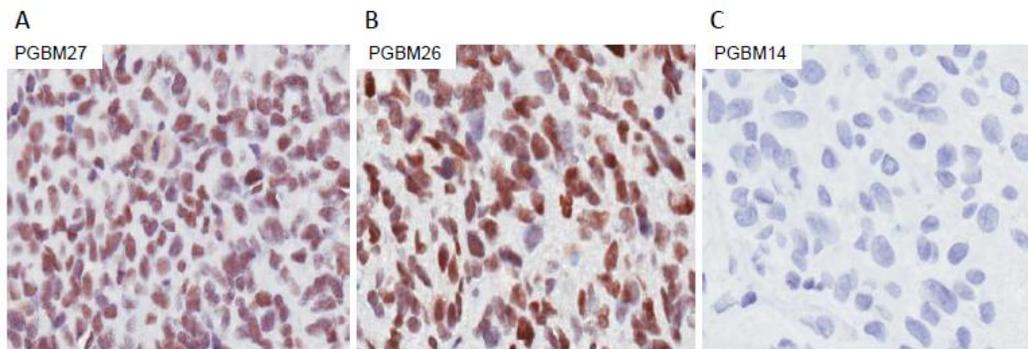
of ALT-positive and -negative staining of a paediatric GBM tissue microarray and a control brain are provided.



### 3.9. SUPPLEMENTARY FIGURES

Figure S3.1: Immunohistochemical staining for ATRX showing concordance between sequencing data and protein expression in samples.

A). & B). ATRX is expressed in these two samples (PGBM27 & PGBM26) with wild type *ATRX* following whole exome sequencing. C). ATRX is not expressed in PGBM14 where mutations in *ATRX* were identified following whole exome sequencing.



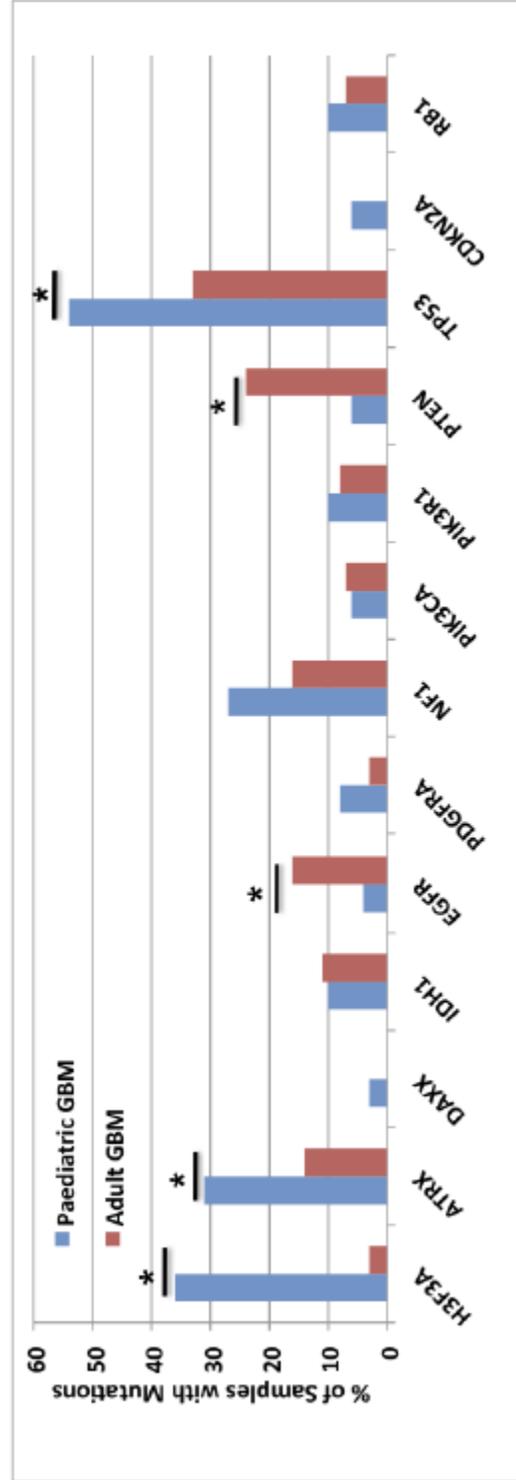
**Figure S3.2a: Comparison of mutated genes between paediatric and adult GBM.**

*ATRX* & *DAXX* mutation data for paediatric GBM were from WES, Sanger sequencing and *ATRX* IHC analysis described in this study, and Heaphy *et al* (Heaphy, de Wilde et al. 2011). Other paediatric GBM results were from our study. *H3F3A* mutation data for adult GBM were obtained using Sanger sequencing. Adult GBM *ATRX* results were calculated from our study and Heaphy *et al* (Heaphy, de Wilde et al. 2011). Datasets described in the literature (Parsons, Jones et al. 2008; Verhaak, Hoadley et al. 2010; Heaphy, de Wilde et al. 2011) were used for all the other results of adult GBM.

Pathway	Gene	Paediatric GBM		Adult GBM		p-value <sup>†</sup>
		No. of tumors	% of tumors	No. of tumors	% of tumors	
Chromatin Remodelling	H3F3A	32/90	36	11/318	3	<0.0001
	ATRX	59/190	31	23/161	14	0.0002
	DAXX	2/69	3	0/217*	0*	0.0576
	IDH1	8/83	10	24/221	11	0.8367
Cell Signalling	EGFR	2/48	4	35/221	16	0.0362
	PDGFRA	4/48	8	4/116	3	0.234
	NF1	13/48	27	36/221	16	0.0979
	PIK3CA	3/48	6	16/221	7	1
	PIK3R1	5/48	10	18/221	8	0.575
Cell Cycle	PTEN	3/48	6	54/221	24	0.0034
	TP53	26/48	54	73/221	33	0.008
	CDKN2A	3/48	6	0/22	0	0.5467
	RB1	5/48	10	15/221	7	0.3694

<sup>†</sup>Fisher's two-tailed comparison test for paediatric and adult GBM.

\*105 of 217 samples were sequenced without DAXX mutation reported, 112 of 217 samples don't have DAXX mutation



**Figure S3.2b: Comparison of frequently mutated genes indicated that paediatric GBM is distinct from the previously identified molecular subgroups in adult GBM**

(Verhaak, Hoadley et al. 2010).

Dark blue indicates mutations in *H3F3A*, *ATRX* and *DAXX* we identified and specific to paediatric GBM; *IDH1* (light blue) mutations are representative of proneural subgroup of adult GBM and not paediatric GBM. Mutations in *PDGFRA* and *TP53* in children had similar rates to adult proneural GBM while *NF1* and *RB1* mutations were more similar to the mesenchymal subgroup (green). Bar graphs showed limited overlap between paediatric and adult GBM. Fisher's t-test used to compare subgroups. \*p-value<0.05. NA=not available.

Adult GBM Gene-expression based molecular subtypes																
		Paediatric GBM			Proneural			Neural			Classical			Mesenchymal		
Pathway	Gene	No. of tumors	% of tumors	No. of tumors	% of tumors	p-value*	No. of tumors	% of tumors	p-value*	No. of tumors	% of tumors	p-value*	No. of tumors	% of tumors	p-value*	
Chromatin Remodelling	H3F3A	32/90	36	NA	NA	NA										
	ATRX	59/190	31	NA	NA	NA										
	DAXX	2/69	3	NA	NA	NA										
Cell Signalling	IDH1	8/83	10	11/37	30	0.0125	1/19	5	1	0/22	0	0.1993	0/38	0	0.05554	
	EGFR	2/48	4	6/37	16	0.0734	5/19	26	0.0166	7/22	32	0.0032	2/38	5	1	
	PDGFRA	4/48	8	4/37	11	0.7235	0/19	0	0.5713	0/22	0	0.3008	0/38	0	0.1264	
	NF1	13/48	27	2/37	5	0.0102	3/19	16	0.5259	1/22	5	0.0499	14/38	37	0.3585	
	PIK3CA	3/48	6	3/37	8	1	1/19	5	1	1/22	5	1	1/38	3	0.6266	
Cell Cycle	PIK3R1	5/48	10	7/37	19	0.3498	2/19	11	1	1/22	5	0.6572	0/38	0	0.0636	
	PTEN	3/48	6	6/37	16	0.1687	4/19	21	0.0937	5/22	23	0.098	12/38	32	0.0033	
	TP53	26/48	54	20/37	54	1	4/19	21	0.0162	0/22	0	<0.0001	12/38	32	0.0494	
	CDKN2A	3/48	6	NA	NA	NA										
	RB1	5/48	10	1/37	3	0.2261	1/19	5	0.6666	0/22	0	0.1727	5/38	13	0.7442	

\*Fisher's two tailed test to compare between paediatric GBM and different subtypes of adult GBM (Proneural, Neural, Classical, Mesenchymal).

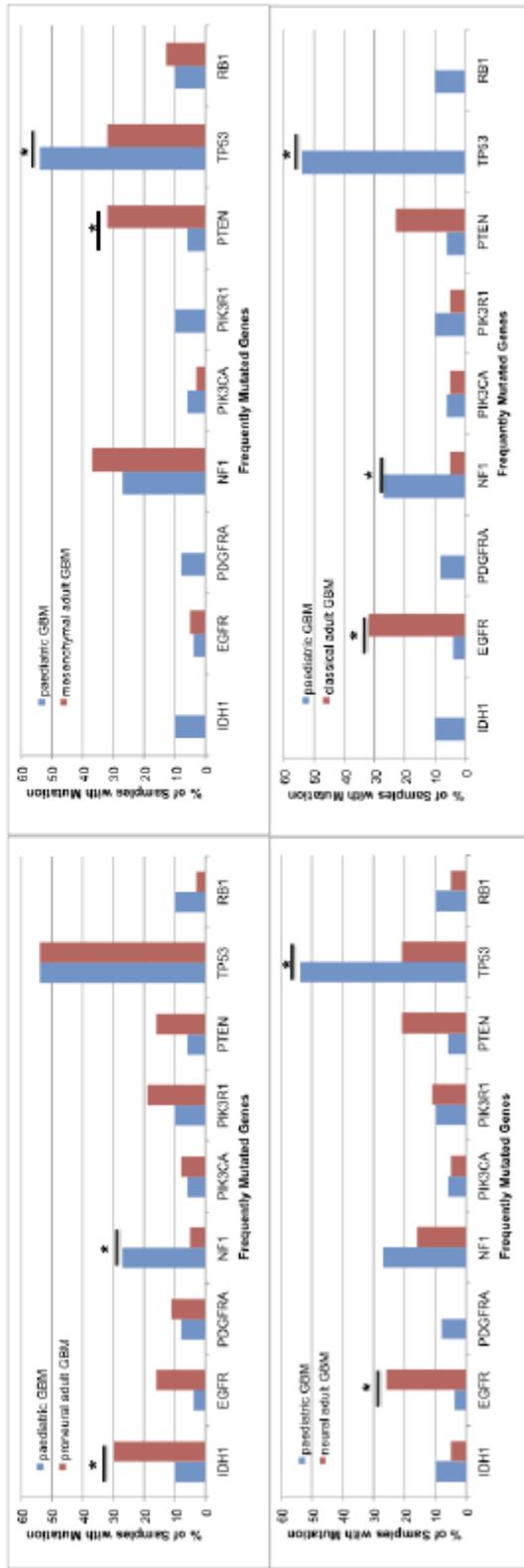
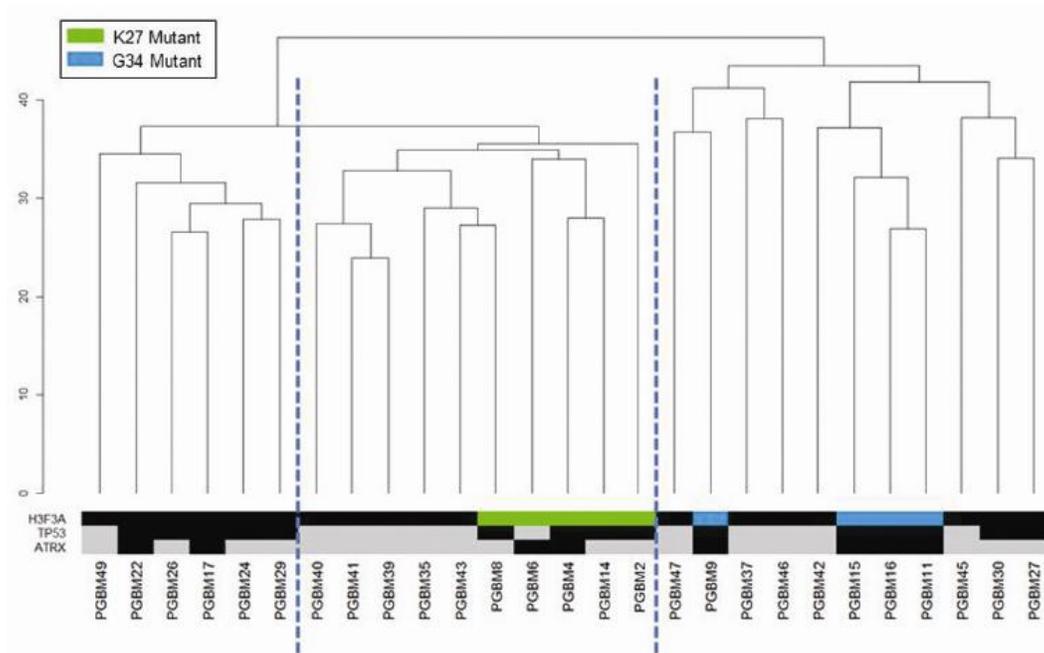


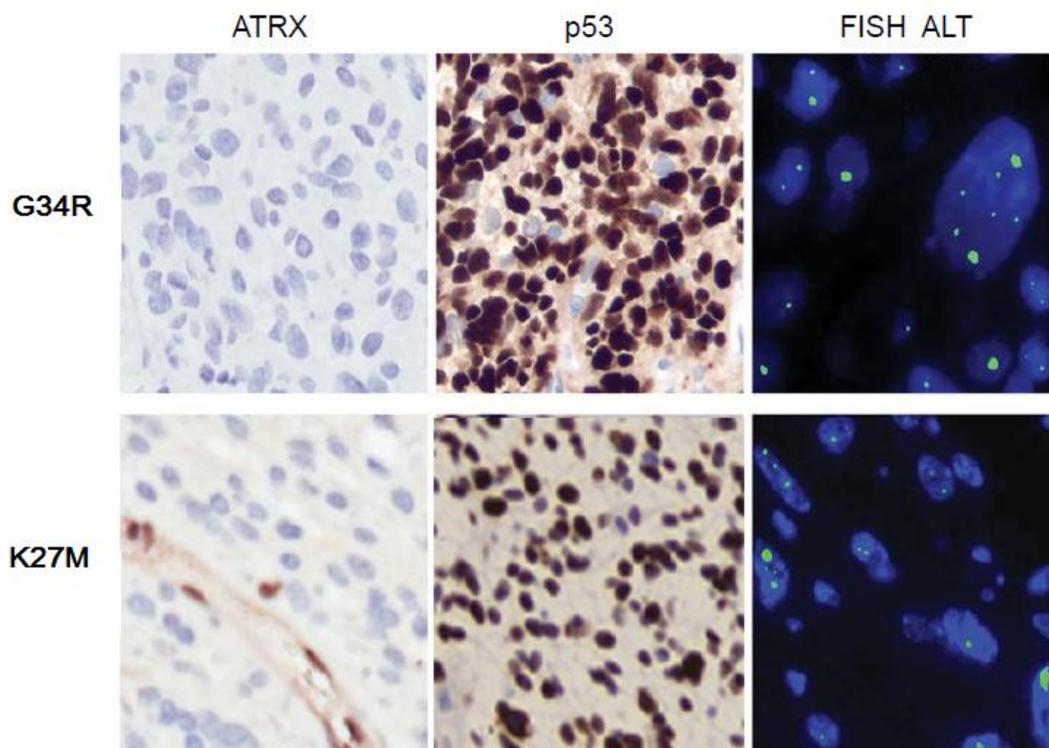
Figure S3.3: Unsupervised clustering of the 100 genes most differentially regulated in 27 paediatric glioblastoma also analyzed by whole exome sequencing.

Samples with K27M H3.3 and G34V H3.3 mutations cluster separately from each other and from samples wild type for H3.3.



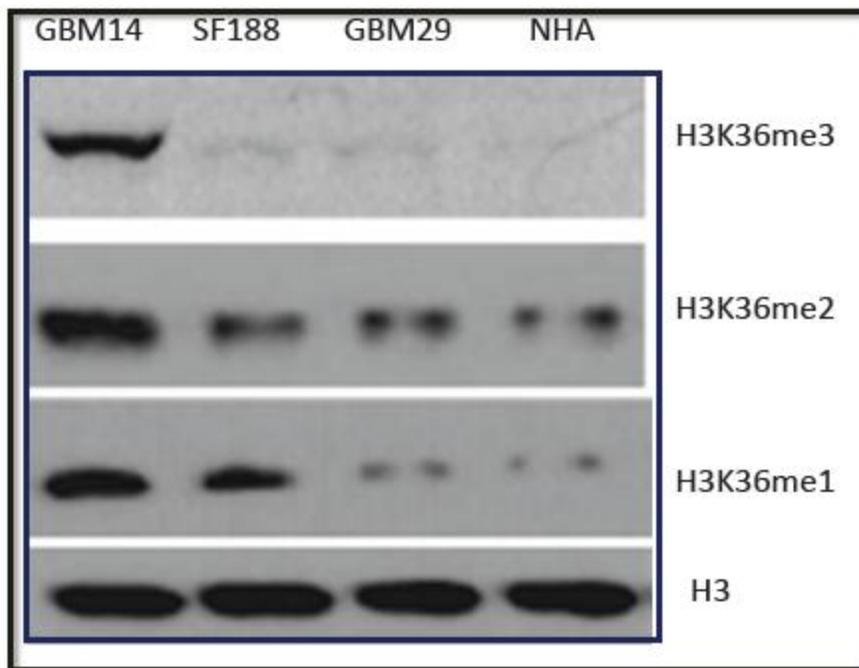
**Figure S3.4: Immunohistochemical staining for ATRX & p53 and ALT FISH of patients with G34V or K27M-H3.3.**

Immunohistochemical staining for ATRX (left panel), p53 (middle panel) of consecutive slides from representative patients with G34R (upper panel) or K27-H3.3 (lower panel) shows homogeneous loss of ATRX and abnormal p53 across the vast majority of tumor cells. Alternative lengthening of telomeres was present in a large proportion of tumor cells shown here by fluorescence *in situ* hybridization (right panel).



**Figure S3.5: H3.3 Lysine 36 is methylated in a G34R mutant (PGBM14).**

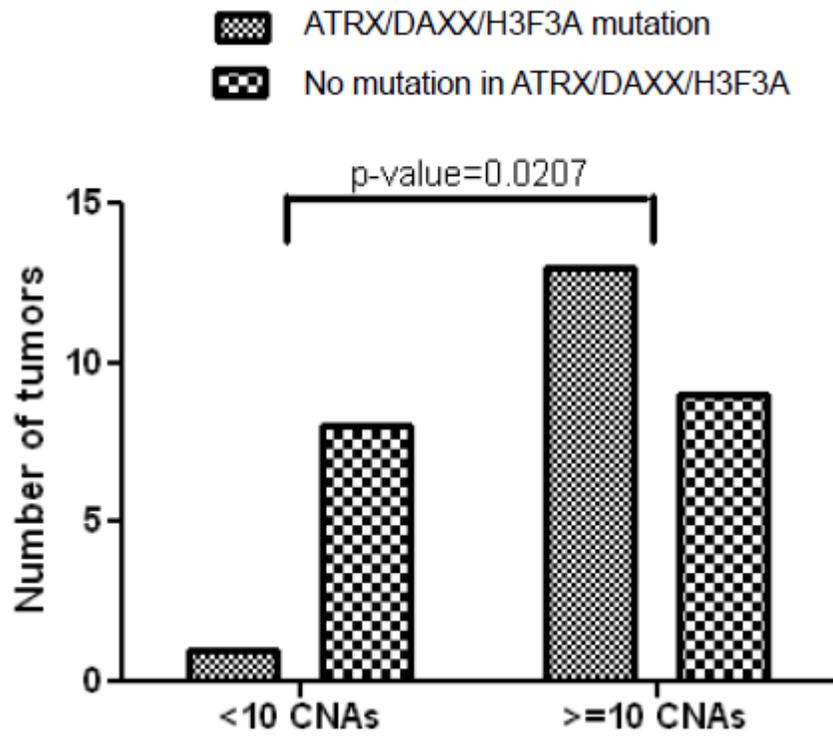
Cell lysates from PGBM14 (which harbours G34R mutation), PGBM29 (wild type for H3.3), a primary paediatric GBM cell line SF199 and normal human astrocytes (both wild type for H3.3) were analyzed with antibodies recognizing the three methylated forms of K36. Even though we cannot differentiate H3.3K36me3 from global H3K36me3 levels, results indicate increased methylation of K36me1, me2 and me3 in the one sample tested carrying the G34R mutation.



**Figure S3.6: Single nucleotide polymorphism (SNP) array profiling reveals differences in copy number aberrations (CNAs) in *ATRX*, *H3F3A* and *TP53* –mutated paediatric glioblastoma.**

Focal losses or gains comprising genes relevant in paediatric GBM overlapped with previous reports (Bax, Mackay et al. 2010; Paugh, Qu et al. 2010; Qu, Jacob et al. 2010). Samples were split into a group with relatively stable genomes (<10 CNAs, bland genome in (Bax, Mackay et al. 2010)) and a group with more unstable genomes ( $\geq 10$  CNAs).

Samples with a mutation in at least one of *ATRX*, *H3F3A* and *TP53* were significantly associated with an unstable genome ( $p=0.0207$ , Fisher's exact test).



## 3.10. SUPPLEMENTARY TABLES

Table S3.1: Clinical information for 48 whole exome sequencing samples

Sample ID	Age	Gender	tumor location	Death	OS (months)	Recurrence	PFS (months)	GEP-Affy	SNP 2.5M Illumina
PGBM1	13	F	thalamic	YES	13	YES	5	NO	YES
PGBM2	5	M	left temporo-parietal	YES	6	YES	4	YES	YES
PGBM3	11	M	intraventricular (I-II)	YES	NA	YES	7	NO	YES
PGBM4	10	M	thalamic+lateral ventricular	YES	7	NA	NA	YES	YES
PGBM5	9	F	NA	YES	36	YES	18	NO	YES
PGBM6	11	M	thalamic	YES	5	NA	NA	YES	YES
PGBM8	6	F	NA	YES	12	YES	7	YES	NO
PGBM9	8	F	NA	YES	7	NO	7	YES	NO
PGBM10	11	M	NA	YES	32	YES	25	NO	NO
PGBM11	13	M	NA	YES	18	NO	18	YES	YES
PGBM12	14	M	left temporal lobe	NO	27	NO	27	NO	YES
PGBM13	14	M	occipital lobe	YES	8	NA	NA	NO	YES
PGBM14	15	M	right temporo-parietal	NO	17	NO	17	YES	YES
PGBM15	13	M	NA	YES	13	YES	9	YES	NO
PGBM16	20	F	parietal occipital	NA	NA	NA	NA	YES	NO
PGBM17	17	M	left frontal and axial	NA	NA	NA	NA	YES	NO
PGBM18	14	M	temporal lobe	NO	27	NO	27	NO	YES
PGBM19	20	M	NA	NA	NA	1	NA	NO	YES
PGBM20	11	M	NA	NA	NA	NA	NA	NO	YES
PGBM21	14	F	temporal lobe	YES	34	YES	22	NO	YES
PGBM22	NA	NA	NA	NA	NA	NA	NA	YES	NO
PGBM23	13	M	NA	NO	14	NA	NA	NO	YES
PGBM24	14	M	NA	YES	14	YES	5	YES	YES
PGBM25	12	M	temporal lobe	YES	17	YES	15	NO	YES
PGBM26	14	M	NA	YES	5	YES	5	YES	YES
PGBM27	9	F	NA	YES	10	YES	9	YES	YES
PGBM28	14	M	left temporo-parietal	NA	NA	NA	NA	NO	NO
PGBM29	15	M	NA	NA	NA	NA	NA	YES	NO

Table S3.1: Clinical information for 48 whole exome sequencing samples,  
continued.

Sample ID	Age	Gender	tumor location	Death	OS (months)	Recurrence	PFS (months)	GEP-Affy	SNP 2.5M Illumina
PGBM30	6	M	thalamic	NA	NA	NA	NA	YES	NO
PGBM31	7	F	NA	YES	12	YES	8	NO	YES
PGBM32	4	M	NA	YES	11	YES	10	NO	YES
PGBM33	12	M	NA	NO	8	NA	NA	NO	NO
PGBM34	12	F	NA	YES	8	NA	NA	NO	YES
PGBM35	7.3	M	parietal lobe	YES	NA	YES	7	YES	YES
PGBM36	7	M	NA	YES	25	YES	15	NO	YES
PGBM37	7	M	left cerebellar	NA	NA	NA	NA	YES	YES
PGBM38	11	M	NA	YES	NA	YES	7	NO	NO
PGBM39	12	F	parietal lobe	NO	24	NA	NA	YES	YES
PGBM40	14	F	thalamic	NO	16	NA	NA	YES	YES
PGBM41	7	F	left thalamic	YES	10	YES	9	YES	YES
PGBM42	2	F	NA	YES	8	YES	4	YES	YES
PGBM43	16	F	NA	YES	12	YES	8	YES	YES
PGBM44	6	F	NA	YES	37	YES	23	NO	NO
PGBM45	9	M	right frontal	YES	12	YES	10	YES	YES
PGBM46	14	M	NA	NO	55	NO	55	YES	NO
PGBM47	14	F	NA	YES	14	YES	12	YES	NO
PGBM48	2	M	NA	NO	117	NO	117	NO	NO
PGBM49	5.4	M	frontal lobe	NA	NA	NA	NA	YES	NO

Table S3.2: Summary of sequence analysis of pediatric GBMs.

Sample	Bases sequenced (after quality filtering)	Median # of reads per base in CCDS†	† Using CCDS version 2011/04/22	
			Median # of reads per base in CCDS after duplicate removal†	CCDS bases with at least 10 reads (%)
PGBM1	13,505,091,987	94	85	92.4
PGBM2	17,119,601,726	109	70	91.3
PGBM3	17,792,909,823	111	72	91.0
PGBM4	13,363,577,977	64	52	90.5
PGBM4-blood	14,066,040,787	74	59	91.9
PGBM5	14,504,723,839	75	43	88.9
PGBM6	11,287,727,427	59	46	88.1
PGBM6-blood	13,999,868,369	67	53	88.9
PGBM8	11,897,621,735	109	88	93.5
PGBM9	11,045,904,509	104	85	93.1
PGBM10	11,619,534,201	100	82	93.1
PGBM11	16,935,710,296	112	104	93.9
PGBM12	18,612,864,498	95	56	91.1
PGBM13	10,904,833,155	51	41	87.3
PGBM13-blood	13,552,900,813	73	58	91.9
PGBM14	15,701,377,658	86	53	91.0
PGBM14-blood	10,213,821,624	50	30	83.4
PGBM15	10,582,247,277	86	67	92.1
PGBM16	11,521,709,389	106	80	92.3
PGBM17	11,870,074,056	68	57	91.9
PGBM18	16,596,170,697	113	104	94.3
PGBM19	11,687,545,184	65	53	91.3
PGBM20	13,400,490,858	69	56	91.5
PGBM21	16,068,676,400	102	95	93.8
PGBM22	11,061,729,809	93	74	92.8
PGBM23	14,088,721,409	68	44	89.1
PGBM24	10,190,203,445	49	41	87.9
PGBM25	11,094,215,054	62	52	90.5
PGBM26	13,718,043,045	123	81	91.7
PGBM27	17,672,965,295	98	62	90.5
PGBM28	13,235,175,617	88	65	92.0
PGBM29	4,376,261,391	38	22	83.9
PGBM30	11,331,964,823	100	78	92.5
PGBM31	9,457,007,712	48	29	85.4
PGBM32	15,996,029,641	101	94	94.1
PGBM33	10,601,377,271	89	72	92.7
PGBM34	9,753,010,363	46	27	83.7
PGBM35	16,828,099,833	115	68	91.3
PGBM36	11,731,008,997	68	38	88.2
PGBM37	20,336,444,728	134	78	92.9
PGBM38	9,929,368,120	51	29	85.5
PGBM39	13,628,886,633	65	52	89.7
PGBM40-blood	13,251,854,585	71	59	91.4
PGBM40	11,824,281,050	58	46	87.2
PGBM41-blood	14,095,936,522	68	53	88.5
PGBM41	17,799,081,592	120	78	91.3
PGBM42	11,032,711,376	64	54	89.9
PGBM43	14,904,682,891	88	80	92.4
PGBM44	14,651,870,734	102	94	93.6
PGBM45	17,328,188,664	90	60	91.7
PGBM46	10,137,136,155	82	69	92.7
PGBM47	10,003,457,301	80	66	92.3
PGBM48	11,519,238,552	111	91	93.4
PGBM49	11,864,031,610	103	81	92.5

Table S3.3: Candidate somatic mutations in the 6 tumor samples with matched germline DNA.

Tumor variants were considered to be somatic when matched normal had more than >= 10 reads and 0 variant reads

Sample	# Somatic mutations		
	Normal has >= 0 reads	Normal has >= 5 reads	Normal has >= 10 reads**
PGBM6	20	17	13
PGBM13	32	32	31
PGBM4	14	12	12
PGBM39	19	19	16
PGBM14	29	18	14
PGBM40	10	6	3

\*\*Variants shown in table below

Sample	Gene	Transcript accession	Nucleotide variant	Amino acid change	Mutation type
PGBM6	AHNAK	NM_001620.1	c.10565C>T	p.(Pro3522Leu)	nonsynonymous SNV
PGBM39	AHRR	NM_020731.4	c.496G>A	p.(Asp166Asn)	nonsynonymous SNV
PGBM14	ATRX	NM_000489.3	c.5269G>T	p.(Glu1757*)	stopgain SNV
PGBM4	ATRX	NM_000489.3	c.3168delG	p.(Lys1057Argfs*61)	frameshift deletion
PGBM13	ATRX	NM_000489.3	c.5215C>T	p.(Arg1739*)	stopgain SNV
PGBM6	ATRX	NM_000489.3	c.5399T>C	p.(Met1800Thr)	nonsynonymous SNV
PGBM13	BMPER	NM_133468.3	c.1476G>T	p.(Lys492Asn)	nonsynonymous SNV
PGBM39	BRAF	NM_004333.4	c.1799T>A	p.(Val600Glu)	nonsynonymous SNV
PGBM13	C13orf40	NM_001146197.1	c.3703G>C	p.(Glu1235Gln)	nonsynonymous SNV
PGBM13	C20orf195	NM_024059.2	c.16G>T	p.(Ala6Ser)	nonsynonymous SNV
PGBM13	C8orf73	NM_001100878.1	c.1933G>A	p.(Asp645Asn)	nonsynonymous SNV
PGBM13	CD5L	NM_005894.2	c.568C>T	p.(Arg190Cys)	nonsynonymous SNV
PGBM13	CHMP7	NM_152272.3	c.1012G>T	p.(Asp338Tyr)	nonsynonymous SNV
PGBM13	CMYA5	NM_153610.3	c.2674C>T	p.(Arg892*)	stopgain SNV
PGBM39	COL19A1	NM_001858.4	c.1969A>T	p.(Thr657Ser)	nonsynonymous SNV

Table S3.3: Candidate somatic mutations in the 6 tumor samples with matched germline DNA, continued.

Sample	Gene	Transcript accession	Nucleotide variant	Amino acid change	Mutation type
PGBM13	CR2	NM_001006658.2	c.1559G>A	p.(Arg520His)	nonsynonymous SNV
PGBM14	CSMD3	NM_198123.1	c.1352C>A	p.(Ala451Asp)	nonsynonymous SNV
PGBM39	DSPP	NM_014208.3	c.3447A>C	p.(Glu1149Asp)	nonsynonymous SNV
PGBM6	DUSP6	NM_001946.2	c.848G>A	p.(Arg283Gln)	nonsynonymous SNV
PGBM40	EIF4E1B	NM_001099408.1	c.140G>A	p.(Gly47Glu)	nonsynonymous SNV
PGBM14	FBXW7	NM_033632.2	c.566_567del	p.(Lys189Serfs*66)	frameshift deletion
PGBM4	FCGBP	NM_003890.2	c.14369G>A	p.(Gly4790Asp)	nonsynonymous SNV
PGBM6	FGFR1	NM_023110.2	c.1966A>G	p.(Lys656Glu)	nonsynonymous SNV
PGBM39	GNA5	NM_001077490.1	c.644C>T	p.(Ser215Phe)	nonsynonymous SNV
PGBM39	GPR172A	NM_024531.3	c.1052G>A	p.(Gly351Asp)	nonsynonymous SNV
PGBM4	GRIPAP1	NM_020137.3	c.2414A>G	p.(Lys805Arg)	nonsynonymous SNV
PGBM4	GYS2	NM_021957.3	c.1889C>T	p.(Thr630Met)	nonsynonymous SNV
PGBM4	H3F3A	NM_002107.4	c.83A>T	p.(Lys28Met)	nonsynonymous SNV
PGBM6	H3F3A	NM_002107.4	c.83A>T	p.(Lys28Met)	nonsynonymous SNV
PGBM13	H3F3A	NM_002107.4	c.103G>A	p.(Gly35Arg)	nonsynonymous SNV
PGBM14	H3F3A	NM_002107.4	c.103G>A	p.(Gly35Arg)	nonsynonymous SNV
PGBM13	HMX3	NM_001105574.1	c.622G>T	p.(Gly208Cys)	nonsynonymous SNV
PGBM13	HOOK1	NM_015888.4	c.206A>G	p.(Asp69Gly)	nonsynonymous SNV
PGBM14	KCN52	NM_020697.2	c.395_397del	p.(Glu133del)	nonframeshift deletion
PGBM13	KIAA1217	NM_019590.3	c.3988G>A	p.(Val1330Met)	nonsynonymous SNV
PGBM4	KIAA1826	NM_032424.1	c.904C>T	p.(Arg302*)	stopgain SNV
PGBM39	KRT27	NM_181537.3	c.167G>A	p.(Gly56Glu)	nonsynonymous SNV
PGBM40	LOXL4	NM_032211.6	c.247G>T	p.(Ala83Ser)	nonsynonymous SNV
PGBM6	LPHN2	NM_012302.2	c.3287C>A	p.(Pro1096Gln)	nonsynonymous SNV
PGBM39	LRP1	NM_002332.2	c.2218C>T	p.(Pro740Ser)	nonsynonymous SNV
PGBM13	LSP1	NM_002339.2	c.970G>A	p.(Gly324Arg)	nonsynonymous SNV

Table S3.3: Candidate somatic mutations in the 6 tumor samples with matched germline DNA, continued.

Sample	Gene	Transcript accession	Nucleotide variant	Amino acid change	Mutation type
PGBM39	LUM	NM_002345.3	c.547C>T	p.(Leu183Phe)	nonsynonymous SNV
PGBM4	LYPD5	NM_001031749.2	c.695G>A	p.(Arg232Gln)	nonsynonymous SNV
PGBM14	MARK1	NM_018650.3	c.1259G>A	p.(Arg420Gln)	nonsynonymous SNV
PGBM14	MFG8	NM_005928.2	c.118_120del	p.(Glu40del)	nonframeshift deletion
PGBM40	MTF1	NM_005955.2	c.1532C>A	p.(Ala511Glu)	nonsynonymous SNV
PGBM13	MTUS2	NM_001033602.2	c.1472C>T	p.(Thr491Met)	nonsynonymous SNV
PGBM13	MYO5C	NM_018728.3	c.4626C>A	p.(Asp1542Glu)	nonsynonymous SNV
PGBM39	NCAM2	NM_004540.3	c.2230A>G	p.(Ser744Gly)	nonsynonymous SNV
PGBM4	NDST2	NM_003635.3	c.329G>A	p.(Arg110His)	nonsynonymous SNV
PGBM6	NF1	NM_001042492.2	c.3735_3744del	p.(Phe1247Glyfs*16)	frameshift deletion
PGBM6	NF1	NM_001042492.2	c.6746_6748del	p.(Val2251del)	nonframeshift deletion
PGBM13	NLRP2	NM_017852.3	c.1379C>T	p.(Ala460Val)	nonsynonymous SNV
PGBM6	OR1E1	NM_003553.2	c.437C>T	p.(Ala146Val)	nonsynonymous SNV
PGBM13	OR4C6	NM_001004704.1	c.662G>T	p.(Cys221Phe)	nonsynonymous SNV
PGBM6	OR51A7	NM_001004749.1	c.136C>T	p.(Leu46Phe)	nonsynonymous SNV
PGBM4	PCDH814	NM_018934.2	c.1966G>A	p.(Ala656Thr)	nonsynonymous SNV
PGBM14	PHF3	NM_015153.2	c.310_312del	p.(Glu106del)	nonframeshift deletion
PGBM4	PIK3C2A	NM_002645.2	c.458C>T	p.(Ala153Val)	nonsynonymous SNV
PGBM13	PRIC285	NM_001037335.2	c.4842C>A	p.(Asp1614Glu)	nonsynonymous SNV
PGBM6	PTEN	NM_000314.4	c.634-2A>C	splicing	splicing
PGBM13	PTGDR	NM_000953.2	c.146G>T	p.(Cys49Phe)	nonsynonymous SNV
PGBM39	RAB23	NM_016277.3	c.551C>T	p.(Thr184Met)	nonsynonymous SNV
PGBM13	RANBP2	NM_006267.4	c.7106G>A	p.(Arg2369His)	nonsynonymous SNV
PGBM13	REX1	NM_001042681.1	c.8C>T	p.(Ala3Val)	nonsynonymous SNV
PGBM13	RGMA	NM_020211.2	c.1248G>T	p.(Arg416Ser)	nonsynonymous SNV
PGBM13	RHOB1	NM_014836.4	c.1502C>T	p.(Pro501Leu)	nonsynonymous SNV

Table S3.3: Candidate somatic mutations in the 6 tumor samples with matched germline DNA, continued.

Sample	Gene	Transcript accession	Nucleotide variant	Amino acid change	Mutation type
PGBM13	<i>RYR2</i>	NM_001035.2	c.13130C>T	p.(Ser4377Leu)	nonsynonymous SNV
PGBM39	<i>SDHA</i>	NM_004168.2	c.772G>C	p.(Gly258Arg)	nonsynonymous SNV
PGBM14	<i>SESN3</i>	NM_144665.2	c.649_650del	p.(Asp217Serfs*19)	frameshift deletion
PGBM13	<i>SFXN4</i>	NM_213649.1	c.971C>A	p.(Ser324Tyr)	nonsynonymous SNV
PGBM14	<i>TKT</i>	NM_001135055.2	c.1644C>T	p.(Trp548Cys)	nonsynonymous SNV
PGBM6	<i>TMC2</i>	NM_080751.2	c.2173C>A	p.(Pro725Thr)	nonsynonymous SNV
PGBM13	<i>TMEM132D</i>	NM_133448.2	c.89G>T	p.(Gly30Val)	nonsynonymous SNV
PGBM6	<i>TNP2</i>	NM_005425.4	c.62C>T	p.(Pro21Leu)	nonsynonymous SNV
PGBM14	<i>TP53</i>	NM_000546.4	c.817C>T	p.(Arg273Cys)	nonsynonymous SNV
PGBM14	<i>TP53</i>	NM_000546.4	c.743G>A	p.(Arg248Gln)	nonsynonymous SNV
PGBM4	<i>TP53</i>	NM_000546.4	c.785delG	p.(Gly262Valfs*83)	frameshift deletion
PGBM13	<i>TP53</i>	NM_000546.4	c.767delC	p.(Thr256Asnfs*89)	frameshift deletion
PGBM39	<i>TRIM28</i>	NM_005762.2	c.499G>A	p.(Val167Met)	nonsynonymous SNV
PGBM13	<i>TTN</i>	NM_133378.4	c.24060C>A	p.(Phe8020Leu)	nonsynonymous SNV
PGBM4	<i>UBE2I</i>	NM_194261.2	c.28G>A	p.(Ala10Thr)	nonsynonymous SNV
PGBM13	<i>UBE3A</i>	NM_000462.3	c.1619T>G	p.(Leu540Arg)	nonsynonymous SNV
PGBM14	<i>URB2</i>	NM_014777.2	c.156G>T	p.(Leu52Phe)	nonsynonymous SNV
PGBM39	<i>USP26</i>	NM_031907.1	c.2138T>A	p.(Ile713Asn)	nonsynonymous SNV
PGBM13	<i>ZCCHC4</i>	NM_024936.2	c.100G>T	p.(Ala34Ser)	nonsynonymous SNV
PGBM14	<i>ZCCHC5</i>	NM_152694.2	c.1085A>T	p.(Gln362Leu)	nonsynonymous SNV
PGBM39	<i>ZNF622</i>	NM_033414.2	c.525G>T	p.(Glu175Asp)	nonsynonymous SNV
PGBM39	<i>ZNF622</i>	NM_033414.2	c.327G>C	p.(Met109Ile)	nonsynonymous SNV

Table S3.4: Summary of somatic mutations in pediatric glioblastoma and 5 cancer types from Parsons et al.

	Pediatric GBM	Adult GBM	Medulloblastoma	Pancreas	Colorectal	Breast
Number of samples analyzed	6	21	22	24	11	11
Number of mutated genes	80	685	218	1007	769	1026
Number of nonsilent mutations	87	748	183	1163	849	1112
Missense	71 (81.6%)	622 (83.2%)	130 (71.0%)	974 (83.7%)	722 (85%)	909 (81.7%)
Nonsense	4 (4.6%)	43 (5.7%)	18 (9.8%)	60 (5.2%)	48 (5.7%)	64 (5.8%)
Insertion	0	3 (0.4%)	5 (2.7%)	4 (0.3%)	4 (0.5%)	5 (0.4%)
Deletion	10 (11.5%)	46 (6.1%)	14 (7.7%)	43 (3.7%)	27 (3.2%)	78 (7.0%)
Duplication	0	7 (0.9%)	7 (3.8%)	31 (2.7%)	18 (2.1%)	3 (0.3%)
Splice site or UTR	2 (2.3%)	27 (3.6%)	9 (4.9%)	51 (4.4%)	30 (3.5%)	53 (4.8%)
Average number of nonsilent mutations per sample	15	36	8	48	77	101
Observed/expected number of nonsense alterations		1	2.48	1.18	1.25	1.37
Total number of substitutions	77	937	199	1486	893	1157
Substitutions at C:G base pairs						
C:G to T:A**	40 (50.6%)	601 (64.1%)	109 (54.8%)	798 (53.8%)	534 (59.8%)	422 (36.5%)
C:G to G:C**	3 (3.8%)	67 (7.2%)	12 (6.0%)	142 (9.6%)	61 (6.8%)	325 (28.1%)
C:G to A:T**	21 (26.6%)	114 (12.1%)	41 (20.6%)	246 (16.6%)	130 (14.6%)	175 (15.1%)
Substitutions at T:A base pairs						
T:A to C:G**	5 (6.3%)	87 (9.3%)	19 (9.5%)	142 (9.6%)	69 (7.7%)	102 (8.8%)
T:A to G:C**	3 (3.8%)	24 (2.6%)	14 (7.0%)	79 (5.3%)	59 (6.6%)	57 (4.9%)
T:A to A:T**	7 (8.9%)	44 (4.7%)	4 (2.0%)	77 (5.2%)	40 (4.5%)	76 (6.6%)
Substitutions at specific dinucleotides						
5'-CpG-3'***	no data	404 (43.1%)	85 (42.7%)	563 (37.9%)	427 (47.8%)	195 (16.9%)
5'-TpC-3'***	no data	102 (10.9%)	14 (7.0%)	218 (14.7%)	99 (11.1%)	395 (34.1%)

Table S3.5: Mutations in selected genes H3F3A, ATRX, DAXX, IDH1, PDGFRA, EGFR, TP53.

Sample	Gene	Transcript accession	Nucleotide variant	Amino acid change	Mutation type
PGBM1	H3F3A	NM_002107.4	c.83A>T	p.(Lys27Met)	Missense
PGBM2	H3F3A	NM_002107.4	c.83A>T	p.(Lys27Met)	Missense
PGBM3	H3F3A	NM_002107.4	c.83A>T	p.(Lys27Met)	Missense
PGBM5	H3F3A	NM_002107.4	c.83A>T	p.(Lys27Met)	Missense
PGBM6	H3F3A	NM_002107.4	c.83A>T	p.(Lys27Met)	Missense
PGBM4	H3F3A	NM_002107.4	c.83A>T	p.(Lys27Met)	Missense
PGBM8	H3F3A	NM_002107.4	c.83A>T	p.(Lys27Met)	Missense
PGBM9	H3F3A	NM_002107.4	c.83A>T	p.(Lys27Met)	Missense
PGBM10	H3F3A	NM_002107.4	c.83A>T	p.(Lys27Met)	Missense
PGBM11	H3F3A	NM_002107.4	c.103G>A	p.(Gly34Arg)	Missense
PGBM14	H3F3A	NM_002107.4	c.103G>A	p.(Gly34Arg)	Missense
PGBM12	H3F3A	NM_002107.4	c.103G>A	p.(Gly34Arg)	Missense
PGBM13	H3F3A	NM_002107.4	c.103G>A	p.(Gly34Arg)	Missense
PGBM15	H3F3A	NM_002107.4	c.103G>A	p.(Gly34Arg)	Missense
PGBM16	H3F3A	NM_002107.4	c.104G>T	p.(Gly34Val)	Missense
PGBM1	ATRX	NM_000489.3	c.3364delT	p.(Cys1122Valfs*8)	Frameshift indel
PGBM4	ATRX	NM_000489.3	c.3168delG	p.(Lys1057Argfs*61)	Frameshift indel
PGBM6	ATRX	NM_000489.3	c.5399T>C	p.(Met1800Thr)	Missense
PGBM11	ATRX	NM_000489.3	c.4179_4182del	p.(Ser1394Asnfs*95)	Frameshift indel
PGBM12	ATRX	NM_000489.3	c.5178_5179insA	p.(Glu1727Argfs*7)	Frameshift indel
PGBM13	ATRX	NM_000489.3	c.5215C>T	p.(Arg1739*)	Nonsense
PGBM14	ATRX	NM_000489.3	c.5269G>T	p.(Glu1757*)	Nonsense
PGBM15	ATRX	NM_000489.3	c.6761A>G	p.(His2254Arg)	Missense
PGBM16	ATRX	NM_000489.3	c.6331C>T	p.(Arg2111*)	Nonsense
PGBM17	ATRX	NM_000489.3	c.4766G>T	p.(Gly1589Val)	Missense
PGBM18	ATRX	NM_000489.3	c.4276C>T	p.(Arg1426*)	Nonsense
PGBM19	ATRX	NM_000489.3	c.4745_4746insA	p.(Lys1584Glufs*17)	Frameshift indel
PGBM20	ATRX	NM_000489.3	c.7327A>G	p.(Asn2443Asp)	Missense
PGBM22	ATRX	NM_000489.3	c.4745_4746insA	p.(Lys1584Glufs*17)	Frameshift indel
PGBM22	ATRX	NM_000489.3	c.3904delA	p.(Arg1302Glufs*44)	Frameshift indel
PGBM22	ATRX	NM_000489.3	c.6406G>A	p.(Asp2136Asn)	Missense
PGBM19	DAXX	NM_001350.4	c.1885_1886insC	p.(Cys629Serfs*29)	Frameshift indel
PGBM21	DAXX	NM_001350.4	c.712C>T	p.(Arg238*)	Nonsense
PGBM17	IDH1	NM_005896.2	c.395G>A	p.(Arg132His)	Missense
PGBM18	IDH1	NM_005896.2	c.395G>A	p.(Arg132His)	Missense
PGBM23	IDH1	NM_005896.2	c.395G>A	p.(Arg132His)	Missense
PGBM29	IDH1	NM_005896.2	c.395G>A	p.(Arg132His)	Missense
PGBM2	PDGFRA	NM_006206.4	c.[1154A>T;1155G>A]	p.(Lys385Ile)	Missense
PGBM16	PDGFRA	NM_006206.4	c.1154A>T	p.(Lys385Met)	Missense
PGBM34	PDGFRA	NM_006206.4	c.2525_2527del	p.(Asp842_Ile843delins Val)	Nonframeshift indel
PGBM12	PDGFRA	NM_006206.4	c.2545T>G	p.(Tyr849Asp)	Missense
PGBM22	EGFR	NM_005228.3	c.2165C>T	p.(Ala722Val)	Missense
PGBM27	EGFR	NM_005228.3	c.2950G>A	p.(Asp984Asn)	Missense
PGBM1	TP53	NM_000546.4	c.916C>T	p.(Arg306*)	Nonsense
PGBM1	TP53	NM_000546.4	c.455_459del	p.(Pro152Argfs*27)	Frameshift indel
PGBM2	TP53	NM_000546.4	c.637C>T	p.(Arg213*)	Nonsense
PGBM3	TP53	NM_000546.4	c.393_395del	p.(Asn131del)	Nonframeshift indel

Table S3.5: Mutations in selected genes H3F3A, ATRX, DAXX, IDH1, PDGFRA, EGFR, TP53, continued.

Sample	Gene	Transcript accession	Nucleotide variant	Amino acid change	Mutation type
PGBM4	TP53	NM_000546.4	c.785delG	p.(Gly262Valfs*83)	Frameshift indel
PGBM8	TP53	NM_000546.4	c.817C>T	p.(Arg273Cys)	Missense
PGBM9	TP53	NM_000546.4	c.818G>C	p.(Arg273Pro)	Missense
PGBM11	TP53	NM_000546.4	c.488A>G	p.(Tyr163Cys)	Missense
PGBM12	TP53	NM_000546.4	c.1024C>T	p.(Arg342*)	Nonsense
PGBM12	TP53	NM_000546.4	c.524G>A	p.(Arg175His)	Missense
PGBM13	TP53	NM_000546.4	c.767delC	p.(Thr256Asnfs*89)	Frameshift indel
PGBM14	TP53	NM_000546.4	c.817C>T	p.(Arg273Cys)	Missense
PGBM14	TP53	NM_000546.4	c.743G>A	p.(Arg248Gln)	Missense
PGBM15	TP53	NM_000546.4	c.548_549insGCCCCCA	p.(Asp184_Asp393delinsProPro)	Nonframeshift indel
PGBM16	TP53	NM_000546.4	c.1024C>T	p.(Arg342*)	Nonsense
PGBM17	TP53	NM_000546.4	c.659A>G	p.(Tyr220Cys)	Missense
PGBM18	TP53	NM_000546.4	c.586C>T	p.(Arg196*)	Nonsense
PGBM18	TP53	NM_000546.4	c.817C>T	p.(Arg273Cys)	Missense
PGBM19	TP53	NM_000546.4	c.800G>A	p.(Arg267Gln)	Missense
PGBM19	TP53	NM_000546.4	c.689C>T	p.(Thr230Ile)	Missense
PGBM20	TP53	NM_000546.4	c.742C>T	p.(Arg248Trp)	Missense
PGBM21	TP53	NM_000546.4	c.799C>T	p.(Arg267Trp)	Missense
PGBM21	TP53	NM_000546.4	c.455C>T	p.(Pro152Leu)	Missense
PGBM22	TP53	NM_000546.4	c.1009C>T	p.(Arg337Cys)	Missense
PGBM22	TP53	NM_000546.4	c.524G>A	p.(Arg175His)	Missense
PGBM23	TP53	NM_000546.4	c.761T>G	p.(Ile254Ser)	Missense
PGBM24	TP53	NM_000546.4	c.586C>T	p.(Arg196*)	Nonsense
PGBM25	TP53	NM_000546.4	c.1024C>T	p.(Arg342*)	Nonsense
PGBM26	TP53	NM_000546.4	c.524G>A	p.(Arg175His)	Missense
PGBM27	TP53	NM_000546.4	c.751A>C	p.(Ile251Leu)	Missense
PGBM28	TP53	NM_000546.4	c.818G>A	p.(Arg273His)	Missense
PGBM29	TP53	NM_000546.4	c.29T>G	p.(Val10Gly)	Missense
PGBM30	TP53	NM_000546.4	c.733G>A	p.(Gly245Ser)	Missense
PGBM5	NF1	NM_000267.3	c.6787_6790del	p.(Tyr2264Thrfs*5)	Frameshift indel
PGBM6	NF1	NM_000267.3	c.3735_3744del	p.(Phe1247Glyfs*16)	Frameshift indel
PGBM6	NF1	NM_000267.3	c.6683_6685del	p.(Val2230del)	Nonframeshift indel
PGBM10	NF1	NM_000267.3	c.2970delA	p.(Met991*)	Nonsense
PGBM18	NF1	NM_000267.3	c.1866T>A	p.(Cys622*)	Nonsense
PGBM18	NF1	NM_000267.3	c.4466delT	p.(Leu1489Hisfs*64)	Frameshift indel
PGBM19	NF1	NM_000267.3	c.1318C>T	p.(Arg440*)	Nonsense
PGBM21	NF1	NM_000267.3	c.5839C>T	p.(Arg1947*)	Nonsense
PGBM22	NF1	NM_000267.3	c.7846C>T	p.(Arg2616*)	Nonsense
PGBM22	NF1	NM_000267.3	c.2659G>A	p.(Ala887Thr)	Missense
PGBM22	NF1	NM_000267.3	c.1381C>T	p.(Arg461*)	Nonsense
PGBM25	NF1	NM_000267.3	c.4575delG	p.(Gly1526Valfs*27)	Frameshift indel
PGBM26	NF1	NM_000267.3	c.6787_6790del	p.(Tyr2264Thrfs*5)	Frameshift indel
PGBM28	NF1	NM_000267.3	c.2026_2027insC	p.(Ile679Aspfs*21)	Frameshift indel
PGBM32	NF1	NM_000267.3	c.4879A>T	p.(Thr1627Ser)	Missense
PGBM33	NF1	NM_000267.3	c.1641+1G>T		Splicing

Table S3.6: Comparison of genes mutated in pediatric GBM and in each of the 4 described adult GBM molecular subgroups.

Pathway	Gene	Pediatric GBM		adult GBM Gene-expression based molecular subtypes											
		No. of tumors	% of tumors	Proneural		Neural		Classical		Mesenchymal					
		No. of tumors	% of tumors	No. of tumors	% of tumors	No. of tumors	% of tumors	No. of tumors	% of tumors	No. of tumors	% of tumors	No. of tumors	% of tumors		
Chromatin Remodelling	H3F3A	33/91	36	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA		
	ATRX	59/191	31	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA		
	DAXX	2/70	3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA		
	IDH1	8/84	10	11/37	30	0.0124	1/19	5	1	0/22	0	0.2006	0/38	0	0.0564
Cell Signalling	EGFR	2/49	4	6/37	16	0.0703	5/19	26	0.0155	7/22	32	0.0029	2/38	5	1
	PDGFRA	4/49	8	4/37	11	0.7208	0/19	0	0.5702	0/22	0	0.3033	0/38	0	0.1283
	NF1	13/49	26	2/37	5	0.011	3/19	16	0.5261	1/22	5	0.0499	14/38	37	0.3542
	PIK3CA	3/49	6	3/37	8	1	1/19	5	1	1/22	5	1	1/38	3	0.6286
	PIK3R1	5/49	10	7/37	19	0.3477	2/19	11	1	1/22	5	0.6583	0/38	0	0.0652
PTEN	3/49	6	6/37	16	0.1646	4/19	21	0.0892	5/22	23	0.0971	12/38	32	0.0032	
Cell Cycle	TP53	27/49	55	20/37	54	1	4/19	21	0.0149	0/22	0	<0.0001	12/38	32	0.0325
	CDKN2A	3/49	6	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	RBI	5/49	10	1/37	3	0.2302	1/19	5	1	0/22	0	0.3151	5/38	13	0.742

\*Fisher's two tailed test to compare between pediatric GBM and different subtypes of adult GBM (Proneural, Neural, Classical, Mesenchymal).

N/A: not available

Table S3.7: Top 100 differentially expressed genes by standard deviation, used for unsupervised hierarchical clustering, ordered as presented in Figure 3.3b.

	G34 Mutant						K27 Mutant						SD	Mean(G34)- Mean(K27)
	PGBM14	PGBM15	PGBM16	PGBM11	PGBM2	PGBM4	PGBM6	PGBM8	PGBM9	PGBM5	PGBM7	PGBM3		
FOXG1	11.79	11.92	11.32	11.89	4.11	0.00	4.17	3.17	1.72	4.96	9.09			
SP8	5.47	9.88	7.95	9.13	0.00	0.68	0.00	2.07	0.00	4.20	7.56			
DLX6-AS1	2.72	10.83	10.20	9.38	0.00	2.70	2.29	0.00	0.00	4.58	7.28			
DLX2	4.03	10.96	9.26	9.25	1.26	3.69	1.20	0.38	0.00	4.28	7.07			
DLX1	5.07	11.45	10.02	9.88	2.74	4.87	0.68	2.51	0.93	4.12	6.76			
DLX6	3.74	10.05	9.02	7.35	1.32	1.58	2.23	0.85	0.00	3.80	6.35			
C14orf23	7.70	6.80	5.78	7.34	0.77	0.00	2.26	0.49	2.20	3.16	5.76			
DLX5	3.87	9.48	7.36	6.77	0.26	0.58	3.55	1.20	0.93	3.39	5.57			
FZD7	8.71	8.54	10.49	9.06	2.98	1.81	3.77	6.37	4.04	3.14	5.41			
PCK1	5.40	6.79	7.78	3.58	0.00	0.00	1.26	0.00	2.91	3.03	5.05			
NPY	7.01	6.22	11.08	11.00	3.20	3.64	4.57	3.87	3.94	3.08	4.99			
MOXD1	5.62	6.93	10.28	10.20	6.64	3.36	1.26	4.53	4.49	2.98	4.20			
TRD@	8.31	4.15	6.92	10.72	1.38	4.89	1.89	5.04	3.83	2.99	4.12			
NEUROD1	3.32	8.05	3.05	9.67	0.00	3.49	1.26	3.41	4.91	3.05	3.41			
GES1	10.48	2.46	4.04	5.86	2.83	0.00	4.14	4.65	0.00	3.19	3.39			
LOC441179	1.38	8.96	5.35	8.56	3.04	0.26	3.79	2.32	4.41	2.99	3.30			
KIRREL3	0.85	5.77	7.36	8.79	0.68	3.68	1.26	2.56	4.69	2.93	3.12			
LOC100292909	4.04	7.45	9.93	5.33	0.00	4.17	1.38	6.41	6.58	3.05	2.98			
TFPI2	0.38	5.25	9.53	6.41	0.00	0.00	0.85	8.74	2.87	3.81	2.90			
LOC100192378	1.14	8.73	7.59	6.04	7.39	0.00	5.00	2.83	0.00	3.40	2.83			
HLA-DQA1	4.52	3.95	6.79	5.76	0.14	0.14	4.96	9.56	0.00	3.33	2.30			
HES5	2.89	9.44	8.08	6.58	1.07	8.77	2.00	5.55	5.86	3.02	2.09			
PLN	0.58	10.65	4.64	7.78	4.03	0.00	3.83	6.28	5.30	3.32	2.03			
HLA-DQB1	0.93	3.98	5.51	5.79	0.58	0.26	0.93	8.58	0.26	3.07	1.93			

Table S3.7: Top 100 differentially expressed genes by standard deviation, used for unsupervised hierarchical clustering, ordered as presented in

Figure 3.3b, continued.

	G34 Mutant				K27 Mutant				SD	Mean(G34)- Mean(K27)	
	PGBM14	PGBM15	PGBM16	PGBM11	PGBM2	PGBM4	PGBM6	PGBM8			PGBM9
LOC100271840	3.83	6.14	2.94	8.37	1.96	1.14	0.68	5.57	8.06	2.88	1.84
AQP9	9.75	1.54	0.58	3.96	1.58	2.07	3.02	6.13	1.68	2.91	1.06
CXCL14	9.76	5.11	4.93	5.49	1.77	1.20	12.02	7.85	6.34	3.48	0.49
OGDHL	5.28	0.49	0.49	8.73	0.77	6.28	6.78	1.32	2.85	3.15	0.14
SLC14A1	1.20	0.68	6.00	2.26	7.90	0.49	0.00	5.75	0.14	3.01	-0.32
CNGA3	0.00	1.43	6.83	5.15	8.57	0.00	1.14	3.35	6.10	3.17	-0.48
DDIT4L	11.53	5.58	2.89	3.14	8.93	3.38	8.05	7.96	4.80	3.02	-0.84
COL6A2	4.22	4.24	4.43	0.00	0.26	0.58	8.39	6.61	5.35	2.95	-1.02
CHIBL1	13.30	4.92	7.45	9.87	8.21	8.10	13.68	12.75	7.74	3.05	-1.21
MET	8.83	5.51	4.41	6.11	3.61	12.84	4.77	11.84	4.14	3.45	-1.22
ASCL1	1.20	11.78	10.24	9.84	11.13	9.68	7.69	8.75	10.35	3.15	-1.25
C8orf34	0.85	5.86	4.73	3.63	10.63	1.63	4.11	7.12	3.52	2.95	-1.64
SLC6A15	1.38	7.21	2.79	0.49	6.18	7.81	0.00	2.63	6.80	3.08	-1.72
CRABP1	0.85	4.82	0.49	1.20	0.00	4.46	1.07	4.93	8.57	2.89	-1.97
C1orf192	1.77	0.68	5.78	0.00	9.55	0.85	1.00	7.12	1.81	3.41	-2.01
IL8	6.79	5.48	5.28	4.15	2.63	3.29	10.42	11.62	9.20	3.21	-2.01
AKR1C1	12.64	3.91	4.78	4.50	9.00	9.68	5.31	9.40	9.05	3.03	-2.03
SLC39A12	0.93	2.96	3.83	4.12	8.75	0.38	6.07	7.77	2.20	2.91	-2.07
FSTL5	0.68	3.93	0.38	4.90	10.54	5.79	0.00	4.83	2.29	3.35	-2.22
LTF	10.01	0.00	5.68	5.88	10.11	0.00	8.63	11.40	8.30	4.22	-2.30
NEFL	4.12	5.89	0.58	1.72	3.91	10.69	1.63	7.85	2.89	3.27	-2.31
C2orf40	2.70	6.16	3.97	5.22	10.44	4.54	0.58	9.94	9.27	3.41	-2.44
CDH13	1.00	0.00	6.84	1.93	6.38	0.00	6.89	6.15	5.92	3.07	-2.63
C7orf57	2.74	0.00	3.00	0.00	9.32	1.43	0.00	6.90	2.98	3.25	-2.69
CCL20	6.00	0.77	1.49	0.00	1.81	0.77	7.35	8.71	5.38	3.28	-2.74
KCNA5	2.61	0.00	0.14	1.26	7.63	0.00	5.40	0.38	5.45	2.91	-2.77
GRIA2	7.99	1.72	9.15	10.53	10.14	11.82	10.06	9.04	9.80	2.90	-2.82
SERPINA3	11.80	1.00	9.61	9.56	12.19	7.09	13.37	11.61	9.99	3.71	-2.86
CDH19	0.00	1.96	5.28	7.79	9.32	5.22	7.31	5.76	5.66	2.87	-2.90

Table S3.7: Top 100 differentially expressed genes by standard deviation, used for unsupervised hierarchical clustering, ordered as presented in Figure 3.3b, continued.

	G34 Mutant				K27 Mutant				SD	Mean(G34)- Mean(K27)	
	PGBM14	PGBM15	PGBM16	PGBM11	PGBM2	PGBM4	PGBM6	PGBM8			PGBM9
SLC44A5	0.00	9.04	5.74	8.13	8.92	8.93	7.99	7.96	9.93	3.01	-3.02
BCHE	1.63	9.34	9.31	9.49	10.45	10.93	10.57	10.45	10.73	2.91	-3.18
LOC157503	0.00	4.22	0.00	6.86	8.43	7.50	3.52	4.99	5.51	3.01	-3.22
SCN7A	0.00	3.07	2.87	7.23	8.69	2.41	8.44	6.54	6.71	3.07	-3.27
OTX2	0.00	5.19	2.49	0.00	8.92	0.00	8.56	6.81	1.68	3.70	-3.27
DPP10	3.10	10.40	2.74	2.23	8.04	8.03	7.10	7.95	8.36	2.95	-3.28
AKR1C2	12.37	1.20	2.29	2.04	8.07	9.35	3.64	9.28	8.41	4.03	-3.28
PAK7	0.00	6.65	3.05	2.26	7.67	8.49	3.12	5.31	6.94	2.86	-3.31
CALB1	7.79	0.00	1.14	2.61	3.32	7.24	6.17	7.92	6.57	3.02	-3.36
RALGAPA2	0.00	0.38	3.57	3.90	7.38	8.85	0.00	5.57	5.08	3.24	-3.41
STMN2	4.09	10.37	2.38	8.83	7.94	11.89	9.93	9.05	10.51	3.13	-3.45
DCC	0.00	7.97	1.89	5.32	6.13	9.29	7.47	6.39	7.94	3.04	-3.65
GRIN2A	2.89	0.77	2.41	0.68	9.86	1.54	4.00	5.29	6.17	2.99	-3.69
CTNNA2	1.26	9.33	6.22	4.86	11.04	8.95	8.03	8.75	8.92	2.94	-3.72
KLRK4	4.25	0.58	0.00	2.91	1.00	6.99	8.50	5.69	7.25	3.17	-3.95
PCDH7	2.00	7.03	2.32	10.51	10.94	8.51	8.87	9.09	9.70	3.32	-3.96
ALDH1A3	0.14	5.89	0.00	1.07	6.85	2.20	6.51	6.33	6.78	3.04	-3.96
RALYL	0.93	5.21	3.90	2.32	8.76	8.21	2.77	7.45	8.43	2.99	-4.04
TTC9B	0.68	2.85	0.68	0.85	1.07	7.62	6.68	4.89	6.42	2.89	-4.07
TMEFF2	0.85	8.22	3.58	5.65	8.61	9.90	7.58	7.85	9.37	2.96	-4.08
INSM1	0.38	0.68	0.00	0.00	0.00	8.04	1.26	5.33	7.61	3.41	-4.18
UGT8	0.49	4.48	5.46	8.07	8.90	9.54	7.29	8.46	10.00	3.03	-4.21
RIT2	0.58	6.96	4.67	6.08	9.99	10.69	7.24	7.15	9.00	3.04	-4.24
OGN	0.00	3.10	0.00	0.14	12.07	0.14	3.56	6.22	3.29	3.99	-4.25
H19	3.32	2.00	4.80	4.38	7.12	6.50	11.57	5.65	8.60	2.89	-4.26
NEGR1	0.00	8.91	1.81	7.28	10.16	9.38	7.69	8.15	8.46	3.52	-4.27

Table S3.7: Top 100 differentially expressed genes by standard deviation, used for unsupervised hierarchical clustering, ordered as presented in Figure 3.3b, continued.

	G34 Mutant						K27 Mutant						SD	Mean(G34)- Mean(K27)		
	PGBM14	PGBM15	PGBM16	PGBM11	PGBM2	PGBM4	PGBM6	PGBM8	PGBM9	PGBM2	PGBM4	PGBM6			PGBM8	PGBM9
SHISA6	4.07	2.85	4.24	0.26	9.25	4.38	8.80	5.71	7.55	9.25	4.38	8.80	5.71	7.55	2.91	-4.28
KLRC3	4.94	0.14	2.70	4.41	4.55	8.46	7.81	7.09	8.83	4.55	8.46	7.81	7.09	8.83	2.89	-4.30
FAM19A5	0.00	2.54	7.12	4.71	8.87	8.64	8.25	6.31	9.57	8.87	8.64	8.25	6.31	9.57	3.23	-4.74
SLC7A2	5.08	0.14	1.93	0.38	7.67	5.77	5.45	7.69	6.82	7.67	5.77	5.45	7.69	6.82	2.98	-4.80
SUSD5	6.98	3.51	0.14	4.36	9.03	8.53	9.60	8.14	7.99	9.03	8.53	9.60	8.14	7.99	3.15	-4.91
ODZ2	2.58	6.04	1.20	4.12	8.22	7.21	10.37	7.23	9.07	8.22	7.21	10.37	7.23	9.07	3.05	-4.93
DLK1	6.85	5.16	0.00	2.43	6.68	3.92	11.63	9.52	11.05	6.68	3.92	11.63	9.52	11.05	3.93	-4.95
MYT1	1.63	4.14	0.93	0.26	4.15	9.02	6.90	5.78	7.93	4.15	9.02	6.90	5.78	7.93	3.14	-5.02
KCND2	2.72	5.26	2.20	6.52	9.36	9.93	8.35	8.46	9.99	9.36	9.93	8.35	8.46	9.99	2.99	-5.04
KCNJ9	0.77	3.70	0.93	0.00	8.38	7.11	7.67	2.83	6.00	8.38	7.11	7.67	2.83	6.00	3.23	-5.05
LHFPL3	6.18	4.49	3.49	7.94	11.34	10.58	10.26	9.37	11.40	11.34	10.58	10.26	9.37	11.40	2.99	-5.07
FAM5C	5.73	2.35	0.26	5.26	8.02	9.32	8.23	8.06	9.05	8.02	9.32	8.23	8.06	9.05	3.16	-5.14
DBC1	1.63	2.10	1.96	3.29	9.49	7.00	10.00	6.35	6.28	9.49	7.00	10.00	6.35	6.28	3.23	-5.57
OPCML	0.49	4.28	4.80	5.00	9.81	9.14	8.76	8.25	10.58	9.81	9.14	8.76	8.25	10.58	3.32	-5.67
CA10	3.51	2.20	2.14	2.04	6.50	8.97	8.03	8.05	9.60	6.50	8.97	8.03	8.05	9.60	3.18	-5.76
GPR17	1.07	2.93	0.00	0.00	6.49	5.31	5.19	8.68	8.63	6.49	5.31	5.19	8.68	8.63	3.43	-5.86
CADM2	1.43	4.88	0.26	5.59	9.30	9.03	9.14	8.26	9.30	9.30	9.03	9.14	8.26	9.30	3.54	-5.97
NXPH1	0.14	4.79	3.55	3.88	8.34	10.20	8.91	8.69	9.68	8.34	10.20	8.91	8.69	9.68	3.48	-6.08
SFRP2	3.49	2.51	2.38	0.49	10.24	7.76	7.35	7.95	8.62	10.24	7.76	7.35	7.95	8.62	3.44	-6.17
OLIG1	4.75	2.93	5.34	4.13	8.65	12.51	11.25	10.75	12.15	8.65	12.51	11.25	10.75	12.15	3.78	-6.77
MEGF11	0.00	0.85	1.43	2.49	8.21	7.80	8.33	7.09	10.24	8.21	7.80	8.33	7.09	10.24	3.91	-7.14

Table S3.8: Numbers of CNAs of each type identified in each tumor sample.

ID	GAINS			DELETIONS			LOH			Total Losses	Total CNAs	CNA Grouping Group 1/2	H3F3A Mut Y/N	ATRX Mut Y/N	DAXX Mut Y/N			
	Whole Chr.	Broad	Focal	Total Gains	Whole Chr.	Broad	Focal	Total Dels	Whole Chr.							Broad	Focal	Total LOH
PGBM1	0	2	2	4	0	0	0	0	1	3	2	6	6	10	2	Y	Y	N
PGBM2	3	1	2	6	0	0	0	0	4	10	2	16	16	22	2	Y	N	N
PGBM3	2	10	7	19	0	0	0	0	3	22	5	30	30	49	2	Y	N	N
PGBM4	0	15	9	24	0	0	0	0	4	17	3	24	24	48	2	Y	Y	N
PGBM5	0	4	0	4	0	0	0	0	2	19	4	25	25	29	2	Y	N	N
PGBM6	0	3	1	4	0	0	0	0	0	5	6	11	11	15	2	Y	Y	N
PGBM11	0	7	11	18	0	0	2	2	0	16	2	18	20	36	2	Y	Y	N
PGBM12	1	2	0	3	0	0	3	3	1	22	8	31	34	37	2	Y	Y	N
PGBM13	1	2	1	4	0	0	0	0	1	4	1	6	6	10	2	Y	Y	N
PGBM14	0	3	1	4	0	0	0	0	11	10	1	22	22	26	2	Y	Y	N
PGBM18	0	2	1	3	0	0	0	0	1	3	2	6	6	9	1	N	Y	N
PGBM19	1	0	0	1	0	0	0	0	17	3	0	20	20	21	2	N	Y	Y
PGBM20	1	5	4	10	0	0	0	0	1	23	17	41	41	51	2	N	Y	N
PGBM21	0	4	3	7	0	0	0	0	13	7	1	21	21	28	2	N	N	Y
PGBM23	0	1	0	1	0	0	0	0	0	3	0	3	3	4	1	N	N	N
PGBM24	0	10	1	11	0	0	2	2	8	15	3	26	28	39	2	N	N	N
PGBM25	0	6	11	17	0	1	4	5	11	14	3	28	33	50	2	N	N	N
PGBM26	2	4	11	17	0	0	0	0	8	5	3	16	16	33	2	N	N	N
PGBM27	3	5	14	22	0	1	0	1	9	14	5	28	29	51	2	N	N	N
PGBM31	0	7	5	12	0	0	0	0	1	3	0	4	4	16	2	N	N	N
PGBM32	0	2	2	4	0	0	0	0	0	2	3	5	5	9	1	N	N	N
PGBM34	3	11	8	22	0	1	1	2	4	18	2	24	26	48	2	N	N	N
PGBM35	0	1	1	2	0	0	0	0	3	1	0	4	4	6	1	N	N	N

Table S3.8: Numbers of CNAs of each type identified in each tumor sample, continued.

ID	GAINS				DELETIONS				LOH				Total Losses	Total CNAs	CNA Grouping Group 1/2	H3F3A Mut Y/N	ATRXL Mut Y/N	DAJXX Mut Y/N
	Whole Chr.	Brood	Focal	Total Gains	Whole Chr.	Brood	Focal	Total Deis	Whole Chr.	Brood	Focal	Total LOH						
PGBM36	0	0	2	2	0	0	0	0	0	1	2	3	3	5	1	N	N	N
PGBM37	0	1	0	1	0	0	0	0	0	5	2	7	7	8	1	N	N	N
PGBM39	3	1	0	4	0	0	3	3	3	2	12	17	20	24	2	N	N	N
PGBM40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	N	N	N
PGBM41	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	N	N	N
PGBM42	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	N	N	N
PGBM43	0	5	16	21	0	0	4	4	5	18	9	32	36	57	2	N	N	N
PGBM45	0	1	6	7	0	0	0	0	0	1	6	7	7	14	2	N	N	N
<b>Totals</b>	<b>20</b>	<b>115</b>	<b>119</b>	<b>254</b>	<b>0</b>	<b>3</b>	<b>19</b>	<b>22</b>	<b>111</b>	<b>267</b>	<b>105</b>	<b>483</b>	<b>505</b>	<b>759</b>				
<b>Mean</b>	<b>0.65</b>	<b>3.71</b>	<b>3.84</b>	<b>8.19</b>	<b>0.00</b>	<b>0.10</b>	<b>0.61</b>	<b>0.71</b>	<b>3.58</b>	<b>8.61</b>	<b>3.39</b>	<b>15.58</b>	<b>16.29</b>	<b>24.48</b>				



Table S3.9: CNA regions identified in each tumor sample, continued.

Sample	GAINS			DELETIONS			LOH		
	Whole Chr	Broad	Focal	Whole Chr	Broad	Focal	Whole Chr	Broad	Focal
PGBM4		1p13.3-1q44	4q12				9	3q11.2-3q25.33	7p22.1-7p22.3
		2p15-2p25.3	4q31.21				11	4q12-4q25	7q31.2-7q31.31
		3q25.33-3q29	4q31.23				16	4q32.3-4q35.2	8p11.22-8p11.21
		3q26.32-3q29	4q34.1				20	5q21.1-5q35.3	
		4q25-4q26	7q31.2					6q13-6q27	
		4q26-4q28.2	8q24.21					7p21.3-7p21.1	
		4q31.3-4q32.1	10q26.13					7p21.1-7p11.2	
		4q32.1-4q32.3	10q26.2					7q35.1-7q36.3	
		10q23.31-10q23.33	20p13					8p12-8p23.3	
		12q12-12q13.13						8q	
		12q23.3-12q24.31						10q21.2-10q22.3	
		14q11.2						12p	
		14q21.1-14q21.2						12q	
		14q21.2-14q21.3						13q	
		16q11.2-16q23.1						14q	
								17p	
								22q	
PGBM5		15q11.2-15q14					1	2p	13q31.1
		15q25.1-15q26.3					6	2q11.2-2q12.3	13q31.3
		17q21.32-17q25.3						2q31.1-2q34	13q32.1-13q32.2
		19p13.11-19p12						2q34-2q37.3	17q11.2
								3p21.31-3p26.3	
								5q	
								9p21.2-9p24.3	
								10p14-10p15.3	
								10p12.1-10q26.3	
								11p	
								12q13.11-12q24.33	
								14q23.3-14q32.33	
								15q15.1-15q22.2	
								16q	
								17p	
								17q11.2-17q21.32	
								19q	
							21q11.2-21q21.1		
							21q21.2-21q21.3		

Table S3.9: CNA regions identified in each tumor sample, continued.

Sample	GAINS			DELETIONS			LOH		
	Whole Chr	Brood	Focal	Whole Chr	Brood	Focal	Whole Chr	Brood	Focal
PGBM6	1q		10q22.3				1p31.1-1p11.2		4q25-4q27
	9p23-9p21.1						9p23-9p24.3		10q23.1
	9q31.1-9q34.3						9p21.1-9q31.1		10q23.1
							10q23.31-10q25.1		10q23.1
							19q		10q22.3
									10q22.3
PGBM11	2p24.2-2p25.3		5q31.3			10q23.31	1q31.1-1q44		2q22.2-2q22.1
	2q24.2-2q37.3		7p21.3			Xp11.3	1q24.2-1q21.1		5q13.1
	5p15.2-5p15.33		7p22.2				9p22.1-9p24.3		
	5q13.2-5q31.3		7p22.3				9p21.3-9p21.2		
	7p21.1-7p21.3		7p22.3				10q23.31-10q26.3		
	10p12.31-10p15.3		9p21.3				10q21.1-10q23.1		
	17q24.2-17q25.3		10q22.3				11p		
			10p11.22-10p11.21				12q		
			16q21-16q22.1				13q		
			20p13				16p11.2-16q24.3		
			Xp11.22-Xp11.21				17p11.2-17p13.3		
							18q11.2-18q12.2		
							18q12.2-18q23		
							20p		
							20q		
							21q		

Table S3.9: CNA regions identified in each tumor sample, continued.

Sample	GAINS			DELETIONS			LOH		
	Whole Chr	Broad	Focal	Whole Chr	Broad	Focal	Whole Chr	Broad	Focal
PGBM12	7	3q26.31-3q29				4q31.23	8	1q23.2-1q43	5q34
		22q13.2-22q13.33				10q22.1		2p	11q14.1
						17p13.3		3q12.1-3q26.1	11q22.1
								3q26.1-3q26.31	11q22.1
								4q	11q22.1-11q24.3
								8p21.1-8p12	15q26.3
								8p11.22-8p11.21	16q24.3
								9p24.1-9p24.3	18q23
								9p24.1-9p13.1	
								9q31.1-9q21.11	
								9q31.1-9q33.2	
								9q33.2-9q34.3	
								10p	
							10q11.21-10q25.1		
							10q25.1-10q26.3		
							14q		
							14q31.3-14q32.31		
							15q		
							16p		
							18q21.2-18q23		
							19p13.11-19p13.3		
							19q13.32-19q13.43		
PGBM13	20	1q	5q13.2-5q13.3				17	3p12.1-3q26.1	6q12
		12p12.3-12p13.33						4p15.31-4p16.3	
								4p15.31-4q35.2	
							11q24.3-11q25		

Table S3.9: CNA regions identified in each tumor sample, continued.

Sample	GAINS		DELETIONS		LOH	
	Whole Chr	Broad	Whole Chr	Broad	Whole Chr	Broad
PGBM14		4p15.33-4q13.1			2	3q
		4q26-4q35.2			6	4q13.1-4q26
		17q12-17q25.3			7	9p
					8	9q
					10	13q
					11	14q
					12	15q
					16	19q13.42-19q13.43
					18	21q
					19	22q
				20		
PGBM18						
		1q			10	3q27.2-3q29
		9q				8p23.2-8p23.3 12q24.33
PGBM19						
	X					
					1	13q
					2	14q
					3	15q
					4	
					5	
					6	
					8	
					9	
				10		
				11		
				12		
				16		
				17		
				19		
				20		
				21		
				22		

Table S3.9: CNA regions identified in each tumor sample, continued.

Sample	GAINS			DELETIONS			LOH		
	Whole Chr	Broad	Focal	Whole Chr	Broad	Focal	Whole Chr	Broad	Focal
PGBM20	19	10p12.2-10p15.3 15q25.1-15q26.3 16p13.12-16p13.3 18p11.31-18p11.21 20p13-20q11.23	2p24.3-2p24.2 2q14.2 2q14.3 15q23				9	1p 3q11.2-3q28 5q31.1-5q34 5q35.2-5q35.3 6p22.3-6p21.32 6q12-6q13 6q13-6q14.1 6q23.3-6q24.2 8p23.1-8p23.3 10q25.3-10q26.3 11p12-11q25 11p13-11p12 11p14.1-11p15.5 12p12.1-12p13.33 14q11.2-14q23.3 14q23.3-14q31.3 16p13.12-16q24.3 17p 17q 18q11.2-18q12.1 18q11.1-18q12.3 18q12.3 22q	5q23.2 6p24.3 6p22.3 6p21.2-6p21.1 6p21.1 6p12.1-6p11.2 6q15 6q22.31 6q24.3-6q25.1 6q25.3 6q25.3 6q26 6q27 6q27 11p14.111p13 18q12.3 18q21.1



Table S3.9: CNA regions identified in each tumor sample, continued.

Sample	GAINS			DELETIONS			LOH		
	Whole Chr	Broad	Focal	Whole Chr	Broad	Focal	Whole Chr	Broad	Focal
PGBM25		4q26-4q28.1	2q33.2-2q33.3		9p21.3-9p21.2	2q22.1-2q22.2	1	4q32.3-4q35.2	4p14-4p13
		7p	3q13.32			3q13.2	2	5q15-5q35.3	4q28.2
		8p	3q26.33-3q27.1			9p23	3	6p-6q13	17p13.3
		8q24.13-8q24.22	4q12			Xp21.2	9	6p12.1-6p11.2	
		10q26.3	8q11.21-8q11.23				10	6q13-6q25.3	
		15q25.1	8q24.22				11	6q25.3-6q27	
			9p24.3-9p24.2				16	8q12.1-8q23.3	
			10q21.1				17	8q24.22-8q24.3	
			10q25.1				18	12q	
			10q25.3				19	13q	
		10q26.2-10q26.3				20	14q		
							15q14-15q26.3		
							21q		
							22q		
PGBM26	16	9q33.3-9q34.3	5p13.2				2	1q42.13-1q44	19p13.2
	20	12q24.31-12p13.33	7q21.2				3	13q	19p13.2
		14q	10p12.2				4	15q	19p12-19p11
		21q	13q14.3				5	19p13.3	
			13q21.1				6	22q	
			13q31.1				8		
			13q33.1				10		
			13q33.2				17		
			13q34						
			19p13.12						
		19p13.11-19p13.12							



Table S3.9: CNA regions identified in each tumor sample, continued.

Sample	GAINS			DELETIONS			LOH		
	Whole Chr	Broad	Focal	Whole Chr	Broad	Focal	Whole Chr	Broad	Focal
PGBM34	3	1q32.1-1q21.2	1p36.22-1p36.21		9p21.3-9p21.1	6q12	10	1p	19p13.3
	4	1p36.21-1p33	1p32.2-1p32.1				11	2q14.3-2p23.1	19p13.2
	7	8q21.13-8q24.3	1p31.1				12	2q34-2q37.3	
		9q21.11-9q31.3	1p13.3				17	3p25.3-3p26.3	
		19p13.2-19p12	19p13.3					4q28.1-4q23.1	
		20p13-20q11.21	22q11.21-22q11.22					5q23.1-5q35.3	
		20q11.22-20q13.12	22q11.21					6p	
		21q	22q11.21					6q	
		22q13.2-22q13.31						8p12-8p23.3	
		Xp21.3-Xq27.3						9p	
		Xp21.3-Xp22.33						13q	
								14q	
								15q11.2-15q22.2	
							15q22.2-15q26.3		
							19p12-19q12		
							20q13.12-20q13.2		
							22q		
							Xq27.3-Xq28		
PGBM35		17q21.32-17q25.3	10q11.22				9	22q	
							16		
							18		
PGBM36			14q32.33						7p22.1
			18q23						7q22.1

Table S3.9: CNA regions identified in each tumor sample, continued.

Sample	GAINS			DELETIONS			LOH		
	Whole Chr	Broad	Focal	Whole Chr	Broad	Focal	Whole Chr	Broad	Focal
PGBM37		1q						2q14.1-2q31.3 6q12-6q27 8q24.22-8q24.3 9p21.1-9p24.3 16q22.1-16q24.3	9q34.3 20p12.1
PGBM39	3	21q			9p21.3		10	12p13.31-12p13.33	5p15.32-5p15.33
	7				9p21.3		12	12p12.3-12p11.23	5p15.2
	19				9p21.3		18		5p15.1 5p14.3 9p23 9p23 9p22.3 9p22.3-9p22.2 9p21.2 9p21.2-9p13.2 12p13.31 12p12.3
PGBM40									
PGBM41									8p11.22
PGBM42								22q11.23-22q13.33 (CN)	

Table S3.9: CNA regions identified in each tumor sample, continued.

Sample	GAINS			DELETIONS			LOH		
	Whole Chr	Broad	Focal	Whole Chr	Broad	Focal	Whole Chr	Broad	Focal
PGBM43		5p14.3-5p13.3	4q12			4q34.3	1	2p	4p12
		6q15-6p25.3	4q12			7p14.1	4	3q13.33-3q29	5p15.32-5p15.31
		9p13.3-9q34.3	4q12			11q22.3-11q23.1	7	5q11.2-5q21.3	5q31.3-5q32
		9p24.1-9p24.3	5p15.33			20q13.31	11	9p22.1-9p21.2	6q24.3
		10p12.1-10p13	5p15.33-5p15.32				18	10q	9p23-9p22.3
			5p15.2					12q21.2-12q23.1	10p11.23
			5p15.1					13q	12q24.31-12q24.32
			5p12					14q	16q23.2-16q23.3
			7q31.2-7q31.31					15q	19p13.3
			9p13.3					16q	
			9p21.1					17p	
			10q23.33-10q24.1						17q11.2-17q23.2
			10p11.21-10p11.1						19q
			10p11.22						20p
		13q13.3						20q	
		19p13.3						22q	
								Xp	
								Xq11.2-Xq21.31	
PGBM45									
		1q	2p24.3					1p32.2-1p34.1	2p23.3-2p24.1
			2p25.1						3p12.3-3p12.1
			2p25.1						6p25.2-6p25.3
			2p25.3						9p23
			7p11.2						10p15.1-10p14
		9p24.3						12q21.1-12q21.2	

### 3.11. CONNECTING TEXT CHAPTER 3 TO 4

Using whole exome sequencing, we identified recurrent mutations in *H3F3A* and mutations in *ATRX* in pediatric glioblastoma, described in Chapter 3. *H3F3A* mutations were highly specific to pediatric high-grade gliomas and only seen in 3.4% of adult GBM. *ATRX* mutations/loss of expression was found in 33.5% of pediatric GBM in our cohort, and another group identified *ATRX* mutations in 7% of adult GBM (Heaphy, de Wilde et al. 2011). Therefore, we question ourselves whether *ATRX* mutations/loss of expression was specific to pediatric disease. Given that *ATRX* mutations were mainly assessed in GBM, it is important to assess their incidence in lower grade and other histological subtypes of gliomas across age. We will address these questions regarding *ATRX* in Chapter 4 of this thesis, using targeted sequencing and immunohistochemistry. This will enable us to better understand the distinct pathogenesis of different glioma entities and to develop tailored therapy in the near future.

Chapter 4: Frequent *ATRX* mutations and loss of expression in  
adult diffuse astrocytic tumors carrying *IDH1/IDH2* and *TP53*  
mutations

Xiao-Yang Liu, Noha Gerges, Andrey Korshunov, Nesrin Sabha, Dong-  
Anh Khuong-Quang, Adam M Fontebasso, Adam Fleming, Djihad Hadjadj;  
Jeremy Schwartzentruber, Jacek Majewski, Zhifeng Dong, Peter Siegel,  
Steffen Albrecht, Sidney Croul, David TW Jones, Marcel Kool, Martje  
Tonjes, Guido Reifenberger, Damien Faury, Gelareh Zadeh, Stefan  
Pfister, Nada Jabado

*Acta Neuropathologica* (124(5):615-25, 2012)

Permission was granted to reproduce the manuscript in Chapter 4.

#### 4.1. ABSTRACT

Gliomas are the most common primary brain tumors in children and adults. We recently identified in pediatric and young adult glioblastoma (GBM, WHO grade IV astrocytoma) frequent alterations in chromatin remodelling pathways including recurrent mutations in *H3F3A* and mutations in *ATRX* ( $\alpha$ -thalassemia/mental-retardation-syndrome-X-linked). *H3F3A* mutations were specific to pediatric high-grade gliomas and identified in only 3.4% of adult GBM. Using sequencing and/or immunohistochemical analyses, we investigated *ATRX* alterations (mutation/loss of expression) and their association with *TP53* and *IDH1* or *IDH2* mutations in 140 adult WHO grade II, III and IV gliomas, 17 pediatric WHO grade II and III astrocytomas and 34 pilocytic astrocytomas. In adults, *ATRX* aberrations were detected in 33% of grade II and 46% of grade III gliomas, as well as in 80% of secondary and 7% of primary GBMs. They were absent in the 17 grade II and III astrocytomas in children, and the 34 pilocytic astrocytomas. *ATRX* alterations closely overlapped with mutations in *IDH1/2* ( $p < 0.0001$ ) and *TP53* ( $p < 0.0001$ ) in samples across all WHO grades. They were prevalent in astrocytomas and oligoastrocytomas but were absent in oligodendrogliomas ( $p < 0.0001$ ). No significant association of *ATRX* mutation/loss of expression and

alternative lengthening of telomeres was identified in our cohort. In summary, our data show that *ATRX* alterations are frequent in adult diffuse gliomas and are specific to astrocytic tumors carrying *IDH1/2* and *TP53* mutations. Combined alteration of these genes may contribute to drive the neoplastic growth in a major subset of diffuse astrocytomas in adults.

## 4.2. INTRODUCTION

Gliomas account for more than 70% of all primary central nervous system neoplasms and are the most common primary brain tumors in both children and adults. This heterogeneous group of tumors is classified by the World Health Organization (WHO) (Louis 2007) according to the presumed cell of origin into four major groups (astrocytic gliomas, oligodendroglial tumors, mixed oligoastrocytomas, and ependymal tumors). They are then further segregated into distinct histological grades, ranging from WHO grade I to WHO grade IV. WHO grading is based on defined cytologic and histologic features, including cellularity, mitotic activity, nuclear atypia, microvascular proliferation and necrosis. WHO grade I and II gliomas are commonly referred to as low grade gliomas (LGG), while WHO grade III (anaplastic gliomas) and IV (glioblastoma multiforme, GBM) tumors are regarded as high grade gliomas (HGG) (Louis 2007). The behavior and genetic alterations identified in gliomas vary according to age, histological type and tumor grade. Pilocytic astrocytoma, the most common WHO grade I glioma entity, is most prevalent in children, rarely progresses to higher grade tumors and is characterized by recently identified genetic alterations in the MAPK pathway, mainly *BRAF* gene fusions or activating point mutations (Jones,

Gronych et al. 2011; Ohgaki and Kleihues 2011). In contrast, WHO grade II (diffuse) astrocytomas are more common in adults and tend to progress to higher grade tumors, i.e., anaplastic astrocytoma and secondary GBM (Ohgaki and Kleihues 2011). The majority of GBM, however, are primary tumors that arise *de novo* and typically manifest in older patients.

Recent molecular markers have allowed for an improved stratification of adult gliomas. Mutations in the isocitrate dehydrogenase (IDH) 1 and 2 genes are detected in approximately 80% of diffuse and anaplastic gliomas as well as secondary GBM, but are rare in primary GBM (Parsons, Jones et al. 2008; Yan, Parsons et al. 2009). WHO grade II and III gliomas are characterized by two additional mutually exclusive genetic alterations, both of which are tightly associated with *IDH1/2*-mutant gliomas: *TP53* mutations, which are found in approximately 60% of diffuse and anaplastic astrocytic gliomas, and 1p19q co-deletions, which are detected in up to 80% of oligodendrogliomas (Bourne and Schiff 2010; Ohgaki and Kleihues 2011). Mixed oligoastrocytomas may carry either of these hallmark alterations in association with *IDH1/2* mutation. Thus, diffuse and anaplastic gliomas can be molecularly divided according to *IDH1/2*, *TP53*, and 1p19q status into at least four major groups,

corresponding to *IDH1/2+/TP53-/1p19q-*, *IDH1/2+/TP53+/1p19q-*, *IDH1/2+/TP53-/1p19q+* and “triple negative” tumors.

Pediatric gliomas rarely show *IDH1* and *IDH2* mutations, less commonly progress from lower to higher grade tumors, infrequently show oligodendroglial differentiation, and, in general, appear to carry distinct molecular and genetic alterations when compared to adult gliomas (Faury, Nantel et al. 2007; Yan, Parsons et al. 2009; Kieran, Walker et al. 2010; Paugh, Qu et al. 2010; Pollack, Hamilton et al. 2010; Qu, Jacob et al. 2010). Recently, we and others identified recurrent driver mutations in *H3F3A*, which encodes histone 3 variant 3 (H3.3), in 36% of pediatric GBM (Schwartzentruber, Korshunov et al. 2012; Wu, Broniscer et al. 2012). We also identified *ATRX* mutations/loss of expression in 33.5% of pediatric GBM (Schwartzentruber, Korshunov et al. 2012), while another group identified *ATRX* mutations in only 7% of adult GBM (Heaphy, de Wilde et al. 2011). *ATRX* mutations were mainly assessed in GBM, leading us to question their incidence in other gliomas across entity, WHO grade and age. Therefore, using DNA sequencing and immunohistochemistry we investigated the status of *ATRX*, *IDH1*, *IDH2* and *TP53* in a set of 140 WHO grade II, III and IV gliomas of adult

patients, 17 WHO grade II (8) and III (9) gliomas of pediatric patients, and 34 pilocytic astrocytomas. We also investigated the potential association of *ATRX* alterations with *TP53* and *IDH1/2* mutations, and determined the presence of alternative lengthening of telomeres (ALT) in our sample set, based on its previous association with *ATRX* mutations (Heaphy, de Wilde et al. 2011; Schwartzenuber, Korshunov et al. 2012).

### 4.3. MATERIALS AND METHODS

#### **Sample characteristics and pathological review.**

Tissue samples were obtained with informed consent and according to the guidelines of the Institutional Review Boards of the respective clinical centers in Montreal, Toronto, Düsseldorf and Heidelberg. All tumors were histologically classified based on the criteria of WHO classification of tumors of the central nervous system (Louis, Ohgaki et al. 2007). Tumor tissue samples from 140 adult glioma patients aged 18 years or older, as well as 51 pediatric patients below the age of 18 (34 pilocytic astrocytoma patients and 17 WHO grade II or III glioma patients) were included in this study. All samples included in this study were wild-type for *H3F3A*. Clinical

characteristics of the patients are summarized in Table S4.1a (adult) and b (pediatric).

**Tissue microarray (TMA) composition and immunohistochemical (IHC) analyses for expression of ATRX, IDH1R132H and TP53.**

Formalin-fixed, paraffin-embedded sections of three adult glioma TMAs and one pediatric pilocytic astrocytoma TMA were obtained. TMAs were comprised of an average of 3 tumor cores from the same sample with a mean diameter of 1.5 mm for each core. Cores were selected from the original tumor sample and oriented on the TMA by the respective neuropathologist at each center they were made in (SC, GR). All TMA blocks were control-stained with hematoxylin and eosin and independently reviewed for adequate tumor representation in the individual tissue cores. TMA1 contained 11 *de novo* anaplastic astrocytomas, TMA2 contained 63 astrocytomas, oligoastrocytomas and oligodendrogliomas (WHO grades II and III), and TMA3 consisted of 22 WHO grade II to grade IV astrocytomas (Table S4.1a). The pilocytic astrocytoma TMA contained cores from 34 different samples. TMAs were immunohistochemically stained and independently scored for ATRX positivity by three individuals and results merged after consensus scoring was obtained as described

(Jiao, Shi et al. 2011; Schwartzenuber, Korshunov et al. 2012). Briefly, samples were considered as negative for ATRX staining if tumor cells showed no nuclear positivity and the core included a control brain or blood vessel with positive ATRX nuclear staining (Jiao, Shi et al. 2011; Schwartzenuber, Korshunov et al. 2012). IDH1 and p53 immunohistochemical staining and scoring were performed as described (Capper, Weissert et al. 2010).

#### **DNA sequencing.**

Sanger sequencing was used to identify mutations in the *IDH1*, *IDH2*, *TP53* and *ATRX* genes as reported elsewhere (Schwartzenuber, Korshunov et al. 2012). Forty-four adult and 17 pediatric glioma frozen samples were analyzed for ATRX mutations (Table S4.1a and b). In addition, two samples with sufficient quality DNA were analyzed by sequencing for *ATRX* mutation and results were correlated to IHC staining. Samples from TMA1-3 were also analyzed using pyrosequencing for *IDH1* or *IDH2* mutations and by single-strand conformation polymorphism (SSCP) and/or Sanger sequencing for *TP53* mutations whenever possible (Table S4.1a).

**Statistical analysis.**

Fisher's Exact Test was used to identify statistically significant associations or differences between two groups.

**Telomere specific fluorescence *in situ* hybridization (FISH).**

Telomere-specific FISH was done with a standard formalin-fixed paraffin-embedded FISH protocol (Heaphy, de Wilde et al. 2011; Schwartzenuber, Korshunov et al. 2012), using a FITC peptide nucleic acid telomere probe from Dako.

**Gene expression profile (GEP) analysis.**

GEP data from two previously studied cohorts of glioma samples were used for analysis: 284 in Gravendeel *et al* (Gravendeel, Kouwenhoven et al. 2009) and 484 untreated primary GBM from the Cancer Genome Atlas Consortium (Parsons, Jones et al. 2008; TCGA 2008; Verhaak, Hoadley et al. 2010). R2 microarray analysis and visualization platform was used to analyze GEP (<http://r2.amc.nl>).

#### 4.4. RESULTS

##### **ATRX alterations characterize subgroups of adult gliomas and are not identified in WHO grade I, II and III pediatric gliomas**

We assessed alterations in *ATRX* in 140 adult glioma samples using either Sanger sequencing (46 samples) or immunohistochemical (IHC) staining (96 samples, including 2 with sequencing data), which were distributed across grade and subtype (Table S4.1a and b). We also performed IHC on 34 pilocytic astrocytomas. We found non-silent mutations in *ATRX* in 22 of 46 (47.8%) adult glioma samples investigated by Sanger sequencing (Table 1, 37.0% truncating and 11.4% missense mutations), while all 17 pediatric samples that we sequenced were wild-type for this gene (Table S4.1b). Immunohistochemical analysis showed loss of expression of *ATRX* in tumor cells but not in neighbouring blood vessels in 31/96 (32%) of adult gliomas and no loss of expression in the 34 pilocytic astrocytomas (Figure S4.1, Table S4.1a-b, Table 4.2). The 2 samples mutant for *ATRX* by sequencing had truncating mutations and lacked detectable *ATRX* expression.

These findings confirm a strong correlation between *ATRX* mutation and its loss of expression as detected by immunohistochemistry, as we and

others reported previously (Heaphy, de Wilde et al. 2011; Jiao, Shi et al. 2011; Schwartzentruber, Korshunov et al. 2012). The lower percentage of ATRX loss found by IHC analysis might be explained by cases having ATRX expression but with missense mutations (likely loss-of-function), which were found in 11% of our sequencing cohort. In agreement with previous reports, *ATRX* missense mutations were uniformly located in evolutionarily highly conserved regions of the gene, mainly in the helicase domains (Figure 4.1). They spared the ADD domain where 50% of the missense mutations identified in the ATRX syndrome occur (Gibbons, Wada et al. 2008). There was no gender bias. We sequenced *ATRX* in constitutive DNA from 7 patients with glioma carrying an *ATRX* mutation where we had access to this material and did not identify germline mutations (4 truncating, 3 missense mutations in the tumors). Based on these results, somatic ATRX alterations are present in at least 51/140 (36.4%) of our adult gliomas. This mutation rate is higher than that observed in pancreatic neuroendocrine tumors (17.6%) (Jiao, Shi et al. 2011), adult GBM (7.1%) (Heaphy, de Wilde et al. 2011), and non-infantile neuroblastomas (20.9%)(Cheung, Zhang et al. 2012), but is similar to that seen in pediatric GBM (33.5%) (Schwartzentruber, Korshunov et al. 2012) (Table 4.1).

**ATRX alterations characterize adult gliomas with *IDH1* or *IDH2* mutation**

Mutations in *IDH1* and *IDH2* are found in approximately 80% of WHO grade II and III gliomas, as well as in secondary GBMs (Parsons, Jones et al. 2008; Yan, Parsons et al. 2009; Ohgaki and Kleihues 2011), and thus are thought to occur at an early stage of tumor development. Our findings are concordant with this, as we identified *IDH1/2* mutations in 83% of the WHO grade II and 78% of the WHO grade III adult gliomas and in 26% of the GBM (our sample set was enriched for secondary GBM) (Table 4.2). Of the 51 adult samples with ATRX alteration, 47 had associated mutations in either *IDH1* or *IDH2*, while 4 were wild-type for both genes, indicating a strong association between ATRX and *IDH1/2* mutations ( $p < 0.0001$ , Fisher's exact test) (Table 4.2, Figure 4.2). Notably, none of the pediatric WHO grade II or III tumors had ATRX or *IDH* mutations.

To further validate preponderance of ATRX alterations in *IDH* mutant gliomas, we interrogated two distinct gene expression datasets derived from adult gliomas (Parsons, Jones et al. 2008; TCGA 2008; Gravendeel, Kouwenhoven et al. 2009; Verhaak, Hoadley et al. 2010). Looking at GBM only (TCGA dataset), we noticed that in particular in proneural GBMs there was a subset of tumors with decreased ATRX expression (Figure 4.3a). K-means cluster analysis of all proneural GBMs ( $n = 106$ ) in this

dataset using gene expression data of 73 genes that significantly correlate with ATRX mRNA expression showed that the cluster with low ATRX expression contains all cases with *IDH1* mutations (ATRX-low 12/39 vs ATRX-high 0/67;  $p < 0.0001$ ) (Figure 4.3b). Extending these analyses to other gliomas using data from Gravendeel et al. also showed that tumors with decreased ATRX mRNA expression (mainly astrocytomas, oligoastrocytomas and glioblastomas) are strongly associated with *IDH1* mutations (ATRX-low 33/63 vs ATRX-high 20/153;  $p < 0.0001$ ) (Figure 4.3c and 4.3d). Thus, significant decrease in ATRX expression was associated with IDH mutant gliomas in both datasets ( $p < 0.0001$ ) further validating the overlap of ATRX alterations with IDH mutant gliomas. Also supporting this, ATRX alterations were found more frequently in younger adults, and the age range mirrored what is seen for *IDH1/2* mutations (Figure S4.2).

**ATRX alterations characterize diffuse adult *IDH1/2*-mutant astrocytic tumors and are absent in oligodendrogliomas**

ATRX alterations were observed in 33% (15/46) of WHO grade II gliomas and 46% (26/56) of WHO grade III gliomas (Table 4.2). When analyzing data based on glioma subgroup, we found ATRX alterations to be present exclusively in tumors of an astrocytic lineage (astrocytomas or

oligoastrocytomas), and did not identify ATRX loss of expression or mutation in the 19 oligodendrogliomas included in our series (Table 4.2). Thus, we observed IDH mutation in respectively 100% (15/15) and 92% (24/26) of grade II and III ATRX-altered-gliomas harboring an astrocytic component suggesting that *ATRX* mutations also occur early in tumorigenesis. In addition, ATRX alterations occurred in 26% (10/38) of adult GBM samples and significantly overlapped with those GBM carrying mutations in *IDH1/2* (8/10, Table 4.2), while limited overlap was seen in primary GBM (2/28,  $p < 0.0001$ ). Increase in the prevalence of ATRX alterations from grade II to IV astrocytomas (33% to 46% to 80% respectively) argues that this loss is a marker of progression however based on small numbers we did not reach statistical significance. Furthermore, when we included two further studies in which *ATRX* was sequenced in a total of 20 oligodendrogliomas (Bettegowda, Agrawal et al. 2011; Heaphy, de Wilde et al. 2011), only one sample of 39 carried an *ATRX* mutation, suggesting that this event does not target cells of the oligodendroglial lineage among the *IDH1/2*-mutant diffuse gliomas ( $p < 0.0001$ , Fisher's exact test) (Figure S4.3). Last, the gene expression data from two published adult gliomas datasets (TCGA and Gravendeel (Gravendeel, Kouwenhoven et al. 2009) also show decreased ATRX

mRNA levels in astrocytomas and oligoastrocytomas, but not in oligodendrogliomas (Figure 4.3c).

### **ATRX alterations are associated with *TP53* mutations in *IDH1/2*-mutant gliomas**

We screened glioma samples in our cohort for *TP53* mutations using DNA sequencing (all pediatric gliomas and 46 adult gliomas) and/or by immunohistochemistry (Table S4.1a), where positive TP53 staining was used as a surrogate for altered *TP53* function as previously described (Tabori, Baskin et al. 2010). *TP53* mutations are known to characterize gliomas of an astrocytic lineage and to be associated with *IDH1/2* mutations (Bourne and Schiff 2010; Ohgaki and Kleihues 2011), and this was true in our dataset ( $p < 0.0001$ , Table 4.2). In the 103 adult WHO grade II, III and IV glioma samples for which *ATRX*, *IDH1/2* and *TP53* status was available we found a significant association between *ATRX* alteration and *TP53* mutation ( $p < 0.0001$ ), as well between *ATRX* and *TP53* mutations in *IDH1/2*-mutant gliomas ( $p = 0.0295$ ) (Table 4.2, Figures 4.2 and 4.4). The observed close association between *ATRX*, *IDH1/2* and *TP53* mutations further suggests that *ATRX* alterations are likely needed with *TP53*

mutations in *IDH1/2*-mutant astrocytomas or oligoastrocytomas (astrocytic lineage) to promote oncogenesis in adult gliomas.

### **Presence of ATRX alteration is not significantly linked to ALT in adult gliomas and characterize young adult**

ALT has been shown to occur in *ATRX*-mutated pancreatic neuroendocrine tumors (PanNETs) and GBM (Heaphy, de Wilde et al. 2011; Schwartzenuber, Korshunov et al. 2012). We investigated the presence of ALT on TMAs from the adult glioma samples included in this study using telomere FISH as reported (Heaphy, de Wilde et al. 2011; Schwartzenuber, Korshunov et al. 2012). ALT was identified in 22%, 24% and 7% of WHO grade II, grade III and grade IV tumors respectively (Table 4.2, Figure S4.4). Unlike PanNETs and pediatric GBMs, we did not find an association between *ATRX* alterations and ALT (TableS4. 2).

## **4.5. DISCUSSION**

We identify herein *ATRX* mutation or loss of expression in 36% of diffuse gliomas of adult patients. *ATRX* alterations significantly overlapped with *IDH1/2* mutations ( $p < 0.0001$ ), *TP53* mutations ( $p < 0.0001$ ), and

*IDH1/2/TP53* mutant adult gliomas ( $p=0.03$ ). As *ATRX* missense mutations may not lead to loss of ATRX expression as detected by immunohistochemistry, the overall frequency of ATRX aberrations in adult gliomas we report is likely to be an underestimation as also suggested by the lower percentage of ATRX alteration we detected using this approach.

ATRX is a critical member of a multiprotein complex that includes DAXX and plays a role in regulating chromatin remodeling, nucleosome assembly, telomere maintenance and deposition of histone H3.3 at transcriptionally silent regions of the genome (Bassett, Cooper et al. 2008; Schneiderman, Sakai et al. 2009; Emelyanov, Konev et al. 2010).

Mutations in this gene lead to the  $\alpha$ -thalassemia/mental retardation X-linked syndrome (Gibbons, Picketts et al. 1995) and are thought in this syndrome to be hypomorphic, retaining partial activity of the protein (Gibbons, Wada et al. 2008). On the other hand, complete loss-of-function mutations have recently been identified in several cancers including PanNETs (Jiao, Shi et al. 2011), neuroblastoma (Cheung, Zhang et al. 2012; Molenaar, Koster et al. 2012), alpha-thalassemia myelodysplasia syndrome (Gibbons, Pellagatti et al. 2003), pediatric GBM (Heaphy, de Wilde et al. 2011; Schwartzenuber, Korshunov et al. 2012), and in 36%

of our samples in this study. The use of temozolomide in adult gliomas has been associated with increased gene mutation rate in tumor tissue samples (TCGA 2008). However, this does not explain the high rate of *ATRX* mutation we observed in adult gliomas, as these were identified in WHO grade II and III gliomas patients at diagnosis prior to therapy and in similar proportion to WHO grade II and III patients who received prior treatment including temozolomide.

Overlap of *IDH1/2* mutations and *ATRX* alterations argues for a specific role of *ATRX* in IDH-driven gliomagenesis. The presence of *ATRX* alterations in WHO grade II adult gliomas (33%) suggest that loss of *ATRX* function is an important driving force for the initial development of this subgroup of *IDH1/2*-mutant gliomas. This event is likely to occur subsequently to *IDH1/2* mutations, as several WHO grade II gliomas carried *IDH1/2* mutations without *ATRX* alteration, while none of the WHO grade II tumors and only two of the WHO grade III adult gliomas had *ATRX* alterations without concomitant *IDH1/2* mutation. Also in keeping with the fact that *ATRX* mutations may not be an initiating event is the lack of increased incidence of cancers including gliomas in patients with the *ATRX* syndrome (Gibbons 2006). Notably, we identify an increase in the

incidence of *ATRX* alterations with grade arguing that this loss may be a marker of progression in diffuse adult astrocytic tumors, however our numbers were too small to reach statistical significance. The H3.3 mutations we identified were mutually exclusive with *IDH1* and *IDH2* mutations, while they significantly overlapped with mutations in *TP53* and *ATRX* (Schwartzentruber, Korshunov et al. 2012). This may account for lack of *ATRX* mutations in pediatric diffuse and anaplastic gliomas as well as pilocytic astrocytomas, as these tumors rarely carry if at all *IDH1/2* or *H3F3A* mutations.

*ATRX* alterations seem to be specific to *IDH1/2* and *TP53* mutant glioma of astrocytic lineage. They did not occur in oligodendrogliomas, which are characterized by *IDH1* or *IDH2* mutations and 1p19q co-deletion. We did not have 1p19q co-deletion data on our sample set. However, co-deletion of these chromosomal regions is mostly restricted to oligodendrogliomas and a subset of oligoastrocytomas with wild-type *TP53*, which we show not to be significantly associated with *ATRX* alterations (Table 4.2, Figure 4.2). This suggests that *ATRX* alterations are mutually exclusive to 1p19q co-deletion, similar to *TP53* mutations. The *IDH1/2* and *TP53* mutant group of gliomas carries an intermediate prognosis between *IDH1/2*-

mutant and 1p19q-deleted gliomas (best prognosis) and *IDH1/2*, *TP53* and 1p19q wild-type gliomas (worst prognosis). Due to the significant overlap with *IDH1/2* and *TP53* mutations, we could not judge whether the added presence of ATRX alteration on its own affected survival or progression in gliomas.

In this study, a larger cohort of 140 glioma identified ALT in 20% of adult glioma samples (Table 4.2, Figure S4.4), a much lower frequency than the 61% in PanNETs or the 36% in pediatric GBM. We found no correlation between ATRX alteration and ALT assessed by FISH, nor between WHO grade and ALT, suggesting that mechanisms other than ATRX alterations can promote ALT in tumors.

Increasing evidence suggests that epigenetic regulators are important drivers for a variety of cancers (Elsasser, Allis et al. 2011), among which ATRX plays a critical role in regulating chromatin states. Loss of function or expression of ATRX can lead to chromosomal instability such as aneuploidy (Baumann, Viveiros et al. 2010). The significant overlap of *ATRX* mutations with *IDH1/2* and *TP53* mutations in adult gliomas mirrors the overlap we identified between *H3F3A*, *ATRX* and *TP53* mutations in

pediatric hemispheric and thalamic GBM. *ATRX* mutations were only present in older children with GBM of the brainstem (Khuong-Quang, Buczkowicz et al. 2012), similar to patients with neuroblastoma (Cheung, Zhang et al. 2012). Our present findings also show preferential *ATRX* mutation in younger age adults. Interestingly, while we identify significant overlap between *ATRX* alterations and *TP53* mutation, in PanNETs the mutations in these genes were mutually exclusive and occurred in cells from a different lineage (Jiao, Shi et al. 2011). This indicates that specific combinations of mutations with different mechanisms of oncogenesis affect chromatin states based on tumor type, lineage and age.

In conclusion, our results argue for the importance of an *ATRX* & *IDH1/2* & *TP53* mutant phenotype in the early development and progression of adult gliomas of astrocytic lineage. The exact role of *ATRX* loss of function and its impact on survival and therapeutic strategies targeting this group of gliomas warrant further studies.

#### 4.6. ACKNOWLEDGEMENTS

This work was supported by the Cole Foundation, and was funded by the Canadian Institute for Health Research (CIHR) (NJ). X.-Y.L. and A.M.F. are recipients of studentship awards from CIHR. N.J. is the recipient of a Chercheur Clinicien Award from Fonds de Recherche en Santé du Québec.

#### 4.7. CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

#### 4.8. AUTHOR CONTRIBUTIONS

XY Liu, A Korshunov, N Sabha, DA Khuong-Quang, AM Fontebasso, D Hadjadj, Z Dong, P Siegel, S Albrecht, S Croul, M Tonjes and D Faury performed experimental work. XY Liu, N Gerges, A Korshunov, J Schwartzentruber, J Majewski, D TW Jones, M Kool, M Tonjes and N Jabado performed data analyses and produced text and figures. A Korshunov, N Sabha, A Fleming, G Reifenberger, G Zadeh, S Pfister collected patient material and data. N Jabado provided leadership for the project.

#### 4.9. FIGURES

**Figure 4.1: Schematic representations of the localization of missense mutations identified in ATRX in this study and other cancers.**

Black: adult gliomas in this study; Blue: PanNET; Red: adult GBM; Green: non-infantile neuroblastoma; Orange: pediatric GBM. (Heaphy, de Wilde et al. 2011; Jiao, Shi et al. 2011; Cheung, Zhang et al. 2012; Schwartzenuber, Korshunov et al. 2012).

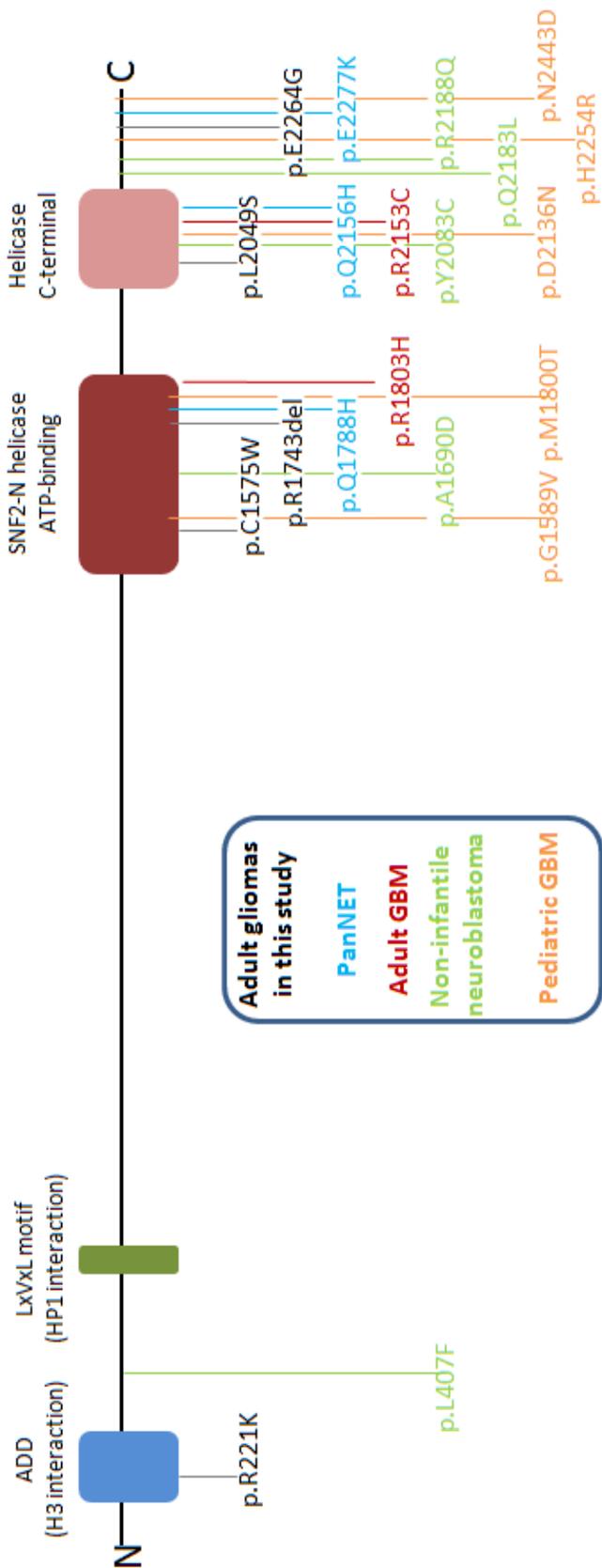
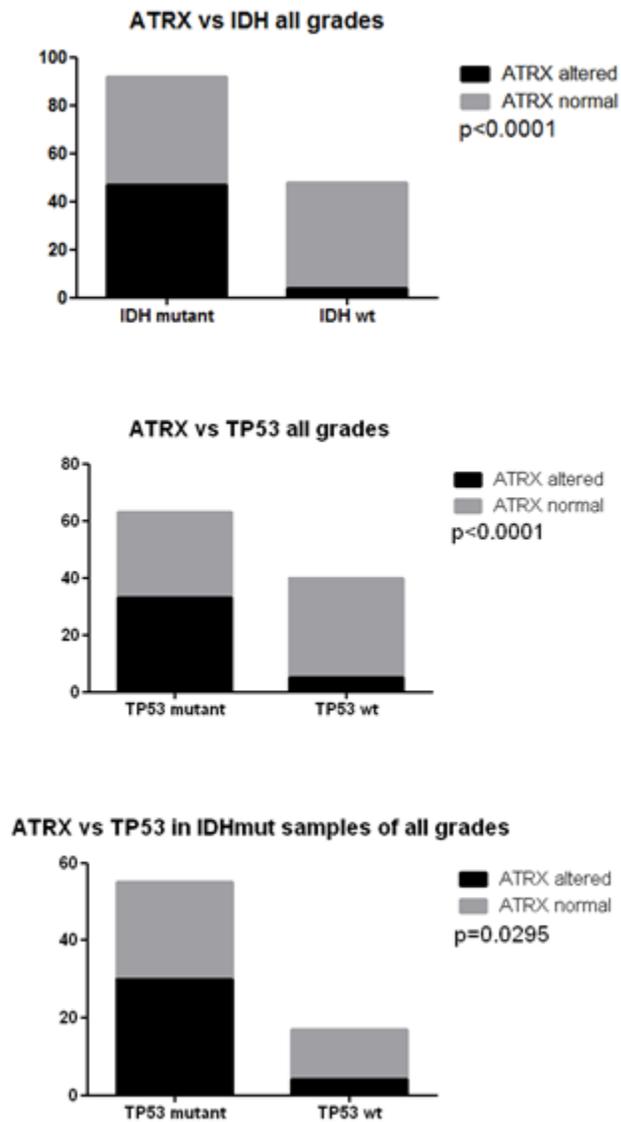


Figure 4.2: ATRX alterations significantly overlap with IDH1 or IDH2, and TP53 mutations in adult gliomas.

140 adult gliomas of all grades are included; due to the rareness of *IDH1* and *IDH2* mutations in pediatric gliomas, they are not included in this figure.



**Figure 4.3: IDH1 mutations are frequent in astrocytomas with decreased ATRX mRNA expression.**

a. Analysis of 484 primary glioblastomas of the TCGA consortium shows that a subset of the proneural GBMs (n=106) have decreased *ATRX* expression. b. K-means cluster analysis of 73 genes that significantly correlate with *ATRX* gene expression in proneural GBMs shows that in the cluster with low *ATRX* expression 12/39 cases have an *IDH1* mutation, whereas none of the 67 cases in the cluster with high *ATRX* expression have an *IDH1* mutation. c. *ATRX* expression is decreased in subsets of astrocytomas grade II and III, oligoastrocytomas grade II and III and glioblastomas grade IV using data from Gravendeel *et al.* (PA: pilocytic astrocytoma. A: astrocytoma. OA: oligoastrocytoma. OD: oligodendroglioma.) d. K-means clustering of 194 genes that significantly correlate with *ATRX* expression in astrocytomas, oligodendrogliomas and glioblastomas identified two clusters: cluster with low *ATRX* expression where 33/63 cases have an *IDH1* mutation and a cluster with high *ATRX* expression where 20/153 cases have an *IDH1* mutation (Parsons, Jones *et al.* 2008; TCGA 2008; Gravendeel, Kouwenhoven *et al.* 2009; Verhaak, Hoadley *et al.* 2010).

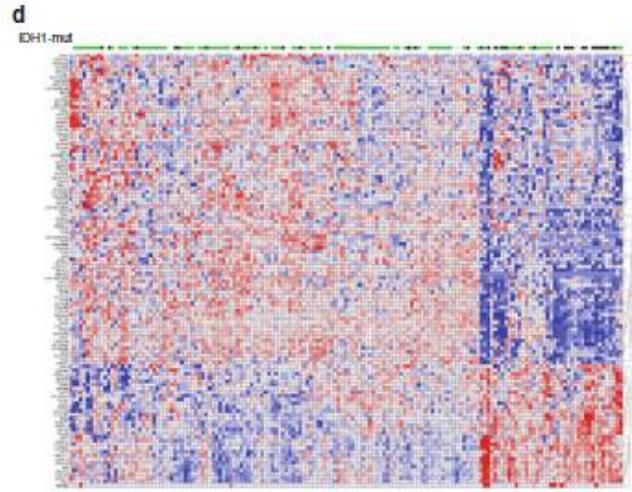
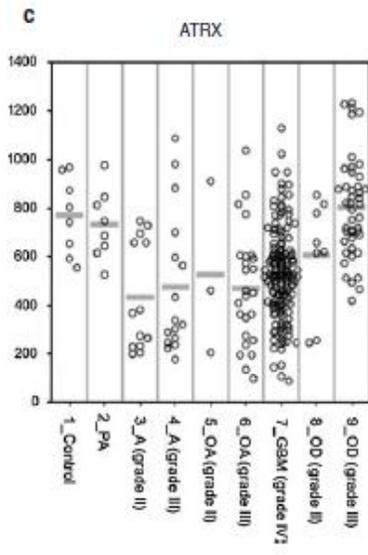
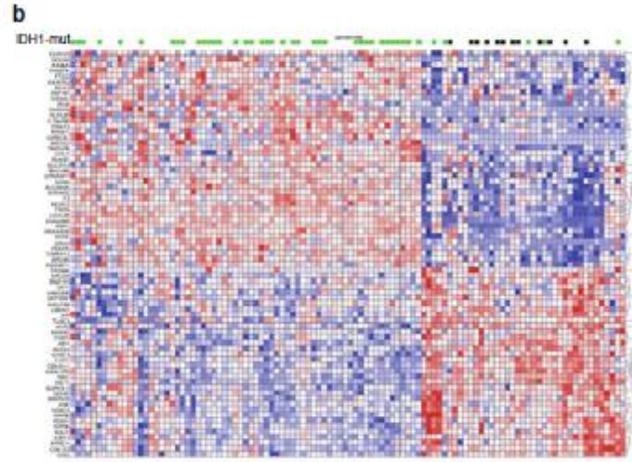
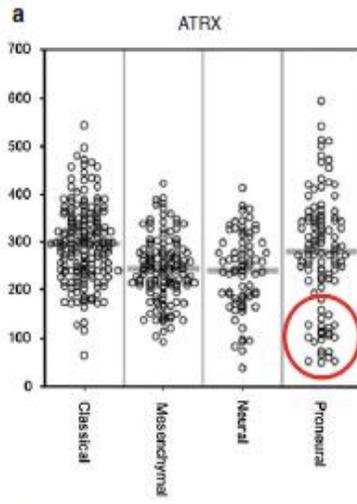
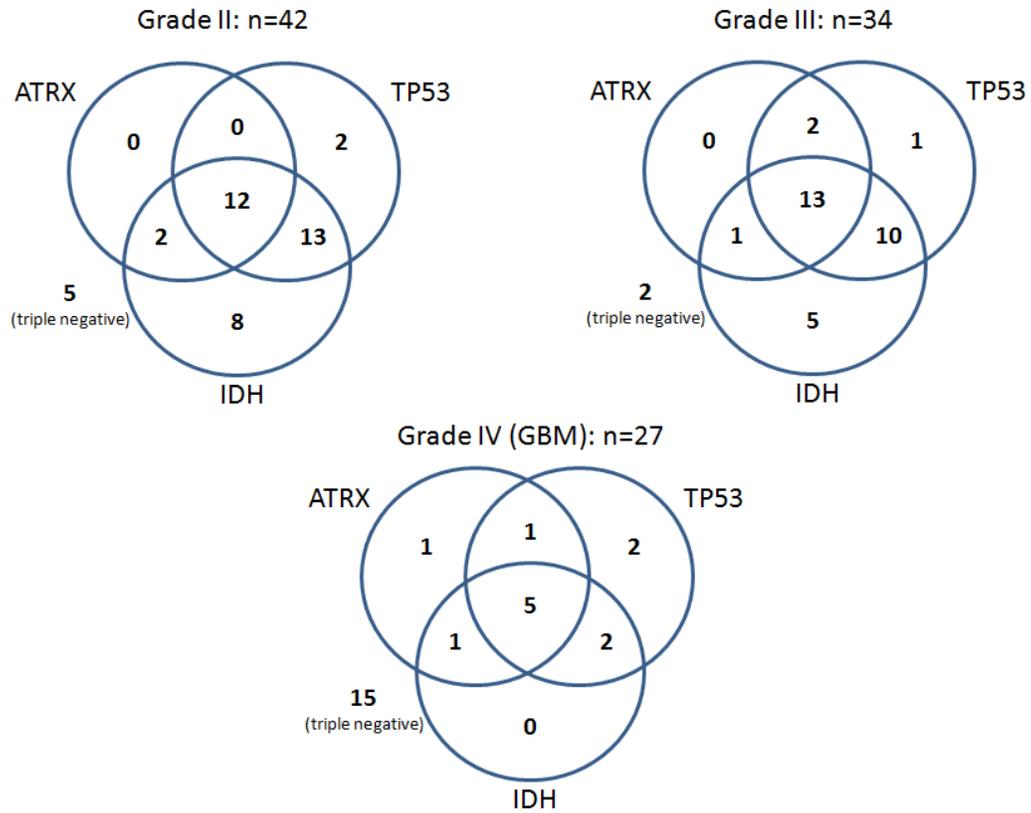


Figure 4.4: ATRX alterations are frequent in IDH1 (or IDH2) and TP53 mutant adult gliomas regardless of grade.



#### 4.10. TABLES

**Table 4.1: ATRX mutations detected by Sanger sequencing in gliomas, other published cancer datasets and the ATRX syndrome.**

(Gibbons, Wada et al. 2008; Bettgowda, Agrawal et al. 2011; Heaphy, de

Wilde et al. 2011; Jiao, Shi et al. 2011; Cheung, Zhang et al. 2012;

Molenaar, Koster et al. 2012; Schwartzentruber, Korshunov et al. 2012).

ATRX mutations	Gliomas						Non-infantile Neuroblastoma (Cheung <i>et al</i> , Molenaar <i>et al</i> ) N=134	ATRX syndrome (Gibbons <i>et al</i> ) N=126 <sup>b</sup>
	adult gliomas (this study) N=46	adult GBM (Heaphy <i>et al</i> ) N=12	oligodendroglioma (Heaphy <i>et al</i> , Bettgowda <i>et al</i> ) N=20	pediatric GBM (Schwartzentruber <i>et al</i> ) N=48	PanNET (Jiao <i>et al</i> ) N=68			
indel	15 (32.6%) <sup>a</sup>	4 (3.6%)	1 (5.0%)	6 (12.5%)	6 (8.8%)	21 (15.7%) <sup>d</sup>	17 (13.5%) <sup>c</sup>	
nonsense	2 (4.5%)	1 (0.9%)	0 (0%)	4 (8.3%)	3 (4.4%)	1 (0.7%)	8 (6.3%)	
missense	4 (11.4%)	2 (1.8%)	0 (0%)	4 (8.3%)	3 (4.4%)	5 (3.7%)	81 (64.3%)	
splicing	1 (2.2%)	1 (0.9%)	0 (0%)	0 (0%)	0 (0%)	1 (0.7%)	18 (14.3%)	
large duplication	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (1.6%)	
<b>total non-silent</b>	<b>22/46(47.8%)</b>	<b>8/112 (7.1%)</b>	<b>1/20 (5.0%)</b>	<b>14/48 (29.2%)</b>	<b>12/68 (17.6%)</b>	<b>28/134 (20.9%)</b>	<b>126 (100%)</b>	

<sup>a</sup> One of the indel mutations is in frame deletion, all the others are frameshift indels.

<sup>b</sup> different types of ATRX mutations identified in alpha thalathemia mental retardation syndrome (ATRX-X syndrome, adapted from Gibbons *et al*).

<sup>c</sup> 5 of the 17 different indel mutations are large deletion/insertions spanning more than 1kb.

<sup>d</sup> 18 non-infantile neuroblastoma samples have large deletions spanning one or more exons.

**Table 4.2: *ATRX*, *IDH1*, *IDH2* and *TP53* alterations and ALT detected in different types of adult gliomas.**

ATRX alterations were identified using immunohistochemistry (n=96) or Sanger sequencing (n=46) in adult gliomas.

Grade	Diagnosis	IDH1/2 mutation	TP53 mutation	IHC		Sequencing			ATRX alteration (IHC/sequencing)	Combined IDH1/2 mutation and ATRX alteration <sup>b</sup>	Combined IDH1/2, TP53 mutation and ATRX alteration <sup>b</sup>	ALT by FISH
				ATRX loss of expression	ATRX loss of expression and IDH mutation	ATRX truncating mutation	Combined ATRX truncating and IDH mutation	ATRX non-silent mutation				
Grade II	All	25/29 (86%)	18/28 (64%)	10/26	10/10	3/3	4/4	4/4	13/29	13/13	10/12	7/26 (27%)
	OAI	6/8 (75%)	5/8 (62%)	1/7	1/1	1/1	1/1	1/1	2/8	2/2	2/2	1/7 (14%)
	OII	7/9 (78%)	4/6 (67%)	0/8	0/1	0/1	0/1	0/1	0/9	0/0	0/0	1/8 (12%)
Grade III	Total grade II	38/46 (83%)	27/42 (64%)	11/41 (27%)	11/11	4/4	5/6	5/5	15/46 (33%) <sup>a</sup>	15/15	12/14	9/41 (22%)
	AIII	22/33 (67%)	19/22 (86%)	16/30	14/16	4/4	4/4	4/4	19/33	17/19	12/15	4/19 (21%)
	OAI	12/13 (92%)	3/4 (75%)	1/3	1/1	3/3	6/10	6/6	7/13	7/7	1/1	0/2 (0%)
Grade IV	OIII	10/10 (100%)	4/8 (50%)	0/8	0/0	0/2	0/2	0/2	0/10	0/0	0/0	3/8 (38%)
	Total grade III	44/56 (78%)	26/34 (76%)	17/41 (41%)	15/17	7/16	7/7	10/10	26/56 (46%) <sup>a</sup>	24/26	13/16	7/29 (24%)
	IDH + GBM	-	-	3/3	NA	5/7	NA	5/7	8/10 (80%)	-	-	-
All grades	IDH - GBM	-	-	0/1	NA	1/7	NA	2/7	2/8 (7%)	-	-	-
	GBM	10/38 (26%)	10/27 (37%)	3/14 (21%)	3/3	6/24	5/6	7/24	10/38 (26%)	8/10	5/8	1/14 (7%)
				31/96 (32%)					51/140 (36%)	47/51		17/24 (20%)

ATRX alterations were identified using immunohistochemistry ( $n = 96$ ) or Sanger sequencing ( $n = 46$ ) in adult gliomas

NA not applicable, IHC immunohistochemistry, A astrocytoma, O oligodendroglioma, GBM glioblastoma, ALT alternative lengthening of telomeres

<sup>a</sup> One sample was analyzed for ATRX alteration by both IHC and sequencing

<sup>b</sup> The numbers are out of the number of samples with ATRX alteration

#### 4.11. SUPPLEMENTARY FIGURES

Figure S4.1: ATRX immunohistochemical staining of a tissue microarray, with examples of a negative and a positive core at high magnification.

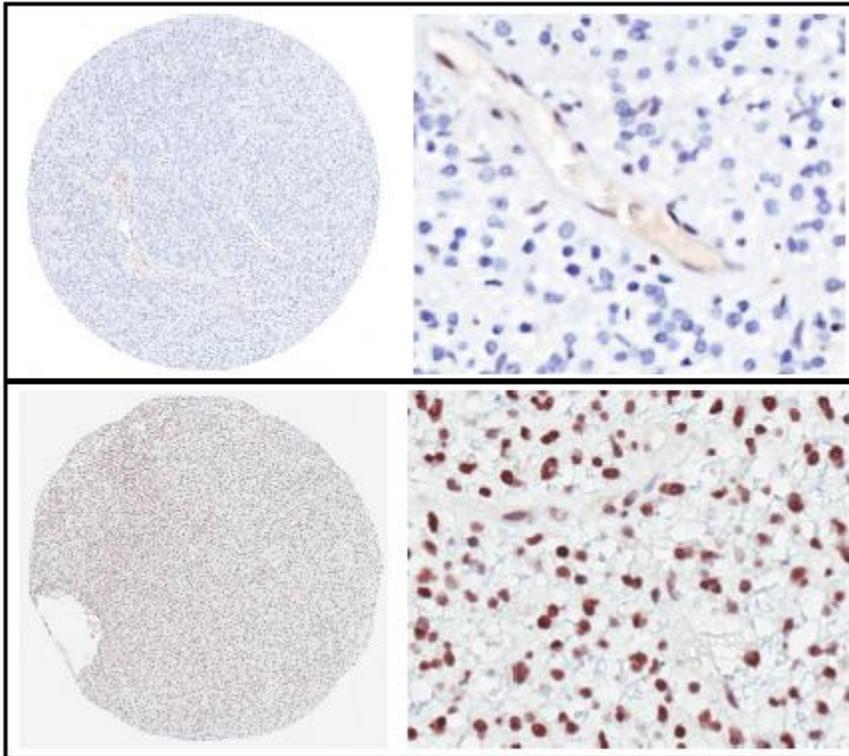


Figure S4.2: GBM samples age distribution between wild type and mutants in ATRX (a) and IDH (b).

Whiskers represent 10-90 percentile.

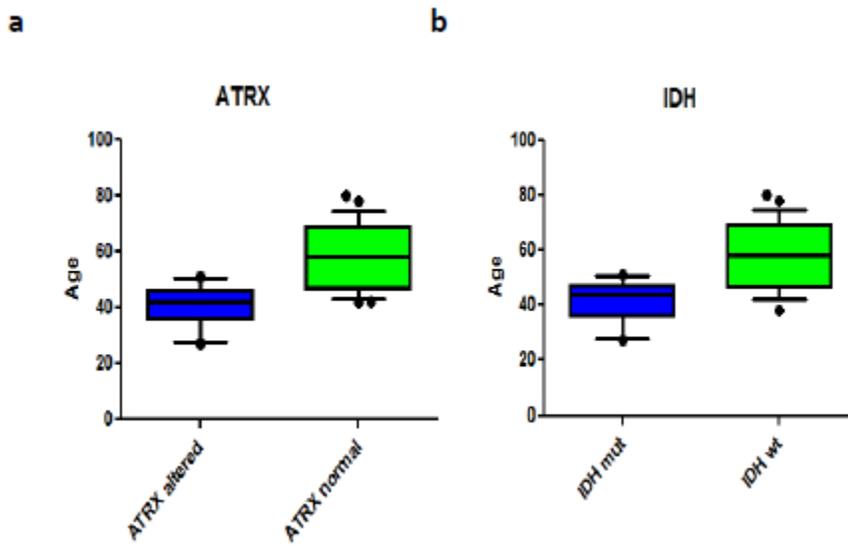


Figure S4.3: ATRX alteration is frequent in astrocytomas and oligoastrocytomas, but does not target oligodendrogliomas.

(Bettegowda, Agrawal et al. 2011; Heaphy, de Wilde et al. 2011).

	Oligodendroglioma (grade II&III)	Astrocytoma/Oligoastrocytoma (grade II&III)	p-value
ATRX alteration	1/39 (2.6%) <sup>a</sup>	41/83 (49%)	<0.0001

<sup>a</sup> Combination of data from this study (n=19) and those from Heaphy *et al* (n=13) and Bettegowda *et al* (n=7).

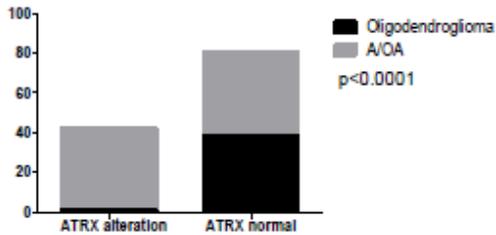
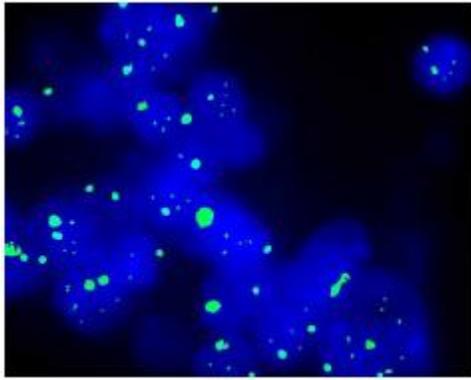
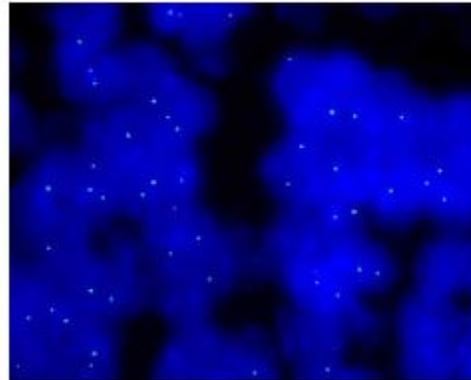


Figure S4.4: Examples of negative and positive alternative lengthening of telomeres in adult gliomas assessed by FISH.



ALT positive



ALT negative

## 4.12. SUPPLEMENTARY TABLES

Table S4.1a: Sample information of the 140 adult gliomas.

Sample ID	Age	Gender	Pathology diagnoses	Detailed information	IDH	TP53	Presence of ATRX alteration	Presence of ALT
Glioma 1	43	M	AII	NA	wild type <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	YES
Glioma 2	26	M	AII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	YES <sup>b</sup>	YES
Glioma 3	31	M	OAI	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	YES <sup>b</sup>	NO
Glioma 4	34	F	AII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	YES <sup>b</sup>	NO
Glioma 5	40	M	OIII	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 6	45	M	OAI	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 7	35	NA	OIII	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 8	51	F	OAI	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NA
Glioma 9	52	F	OIII	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	YES
Glioma 10	37	F	AIII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	YES <sup>b</sup>	NO
Glioma 11	40	M	AII	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	YES
Glioma 12	27	M	AII	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 13	36	M	AII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 14	40	F	OII	NA	R132 <sup>b</sup>	NA	NO <sup>b</sup>	YES
Glioma 15	50	M	OII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 16	43	M	AII	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	YES
Glioma 17	29	NA	AII	NA	wild type <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 18	64	M	OAI	NA	wild type <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 19	38	F	OII	NA	wild type <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 20	37	F	AII	NA	wild type <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 21	46	M	OAI	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 22	37	F	AIII	NA	wild type <sup>b</sup>	mutated <sup>b</sup>	YES <sup>b</sup>	NO
Glioma 23	51	F	AII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 24	36	F	AII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	YES
Glioma 25	38	F	AII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	YES <sup>b</sup>	NO
Glioma 26	38	M	AII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	YES <sup>b</sup>	NO
Glioma 27	34	F	AIII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	YES <sup>c</sup>	YES
Glioma 28	38	M	AIII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	YES <sup>b</sup>	NO
Glioma 29	47	M	AIII	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	YES <sup>b</sup>	NO
Glioma 30	45	F	AII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 31	40	M	OII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 32	59	M	OIII	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 33	50	M	OAI	NA	wild type <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	YES
Glioma 34	49	F	OAI	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	YES <sup>b</sup>	NO
Glioma 35	35	M	AIII	NA	R172H-IDH2 <sup>a</sup>	mutated <sup>b</sup>	YES <sup>b</sup>	YES
Glioma 36	41	M	AII	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	YES <sup>b</sup>	NO
Glioma 37	49	M	OII	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 38	59	F	OIII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 39	28	M	AII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	YES <sup>b</sup>	NO
Glioma 40	46	F	AII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 41	28	F	OII	NA	wild type <sup>b</sup>	NA	NO <sup>b</sup>	NO
Glioma 42	46	F	AII	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 43	27	M	AII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 44	57	M	OIII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	YES
Glioma 45	35	M	AIII	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 46	24	M	AII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	YES <sup>b</sup>	NO
Glioma 47	49	M	AII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	YES <sup>b</sup>	NO
Glioma 48	27	M	AII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	YES <sup>c</sup>	YES
Glioma 49	52	F	AII	NA	wild type <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 50	37	F	AIII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 51	44	M	AII	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 52	38	M	AIII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	YES

Table S4.1a: Sample information of the 140 adult gliomas, continued.

Glioma 53	44	F	AIII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 54	26	M	OAI	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 55	60	M	AIII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	YES <sup>b</sup>	NO
Glioma 56	29	F	OAI	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 57	44	M	AII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 58	65	F	OII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 59	30	M	OII	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 60	51	F	OIII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	YES
Glioma 61	20	M	AII	NA	R132 <sup>b</sup>	wild type <sup>b</sup>	YES <sup>b</sup>	YES
Glioma 62	67	M	OIII	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 63	55	F	OAI	NA	R132 <sup>b</sup>	mutated <sup>b</sup>	NO <sup>b</sup>	NO
Glioma 64	38	F	AIII	de novo	wild type <sup>b</sup>	NA	NO <sup>b</sup>	NA
Glioma 65	67	F	AIII	de novo	wild type <sup>b</sup>	NA	NO <sup>b</sup>	NA
Glioma 66	37	F	AIII	de novo	R132 <sup>b</sup>	NA	YES <sup>b</sup>	NA
Glioma 67	69	F	AIII	de novo	wild type <sup>b</sup>	NA	NO <sup>b</sup>	NA
Glioma 68	77	M	AIII	de novo	wild type <sup>b</sup>	NA	NO <sup>b</sup>	NA
Glioma 69	39	M	AIII	de novo	R132 <sup>b</sup>	NA	YES <sup>b</sup>	NA
Glioma 70	26	M	AIII	de novo	R132 <sup>b</sup>	NA	YES <sup>b</sup>	NA
Glioma 71	43	M	AIII	de novo	wild type <sup>b</sup>	NA	NO <sup>b</sup>	NA
Glioma 72	76	F	AIII	de novo	R132 <sup>b</sup>	NA	YES <sup>b</sup>	NA
Glioma 73	79	F	AIII	de novo	wild type <sup>b</sup>	NA	NO <sup>b</sup>	NA
Glioma 74	47	F	AIII	de novo	wild type <sup>b</sup>	NA	NO <sup>b</sup>	NA
Glioma 75	51	M	OIII	NA	R132 <sup>a</sup>	NA	NO <sup>a</sup>	NA
Glioma 76	34	M	GBM	NA	R132 <sup>a</sup>	NA	YES <sup>a</sup>	NA
Glioma 77	42	F	GBM	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	YES <sup>a</sup>	NA
Glioma 78	46	F	AII	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	YES <sup>a</sup>	NA
Glioma 79	34	M	OAI	recurrent	R132 <sup>a</sup>	NA	YES <sup>a</sup>	NA
Glioma 80	42	F	OAI	NA	R132 <sup>a</sup>	NA	YES <sup>a</sup>	NA
Glioma 81	64	F	OAI	NA	R132 <sup>a</sup>	NA	YES <sup>a</sup>	NA
Glioma 82	73	F	GBM	NA	wild type <sup>a</sup>	NA	NO <sup>a</sup>	NA
Glioma 83	37	M	OAI	NA	R132 <sup>a</sup>	NA	NO <sup>a</sup>	NA
Glioma 84	55	M	OAI	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	YES <sup>a</sup>	NA
Glioma 85	38	M	OIII	NA	R132 <sup>a</sup>	NA	NO <sup>a</sup>	NA
Glioma 86	42	M	GBM	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>a</sup>	NA
Glioma 87	52	M	OAI	NA	R132 <sup>a</sup>	NA	YES <sup>a</sup>	NA
Glioma 88	70	F	GBM	NA	wild type <sup>a</sup>	NA	NO <sup>a</sup>	NA
Glioma 89	54	M	AII	NA	R132 <sup>a</sup>	NA	YES <sup>a</sup>	NA
Glioma 90	47	F	GBM	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	YES <sup>a</sup>	NA
Glioma 91	44	F	OAI	NA	R132 <sup>a</sup>	NA	YES <sup>a</sup>	NA
Glioma 92	72	M	GBM	NA	wild type <sup>a</sup>	NA	NO <sup>a</sup>	NA
Glioma 93	43	M	GBM	recurrent	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>a</sup>	NA
Glioma 94	46	F	GBM	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>a</sup>	NA
Glioma 95	48	F	GBM	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	NO <sup>a</sup>	NA
Glioma 96	46	M	GBM	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>a</sup>	NA
Glioma 97	45	F	GBM	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	NO <sup>a</sup>	NA
Glioma 98	46	M	OII	NA	R132 <sup>a</sup>	NA	NO <sup>a</sup>	NA
Glioma 99	58	M	OAI	NA	R132 <sup>a</sup>	NA	NO <sup>a</sup>	NA
Glioma 100	37	M	OAI	NA	R132 <sup>a</sup>	NA	YES <sup>a</sup>	NA
Glioma 101	42	M	GBM	recurrent	R132 <sup>a</sup>	mutated <sup>a</sup>	YES <sup>a</sup>	NA
Glioma 102	37	M	AIII	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	YES <sup>a</sup>	NA
Glioma 103	60	F	GBM	NA	wild type <sup>a</sup>	NA	NO <sup>a</sup>	NA
Glioma 104	60	F	AIII	recurrent	R132 <sup>a</sup>	mutated <sup>a</sup>	YES <sup>a</sup>	NA
Glioma 105	78	M	GBM	NA	wild type <sup>a</sup>	NA	NO <sup>a</sup>	NA
Glioma 106	51	M	GBM	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	YES <sup>a</sup>	NA

Table S4.1a: Sample information of the 140 adult gliomas, continued.

Glioma 107	48	M	GBM	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>a</sup>	NA
Glioma 108	52	M	GBM	NA	wild type <sup>a</sup>	mutated <sup>a</sup>	NO <sup>a</sup>	NA
Glioma 109	52	F	OAI	NA	R132 <sup>a</sup>	NA	NO <sup>a</sup>	NA
Glioma 110	38	M	GBM	NA	wild type <sup>a</sup>	mutated <sup>a</sup>	YES <sup>a</sup>	NA
Glioma 111	44	M	GBM	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>a</sup>	NA
Glioma 112	67	F	GBM	NA	wild type <sup>a</sup>	NA	NO <sup>a</sup>	NA
Glioma 113	53	M	GBM	recurrent	wild type <sup>a</sup>	NA	NO <sup>a</sup>	NA
Glioma 114	46	F	OAI	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>a</sup>	NA
Glioma 115	74	M	GBM	recurrent	wild type <sup>a</sup>	NA	NO <sup>a</sup>	NA
Glioma 116	36	NA	AIII	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	YES <sup>a</sup>	NA
Glioma 117	44	M	AII	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	YES <sup>a</sup>	NA
Glioma 118	46	F	GBM	NA	R132 <sup>a</sup>	NA	YES <sup>a</sup>	NA
Glioma 119	61	M	AIII	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	YES <sup>b</sup>	NO
Glioma 120	64	M	AIII	NA	wild type <sup>a</sup>	mutated <sup>a</sup>	YES <sup>d</sup>	NO
Glioma 121	35	F	AIII	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	YES <sup>d</sup>	YES
Glioma 122	35	M	AIII	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	NO <sup>b</sup>	NO
Glioma 123	41	F	AIII	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	YES <sup>d</sup>	NO
Glioma 124	37	F	AIII	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>b</sup>	NO
Glioma 125	38	F	AIII	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	YES <sup>d</sup>	NO
Glioma 126	42	F	GBM	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	YES <sup>d</sup>	NO
Glioma 127	27	M	GBM	NA	R132 <sup>a</sup>	wild type <sup>a</sup>	YES <sup>d</sup>	YES
Glioma 128	70	M	GBM	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>b</sup>	NO
Glioma 129	66	F	GBM	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>b</sup>	NO
Glioma 130	78	M	AIII	NA	wild type <sup>a</sup>	mutated <sup>a</sup>	NO <sup>b</sup>	NO
Glioma 131	60	M	GBM	NA	wild type <sup>a</sup>	mutated <sup>a</sup>	NO <sup>b</sup>	NO
Glioma 132	36	M	GBM	NA	R132 <sup>a</sup>	mutated <sup>a</sup>	YES <sup>d</sup>	NO
Glioma 133	63	M	GBM	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>b</sup>	NO
Glioma 134	54	NA	GBM	NA	wild type <sup>a</sup>	NA	NO <sup>b</sup>	NO
Glioma 135	42	F	GBM	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>b</sup>	NO
Glioma 136	58	M	GBM	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>b</sup>	NO
Glioma 137	60	M	GBM	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>b</sup>	NO
Glioma 138	80	M	GBM	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>b</sup>	NO
Glioma 139	58	F	GBM	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>b</sup>	NO
Glioma 140	53	M	GBM	NA	wild type <sup>a</sup>	wild type <sup>a</sup>	NO <sup>b</sup>	NO

<sup>a</sup> indicates samples underwent Sanger sequencing.

<sup>b</sup> indicates samples underwent IHC.

<sup>c</sup> indicates the 2 samples analyzed by both Sanger sequencing and IHC: sequencing identified frameshift mutations, and the 2 samples showed negative ATRX staining by IHC

Table S4.1b: ATRX alterations in the 17 pediatric gliomas and 34 pilocytic astrocytomas analyzed in this study.

	<b>ATRX alteration</b>	<b>Method of screening</b>
<b>pediatric PA</b>	0/34 (0%)	immunohistochemistry
<b>pediatric grade II astrocytoma</b>	0/8 (0%)	sequencing
<b>pediatric grade III astrocytoma</b>	0/7 (0%)	sequencing
<b>pediatric grade III oligodendroglioma</b>	0/2 (0%)	sequencing

Table S4.2: Number of samples with alternative lengthening of telomeres in different subtypes of adult gliomas.

Diagnosis	ALT +	ALT + & ATRX altered	ALT -	ALT - & ATRX altered
All	7	3	19	7
OAI	1	0	6	1
OII	1	0	7	0
AIII	4	3	15	9
OAI	0	0	2	1
OII	3	0	5	0
GBM	1	1	13	2
Total	17	7	67	20

## Chapter 5: DISCUSSION

This thesis work focuses on the understanding of molecular and genetic pathogenesis of pediatric and young adult glioblastoma. We studied the molecular function of YB-1, previously identified to be upregulated in pediatric GBM, in the formation/progression of astrocytoma, using both GBM/normal human astrocytes cell lines and xenograft mouse model. We suggested that the role of YB-1 in astrocytoma genesis depends on its intracellular localization. High nuclear expression of YB-1 led to increased EGFR expression and cell proliferation; whereas cytoplasmic YB-1 limits cell proliferation and tumorigenicity, but promotes cell migration. According to our data, direct targeting of YB-1 for therapy can be challenging, given its seemingly opposite function in the nucleus and cytoplasm, and new therapeutic targets are in need to be discovered.

We then undertook whole exome sequencing study to discover important mutations in pediatric GBM, for a better understanding of its genetic pathogenesis and the identification of new therapeutic targets. We discovered the disruption of the chromatin remodelling H3.3/ATRX-DAXX axis and their association with *TP53* mutations in pediatric GBMs. This is the first report on disease associated recurrent somatic mutations

happening in a histone protein (K27M-H3.3 and G34R/V-H3.3). Since the discovery of *IDH1/2* mutations in adult gliomas, our findings once again attracted the focus of glioblastoma research into the epigenetic regulatory events. Occurring in 44% of our 48 pediatric GBM samples, mutations in H3.3/ATRX-DAXX constitute a major pathogenesis pathway leading to pediatric GBM. The high specificity of H3.3 recurrent somatic mutations in pediatric high grade gliomas compared to adult cases and lower grade tumors help explain the difference of pediatric and adult glioma pathogenesis and provide a good candidate for therapeutic development.

In addition to pediatric GBM, we also investigated the prevalence of ATRX alterations in lower grade and adult gliomas, similar to *H3F3A* mutations. We discovered its prevalence in adult diffuse gliomas (grade II, III and IV) and its specificity to astrocytic tumors carrying *IDH1/2* and *TP53* mutations. We thus revealed an important pathway in the early development and progression of adult gliomas of astrocytic lineage, by the mutations of *IDH1/2* & *TP53* & *ATRX*. Mirroring the *H3F3A* & *ATRX* & *TP53* mutations in pediatric GBM, the mutations of *IDH* & *TP53* & *ATRX* in adult gliomas of astrocytic lineage further support that the molecular

alterations of gliomas are distinct among different age groups and tumor types.

Our results have brought up the importance of epigenetic regulation in gliomagenesis and progression. Further research on the molecular mechanism how mutant H3.3 is involved in pediatric GBM is warranted, such as their effect on histone H3.3 post-translational modification and on overall DNA methylation profile, as well as their binding target of DNA sequences and their collaboration with *ATRX* and *TP53* mutations in driving tumorigenesis.

### 5.1. YB-1: an oncoprotein or anti-oncogenic protein?

YB-1 has long been thought of as an oncoprotein in cancers. Its high expression in a variety of cancers, as well as its prognostic value for tumor aggressiveness and resistance to chemotherapy render it a potential target for therapeutic consideration. However, the function of YB-1 as an oncoprotein seems to relate closely to its localization in the nucleus: nuclear expression of YB-1 is correlated with Her-2 (Human epidermal growth factor receptor 2) expression, MDR1 (multidrug resistance protein 1) synthesis and poor survival in breast cancer, ovarian cancer (Bargou, Jurchott et al. 1997; Kamura, Yahata et al. 1999; Fujita, Ito et al. 2005; Oda, Ohishi et al. 2007; Fujii, Kawahara et al. 2008) and diffuse large B-cell lymphoma (Xu, Zhou et al. 2009); in lung and liver cancers, the nuclear localization of YB-1 has been associated with poor survival (Shibahara, Sugio et al. 2001; Yasen, Kajino et al. 2005; Kashihara, Azuma et al. 2009); increased proliferation was also observed in osteosarcoma and rhabdomyosarcoma with nuclear YB-1 expression (Oda, Sakamoto et al. 1998; Oda, Kohashi et al. 2008). These observations could be explained by nuclear activity of YB-1 in DNA repair, as well as the transcriptional regulation of EGFR, MET, MDR1, PIK3CA,

PDGFB, TP53, p21 and Cyclin A & B1, thus promoting cell proliferation and drug resistance, as discussed in Chapter 1.

It wasn't until recently that YB-1 was discovered to have anti-oncogenic activity, which is associated with its cytoplasmic localization. YB-1 in the cytoplasm binds to the 5'-cap structure of the mRNPs, many of which are involved in cell growth and division, and inhibits their translation. In chicken embryo fibroblasts, the Akt-mediated oncogenic transformation was blocked by YB-1 overexpression through translational inhibition of transcripts necessary for oncogenesis (Bader and Vogt 2008). In breast cancer cells, overexpression of YB-1 in the cytoplasm decreased proliferation rate by translational inhibition of growth-related transcripts (Evdokimova, Tognon et al. 2009). Meanwhile, metastatic ability and epithelial to mesenchymal transition were promoted through translational activation of Snail1 by cytoplasmic YB-1, providing added complexity on the function cytoplasmic YB-1.

Our data described in Chapter 2 indicated a predominantly cytoplasmic localization of ectopically overexpressed YB-1 in glioblastoma and normal human astrocytes cell lines. Orthotopic xenograft mouse model showed

reduced tumorigenicity of cells with overexpressed YB-1 in the cytoplasm, in compliance with its proliferation limiting activity found in our cell lines as well as the breast cancer cell lines (Evdokimova, Tognon et al. 2009).

Silencing of YB-1 using shRNA in GBM cell lines led to a decrease of YB-1 level mainly in the cytoplasm, and was associated with increased cell proliferation. These results demonstrated for the first time in the glioblastoma cells that cytoplasmic YB-1 exhibit anti-oncogenic activity.

Further, we showed the association of increased EGFR expression with nuclear expression of YB-1 and the possible contribution of enriched YB-1 expression in promoting cell proliferation, in compliance with the oncogenic ability of nuclear YB-1 previously described in literature.

The regulation of YB-1 translocation between the nucleus and cytoplasm remains unclear. Although evidence before has indicated that wild type p53 is needed for YB-1 nuclear localization (Zhang, Homer et al. 2003), we observed nuclear YB-1 expression in our pediatric GBM cell line with mutant p53. Other group has shown that phosphorylation of YB-1 by Akt mediates its nuclear localization in breast cancer cells (Sutherland, Kucab et al. 2005). However, in our GBM cell lines, although the Akt pathway is constitutively active, ectopically expressed YB-1 was still mainly located in

the cytoplasm. Therefore, the regulation of YB-1 localization may be cell type specific and requires further investigation.

Therefore, the involvement of YB-1 in glioblastoma genesis is complicated. It can act both as an oncoprotein or anti-oncogenic protein, depending on its nuclear/cytoplasmic localization. When studying this protein as a prognostic marker or developing therapeutic regimen to target it, we have to investigate its sub-cellular localization at the same time. With the regulation of YB-1 shuttling between cytoplasm and nuclear remain controversial and unclear, further research is needed in order to target YB-1 expression in specific intracellular localization.

## 5.2. Whole exome/genome sequencing: current issues.

We employed whole exome sequencing in the study outlined in Chapter 3 and successfully discovered somatic recurrent mutations in *H3F3A*, encoding the histone variant H3.3, as well as *ATRX-DAXX* and *TP53* mutations. Despite of the efficiency of WES technology in the identification of important mutations in several types of cancers, currently unsolved issues still exist.

First of all, WES focuses on the protein-coding regions, which constitutes only 1% of the genome. The large areas of regulatory regions such as promoters and enhancers are not covered, thus mutations important for regulation of gene expression can be missed. Even in whole genome sequencing, discoveries so far have concentrated on the protein coding sequences, due to the difficulty of obtaining functional information in the genetic variants found in the regulatory regions (Downing, Wilson et al. 2012). To overcome this obstacle, WGS is used if increased cost and data size are acceptable, and genomic mutation data should be combined with mRNA expression data from microarray analysis as well as epigenetic information (such as DNA methylation). Genomic sequencing data can also be combined with transcriptome sequencing results (RNA-seq) to

help identify mutations affecting transcription and splicing (Liu, Lee et al. 2012). Integration of these data can provide us with a better knowledge on the effect of mutations of regulatory regions in gene expression, thus a higher opportunity to identify mutations relevant to the disease. With the large amount of data generated from WGS, gene expression and epigenetic evaluation, these integrated studies would require great computing power and improved bioinformatical analysis algorithms.

Second, the tumor samples removed from surgery are usually heterogeneous, consisting of not only contamination from stromal cells, but also vascular cells, such as the case of glioblastoma, as well as possibly different subclones of tumor cells. The content of tumor cells in a sample can be assessed by the H&E (hematoxylin and eosin) staining of the tumor slides. However, it is currently impossible to separate the heterogeneous subclones of tumor cells for sequencing. Recent studies on cancer stem cells argue for the existence of a small population of tumor cells responsible for recurrence after treatment (Bacelli and Trumpp 2012). These cancer stem cells sustain the tumor growth and are highly resistant to radio- and chemotherapy. Due to their very small number in the tumor bulk, the DNAs from these cancer stem cells are under-

represented, and the current depth of whole exome/genome sequencing do not allow the easy identification of mutations in these cells. Therefore, technical improvement on the sequencing coverage at the genomic scale is needed to discover critical pathogenesis-related mutations harbored by a small subpopulation of tumor cells.

In fact, great advancement in sequencing speed is being carried on, allowing for deeper read coverage. The current instruments allow for sequencing of more than 5 human genomes at about 30× coverage each run in about 6 to 11 days. The emerging GridION system from Oxford Nanopore will soon make it possible to sequence the entire human genome in 15 minutes. In addition, the total cost (including labor, reagents, equipments and data analysis) per genome of sequencing is continuously decreasing: from \$10 million at the year of 2007 to less than \$10,000 in 2012 (Wetterstrand). For the purpose of library preparation and sequencing procedures, many institutional based genome centers offer WES to researchers with a cost of less than \$1000 per sample. The rapidly increasing sequencing power and cost-efficiency will make it possible to sequence large cohort of samples with decent coverage in a reasonable amount of time.

The development of bioinformatic tools has allowed for a better understanding of the cancer genomes from WES/WGS data. Not only can they discover genetic mutations such as nucleotide substitutions and small indels (insertion/deletions), they can also identify copy number changes and chromosomal rearrangements in the cancer genomes (Meyerson, Gabriel et al. 2010) (Campbell, Stephens et al. 2008). The identification of copy number changes and chromosome structural variations from next-generation sequencing data is a relatively new field, which blossoms since the last three years. Some computational algorithms such as CNVnator and SeqCBS allows for the analysis of copy number variations; while others like SegSeq and CREST discover chromosomal rearrangement breakpoints. Despite of the numerous bioinformatic tools for the analysis of copy number and structural variants, no standard protocols have been accepted, due to the many statistical challenges and biases faced by the current methods (Teo, Pawitan et al. 2012). Therefore, a great number of efforts from different bioinformatics teams across the world are being put into the development of better algorithms.

WES/WGS offers an unprecedented experience in the discovery of genetic alterations involved in cancer development. It has allowed for the effective identification of driver mutations and new therapeutic targets for cancer treatment. Given its large output and decreasing cost, we could expect shortly for its clinical use for obtaining the genomic background of the patients and thus providing personalized treatment.

### 5.3. Glioma pathogenesis: entering the epigenetic stage.

In this thesis work, we identified the recurrent somatic mutations K27M and G34R/V of the histone variant H3.3 in high grade pediatric glioblastoma and their specificity to pediatric high grade astrocytomas. We also described *ATRX* mutations and/or loss of expression in pediatric GBM as well as adult grade II, III and IV astrocytomas. Combined with the previously identified *IDH1/2* mutations important for adult gliomagenesis, these genetic alterations indicate the key role of epigenetic regulation disruption in the pathogenesis of adult and pediatric gliomas. They have led our attention to the interaction of genetic mutations with epigenetic alterations in driving cancer.

In fact, it is not rare that mutations in epigenetic regulation were recently observed in cancers. For example, somatic mutations of *DNMT3A* (encoding DNA methyltransferase 3A) leading to reduced enzymatic function were found in 20.5% of acute myeloid leukemia (AML) (Yan, Xu et al. 2011). This enzyme catalyzes the DNA methylation reaction, which is an important epigenetic mark often associated with repressed expression of genes. Patients with *DNMT3A* mutations were shown to have worse prognosis, indicating their values as biomarker of this disease.

*IDH1/2* mutations characterize another subgroup of AML, with altered epigenetic state (Rakheja, Konoplev et al. 2012). In 2011, 59% of transitional cell carcinoma, which is a type of bladder cancer, was found to harbor genetic abnormalities in chromatin remodeling genes such as *UTX* (a histone H3 lysine 27 (H3K27) demethylase) and *MLL-MLL3* (a histone methyltransferase) (Gui, Guo et al. 2011). In addition, *UTX* mutations indicated a tumor group with low grades and stages. In renal cell carcinoma, inactivating mutations in *SETD2* (a histone H3K36 methyltransferase), *KDM5C* (a histone H3K4 demethylase) and *UTX* were identified; in diffuse large B-cell lymphomas, recurrent somatic mutations at Tyr641 position of the *EZH2* (a histone methyltransferase trimethylating H3K27) were found in 21.7% of cases (Morin, Johnson et al. 2010), further arguing for the involvement of mutations in chromatin remodeling machinery in cancer development.

With accumulating evidence pointing towards their involvement in cancer, epigenetic regulation happens at several levels: histone modifications (such as methylation, acetylation and phosphorylation) and ATP-dependent chromatin remodeling complexes for remodeling of chromatin structure, as well as DNA methylation to control gene expression pattern.

These epigenetic events are tightly controlled during development and under normal conditions, and it is not surprising to see that disruptions in epigenetic regulation could cause disease such as cancers. For the rest of this section, more discussions will be focused on the epigenetic functions and clinical implications of the mutations in genes important for gliomagenesis: H3.3, IDH and ATRX.

### 5.3.1. H3.3

Recurrent mutations K27M and G34R/V of H3.3 were identified in 9 (18.8%) and 6 (12.5%) of our 48 non-brainstem pediatric GBM samples cohort, respectively. In the case of diffuse intrinsic pontine gliomas (DIPG), which is a type of pediatric high grade astrocytoma occurring in the brainstem, 71% of cases were found to harbor K27M-H3.3 mutation, and no G34V/R-H3.3 or *IDH1/2* mutations were identified (Khuong-Quang, Buczkowicz et al. 2012). K27M-H3.3 mutation is associated with worse survival compared to wild type H3.3 in DIPG patients. Concurrent with our findings, another group also identified somatic K27M-H3 mutations (in their case, either H3.1 or H3.3) in 78% of DIPGs and 22% of non-brainstem pediatric glioblastomas. Similarly, they also found 14% of non-

brainstem pediatric GBMs had G34R-H3.3 mutation. It appears that the type of H3.3 mutations is highly correlated with GBM location: K27M mutation often occurs infratentorially in the thalamus or brainstem, whereas G34R/V mutation, often associated with concurrent ATRX mutations, is mostly found in supratentorial glioblastomas in the cortex. The different localizations of the H3.3 mutations possibly reflect different cells of origin during brain development. This was further proven by the distinct DNA methylation and gene expression pattern of cells harboring K27M or G34R/V mutations, representing the expression profiles of their corresponding regions in developing brain (Sturm, Witt et al. 2012).

The reason H3.3 mutations are mainly seen in the pediatric setting could be explained by the expression and function of H3.3 incorporation during brain development of animal models. H3.3 is the predominant H3 variant incorporated in the histone core during rat brain development (Bosch and Suau 1995). In zebrafish, H3.3 incorporation into the nucleosome is essential for the development of cranial neural crest (Cox, Kim et al. 2012). Therefore, the developing brains in the pediatric cases are probably more sensitive to H3.3 alterations.

In fact, the identification of *H3F3A* mutations in glioblastomas combined with DNA methylation pattern, copy number alterations and expression profile helps to define different epigenetic subgroups of glioblastoma of all age (Sturm, Witt et al. 2012). K27M-H3.3, G34R/V-H3.3 and *IDH* mutations are mutually exclusive in glioblastomas, and each characterizes a subgroup of GBM with distinct DNA methylation pattern (termed “IDH”, “K27” and “G34”). GBMs with *IDH* mutations demonstrated global hypermethylation, in accordance with the previously identified G-CIMP phenotype (Noushmehr, Weisenberger et al. 2010). On the other hand, the G34 subgroup exhibited prominent global DNA hypomethylation. GBMs with none of these three mutations were further classified into the “PDGFRA” subgroup characterized by PDGFRA amplification, the “Classic” subgroup featured by *EGFR* amplification and *CDKN2A* deletion and the “mesenchymal” subgroup with no characteristic copy number aberrations or point mutations. Each of these subgroups displayed its unique DNA methylation pattern. This novel epigenetic subgrouping (IDH, K27, G34, PDGFRA, classic and mesenchymal) of GBM reflects closely and supplements the previous gene-expression-based four molecular subgroups (Verhaak, Hoadley et al. 2010). The proneural expression pattern is represented by the IDH, K27 and PDGFRA epigenetic

subgroups; whereas the expression patterns for classic and mesenchymal epigenetic subgroups correspond to the previous classical and mesenchymal molecular subtypes. The novel epigenetic based subgrouping has demonstrated clinical prognostic values: patients with *IDH* or G34V-H3.3 mutations tend to have better overall survival compared to wild type patients; whereas K27M-H3.3 indicates a slightly worse prognosis. Therefore, this integrated approach of genetic and epigenetic profile has proven to be a better way of subgrouping glioblastomas, with more clinical relevance.

Mutations in both H3.3 and IDH affect the histone methylation state, and it is not surprising that the DNA methylation patterns change accordingly. DNA methylation profiles at CpG islands were shown to correlate closely to histone methylation state at H3K4, H3K9 and H3K27 in mouse embryonic stem cells and neural cells (Meissner, Mikkelsen et al. 2008). Some DNA methyltransferases contain domains recognizing specific histone H3 methylation pattern (unmethylated or methylated); whereas several histone H3 methyltransferases can interact with either methylated DNA through methyl-CpG-binding domain (MBD) or unmethylated DNA via the CpG-binding CXXC domain (Hashimoto, Vertino et al. 2010).

These serve as the molecular explanation for the coupling of DNA and histone methylation pattern. Therefore, the altered histone methylation state caused by H3.3 or IDH mutations probably leads to changes in both chromatin conformation and DNA methylation pattern, thus shaping specific gene expression pattern for the development of cancer.

Although H3.3 mutations seem to alter global DNA methylation patterns, the exact mechanism how H3.3 mutations promote the onset of pediatric glioblastoma is still unclear. Questions remain to be answered: are the mutant and wild type H3.3 incorporated at different loci within the genome? Are other concurrent genetic alterations such as mutations in *TP53* and *ATRX* necessary for H3.3 mutations to drive tumorigenesis? Analysis such as ChIP-seq (chromatin immunoprecipitation-sequencing) can be used to test for the DNA sequences bound to mutant and wild type H3.3; mutant mouse model are being developed to assess the tumorigenicity of these mutations.

Given their prevalence and prognostic value in pediatric glioblastoma, K27M-H3.3 and G34R/V-H3.3 mutations should be tested routinely in pediatric GBM patients. Sequencing can identify the mutations at DNA

level, and antibodies recognizing specifically the two mutations at the protein level are also being developed. These antibodies will allow for the convenient identification of H3.3 mutations, using formalin fixed paraffin embedded tumor slides, and shed light into the development of targeted therapy against mutant H3.3.

### 5.3.2. IDH

*IDH1/2* mutations are more prevalent in adult secondary GBM relative to pediatric cases, and it's found to be common in as early as grade II gliomas. In GBM, they usually occur supratentorially in cerebral cortex, with frontal lobe the most common location (Sturm, Witt et al. 2012). They inhibit histone and DNA methylation through the production of 2-HG, with detailed mechanism discussed in Chapter 1 Introduction. Therefore, similar to H3.3 mutations, *IDH1/2* mutations induce changes in chromatin remodeling and DNA methylation pattern, the so-called G-CIMP, contributing to their mechanism of gliomagenesis. The association of *IDH1/2* mutations with TP53 and ATRX mutations in adult gliomagenesis of astrocytic lineage functionally mirrors the combination of *H3F3A*

mutations with *TP53* and *ATRX* mutations in pediatric GBM. Although the exact genetic alterations are different, it is the epigenetic control that is disrupted in both pediatric GBM and adult glioma.

Other than gliomas, *IDH1/2* mutations are also commonly seen in acute myeloid leukemia (AML). Similar to gliomas, mutations happen at position R132 in IDH1 or R172 in IDH2, but it can also occur at position R140 of IDH2 in AML (Rakheja, Konoplev et al. 2012). *IDH1/2* mutations in AML also induce global DNA hypermethylation through inhibition of DNA demethylation, highlighting a key role of IDH mutation induced disruption of epigenetic regulation in different types of cancers.

Not only are IDH mutations involved in epigenetic regulation, new evidence has pointed out their role in epithelial-mesenchymal transition (EMT). Although IDH mutations are rare in epithelial tumors, such as their occurrence in 3% of prostate cancer (Kang, Kim et al. 2009), epithelial cell lines with endogenous heterozygote IDH mutation have helped to reveal the effect of IDH mutation in promoting EMT (Grassian, Lin et al. 2012). Similar to epigenetic changes, the IDH-induced EMT is also mediated by the oncometabolite 2-HG. The exact mechanism how 2-HG induces EMT

is still unknown, but it seems to depend on the upregulation of ZEB1 transcription factor and decreased mir-200 microRNA expression, both well-characterized EMT drivers. Therefore, IDH mutant-induced 2-HG accumulation has multiple mechanisms in transforming different types of tumor cells.

IDH mutations have shown great diagnostic and prognostic values for glioma patients. IDH mutations separate secondary adult GBM from primary, and are the characteristic feature of diffuse gliomas. They also constitute a group of adult patient with young age of tumor onset and better clinical outcome. Nowadays, molecular classification using genetic markers, such as IDH mutations, TP53 mutations (identifying astrocytic tumors) and 1p19q loss (characterizing oligodendrogliomas), is implementing and gradually replacing the traditional histological and pathological classification of gliomas (Theeler, Yung et al. 2012). These markers provide critical clinical tools to better diagnose different types of gliomas and to indicate patient prognosis.

Due to their important clinical implications, methods for simple, rapid and reliable screening of IDH mutations or 2-HG elevation, other than direct

sequencing, are being studied. High-resolution melting (HRM) has proven to be a reasonable approach. Its underlying principle is to use the different melting temperatures of double strand DNA due to their different nucleotide composition. This process is monitored by fluorescence in real-time, and the separation of melting curve with increasing temperature can differentiate mutant from wild type sequence. This method has successfully and reliably identified *IDH1/2* mutations in AML samples (Ibanez, Such et al. 2012), and is being optimized in our lab for mutation screening in glioma samples. Since the culprit of IDH-driven oncogenesis is the elevated level of 2-HG, its quantity in tumor cells is also an important marker. A recently developed enzymatic assay allows for the rapid and sensitive detection of 2-HG level elevation in paraffin-embedded slides and frozen tumor tissues, as well as serum of AML patients (Balss, Pusch et al. 2012). In this assay, 2-HG is converted enzymatically by 2-hydroxyglutarate dehydrogenase (HGDH) into  $\alpha$ -KG, with the amount of byproduct NADH measured and reflecting the 2-HG substrate quantity. These approaches have enabled large scale routine screening of clinical samples for IDH and 2-HG as biomarkers with low cost.

In order to improve treatment for glioma patients with IDH mutations, targeted therapy specifically inhibit mutant IDH or 2-HG, as well as effective drug delivery of chemotherapeutic agents are being studied (Yen, Bittinger et al. 2010). With the vast amount of efforts from academic institutions and pharmaceutical companies, entrance of IDH and 2-HG inhibitors into clinical trials will not be far in the future.

### **5.3.3. ATRX**

ATRX is an important chromatin remodeling protein, whose mutations and/or loss of expression were frequently seen in our pediatric GBM samples as well as adult grade II, III and IV gliomas. Combined ATRX and TP53 mutations specify an important pathway towards astrocytic lineage of adult diffuse gliomas from IDH mutant precursor cells. Concurrent with our findings, other groups also identified ATRX mutations to be an important genetic alteration in adult diffuse astrocytic gliomas (Jiao, Killela

et al. 2012; Kannan, Inagaki et al. 2012), and IDH mutations combined with CIC/FUBP1 mutations and/or 1p19q codeletion specify the oligodendrocytic lineage of adult diffuse gliomas (Jiao, Killela et al. 2012). These findings provide us with a better understanding of adult glioma pathogenesis, in addition to IDH mutations.

ATRX mutations were associated with alternative lengthening of telomeres (ALT) in our pediatric GBM samples, possibly because ATRX functions to load histone H3.3 into heterochromatic region in the pericentromeres and telomeres. Disrupted ATRX expression or function probably leads to impaired formation of heterochromatin state of the telomere, thus facilitating ALT. In fact, ATRX mutant pediatric GBMs showed prominent hypomethylation at chromosomal ends (Sturm, Witt et al. 2012), which provides a link between the heterochromatin formation perturbation at telomeric region and ALT.

Although simultaneous *ATRX* mutations occur in approximately 40%, 60% and 80% of adult IDH-mutant grade II, III and IV gliomas respectively, we were not able to reach statistical significance for the enrichment of ATRX mutations with increased grade, due to our sample size. It is possible that

ATRX alterations contribute to the progression of gliomas from low grade to high grade. However, there may be other genetic alterations to account for this progression as well as the tumorigenesis of ATRX-wild-type & IDH-mutant astrocytomas. Further studies are necessary to completely decipher the genetic pathways leading to adult glioma onset and progression.

#### 5.4. Conclusion remarks and significance of research

In this thesis work, we revealed the molecular functions of YB-1 protein, previously identified to be overexpressed in pediatric GBM, in astrocytoma pathogenesis, based on its nuclear / cytoplasmic localization (Chapter 2). We also identified novel mutations in epigenetic regulation to drive pediatric glioblastoma genesis: *H3F3A*, *ATRX* and *DAXX* (Chapter 3). In addition, we unraveled the importance of *ATRX* alterations in adult diffuse astrocytic gliomas (Chapter 4). Our results contribute significantly to the better understanding of genetic and molecular alterations in gliomas, the improved subgrouping of glioblastoma, and the development of novel targeted therapies.

Our results have demonstrated the interplay of genetic mutations with disrupted epigenetic regulation as an important mechanism in pediatric and young adult gliomagenesis. We also showed the efficiency of next-generation sequencing and integrated analysis of alterations at genomic, epigenomic and transcriptomic level, in identifying driving force of cancer. Further studies are warranted to understand the molecular mechanisms of driver mutations we identified. Since epigenetic regulation such as chromatin remodeling pathway was proven to be frequently disrupted in

gliomagenesis, further research is needed to discover other important mutations in epigenetic control. These studies will enable us to categorize the patients based on their genetic and molecular background, and to provide personalized medicine and treatment.

## LIST OF REFERENCES

- Astanehe, A., M. R. Finkbeiner, et al. (2009). "The transcriptional induction of PIK3CA in tumor cells is dependent on the oncoprotein Y-box binding protein-1." Oncogene **28**(25): 2406-2418.
- Bacelli, I. and A. Trumpp (2012). "The evolving concept of cancer and metastasis stem cells." J Cell Biol **198**(3): 281-293.
- Bader, A. G., K. A. Felts, et al. (2003). "Y box-binding protein 1 induces resistance to oncogenic transformation by the phosphatidylinositol 3-kinase pathway." Proc Natl Acad Sci U S A **100**(21): 12384-12389.
- Bader, A. G. and P. K. Vogt (2008). "Phosphorylation by Akt disables the anti-oncogenic activity of YB-1." Oncogene **27**(8): 1179-1182.
- Balss, J., S. Pusch, et al. (2012). "Enzymatic assay for quantitative analysis of (D)-2-hydroxyglutarate." Acta Neuropathol.
- Bargou, R. C., K. Jurchott, et al. (1997). "Nuclear localization and increased levels of transcription factor YB-1 in primary human breast cancers are associated with intrinsic MDR1 gene expression." Nat Med **3**(4): 447-450.
- Barrow, J., M. Adamowicz-Brice, 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**(2): 212-222.
- Basaki, Y., F. Hosoi, et al. (2007). "Akt-dependent nuclear localization of Y-box-binding protein 1 in acquisition of malignant characteristics by human ovarian cancer cells." Oncogene **26**(19): 2736-2746.
- Bassett, A. R., S. E. Cooper, et al. (2008). "The Chromatin Remodelling Factor dATRX Is Involved in Heterochromatin Formation." Plos One **3**(5).
- Batra, S. K., S. Castelino-Prabhu, et al. (1995). "Epidermal growth factor ligand-independent, unregulated, cell-transforming potential of a naturally occurring human mutant EGFRvIII gene." Cell Growth Differ **6**(10): 1251-1259.

- Baumann, C., M. M. Viveiros, et al. (2010). "Loss of maternal ATRX results in centromere instability and aneuploidy in the mammalian oocyte and pre-implantation embryo." *Plos Genetics* **6**(9).
- Bax, D. A., N. Gaspar, 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**(18): 5753-5761.
- Bax, D. A., A. Mackay, et al. (2010). "A distinct spectrum of copy number aberrations in pediatric high-grade gliomas." *Clin Cancer Res* **16**(13): 3368-3377.
- Bergmann, S., B. Royer-Pokora, et al. (2005). "YB-1 provokes breast cancer through the induction of chromosomal instability that emerges from mitotic failure and centrosome amplification." *Cancer Res* **65**(10): 4078-4087.
- Bernstein, B. E., T. S. Mikkelsen, et al. (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." *Cell* **125**(2): 315-326.
- Bettegowda, C., N. Agrawal, et al. (2011). "Mutations in CIC and FUBP1 contribute to human oligodendroglioma." *Science* **333**(6048): 1453-1455.
- Bosch, A. and P. Suau (1995). "Changes in core histone variant composition in differentiating neurons: the roles of differential turnover and synthesis rates." *Eur J Cell Biol* **68**(3): 220-225.
- Bourne, T. D. and D. Schiff (2010). "Update on molecular findings, management and outcome in low-grade gliomas." *Nat Rev Neurol* **6**(12): 695-701.
- Bouvet, P., K. Matsumoto, et al. (1995). "Sequence-specific RNA recognition by the Xenopus Y-box proteins. An essential role for the cold shock domain." *J Biol Chem* **270**(47): 28297-28303.
- Broniscer, A. (2006). "Past, present, and future strategies in the treatment of high-grade glioma in children." *Cancer Invest* **24**(1): 77-81.

- Cadioux, B., T. T. Ching, et al. (2006). "Genome-wide hypomethylation in human glioblastomas associated with specific copy number alteration, methylenetetrahydrofolate reductase allele status, and increased proliferation." Cancer Res **66**(17): 8469-8476.
- Cage, T. A., S. Mueller, et al. (2012). "High-grade gliomas in children." Neurosurg Clin N Am **23**(3): 515-523.
- Campbell, P. J., P. J. Stephens, et al. (2008). "Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing." Nat Genet **40**(6): 722-729.
- Capper, D., S. Weissert, et al. (2010). "Characterization of R132H mutation-specific IDH1 antibody binding in brain tumors." Brain Pathol **20**(1): 245-254.
- Cervera, A. M., J. P. Bayley, et al. (2009). "Inhibition of succinate dehydrogenase dysregulates histone modification in mammalian cells." Mol Cancer **8**: 89.
- Chatterjee, M., C. Rancso, et al. (2008). "The Y-box binding protein YB-1 is associated with progressive disease and mediates survival and drug resistance in multiple myeloma." Blood **111**(7): 3714-3722.
- Cheung, N. K., J. Zhang, et al. (2012). "Association of age at diagnosis and genetic mutations in patients with neuroblastoma." JAMA **307**(10): 1062-1071.
- Chi, P., C. D. Allis, et al. (2010). "Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers." Nat Rev Cancer **10**(7): 457-469.
- Choi, M., U. I. Scholl, et al. (2009). "Genetic diagnosis by whole exome capture and massively parallel DNA sequencing." Proc Natl Acad Sci U S A **106**(45): 19096-19101.
- Chowdhury, R., K. K. Yeoh, et al. "The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases." EMBO Rep **12**(5): 463-469.

- Cohen, K. J., I. F. Pollack, et al. (2011). "Temozolomide in the treatment of high-grade gliomas in children: a report from the Children's Oncology Group." Neuro Oncol **13**(3): 317-323.
- Coward, P., H. G. Wada, et al. (1998). "Controlling signaling with a specifically designed Gi-coupled receptor." Proc Natl Acad Sci U S A **95**(1): 352-357.
- Cox, S. G., H. Kim, et al. (2012). "An essential role of variant histone H3.3 for ectomesenchyme potential of the cranial neural crest." PLoS Genet **8**(9): e1002938.
- Dang, L., D. W. White, et al. (2010). "Cancer-associated IDH1 mutations produce 2-hydroxyglutarate." Nature **465**(7300): 966.
- Das, K. K., A. Mehrotra, et al. (2012). "Pediatric glioblastoma: clinico-radiological profile and factors affecting the outcome." Childs Nerv Syst.
- Das, S., R. Chattopadhyay, et al. (2007). "Stimulation of NEIL2-mediated oxidized base excision repair via YB-1 interaction during oxidative stress." J Biol Chem **282**(39): 28474-28484.
- De La Fuente, R., M. M. Viveiros, et al. (2004). "ATRX, a member of the SNF2 family of helicase/ATPases, is required for chromosome alignment and meiotic spindle organization in metaphase II stage mouse oocytes." Dev Biol **272**(1): 1-14.
- Dhayalan, A., R. Tamas, et al. (2011). "The ATRX-ADD domain binds to H3 tail peptides and reads the combined methylation state of K4 and K9." Hum Mol Genet **20**(11): 2195-2203.
- Di Costanzo, A., A. Troiano, et al. (2012). "The p63 protein isoform DeltaNp63alpha modulates Y-box binding protein 1 in its subcellular distribution and regulation of cell survival and motility genes." J Biol Chem **287**(36): 30170-30180.
- Didier, D. K., J. Schiffenbauer, et al. (1988). "Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class II Y box." Proc Natl Acad Sci U S A **85**(19): 7322-7326.

- Dolecek, T. A., J. M. Propp, et al. (2012). "CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2005-2009." Neuro Oncol **14 Suppl 5**: v1-v49.
- Downing, J. R., R. K. Wilson, et al. (2012). "The Pediatric Cancer Genome Project." Nat Genet **44**(6): 619-622.
- Dubuc, A. M., S. Mack, et al. (2012). "The epigenetics of brain tumors." Methods Mol Biol **863**: 139-153.
- Edmunds, J. W., L. C. Mahadevan, et al. (2008). "Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation." EMBO J **27**(2): 406-420.
- Ekstrand, A. J., N. Sugawa, et al. (1992). "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 **89**(10): 4309-4313.
- Eliseeva, I. A., E. R. Kim, et al. (2011). "Y-box-binding protein 1 (YB-1) and its functions." Biochemistry (Mosc) **76**(13): 1402-1433.
- Elsasser, S. J., C. D. Allis, et al. (2011). "Cancer. New epigenetic drivers of cancers." Science **331**(6021): 1145-1146.
- Emelyanov, A. V., A. Y. Konev, et al. (2010). "Protein Complex of Drosophila ATRX/XNP and HP1a Is Required for the Formation of Pericentric Beta-heterochromatin in Vivo." Journal of Biological Chemistry **285**(20): 15027-15037.
- En-Nia, A., E. Yilmaz, et al. (2005). "Transcription factor YB-1 mediates DNA polymerase alpha gene expression." J Biol Chem **280**(9): 7702-7711.
- Evdokimova, V., L. P. Ovchinnikov, et al. (2006). "Y-box binding protein 1: providing a new angle on translational regulation." Cell Cycle **5**(11): 1143-1147.
- Evdokimova, V., P. Ruzanov, et al. (2006). "Akt-mediated YB-1 phosphorylation activates translation of silent mRNA species." Mol Cell Biol **26**(1): 277-292.

- Evdokimova, V., P. Ruzanov, et al. (2001). "The major mRNA-associated protein YB-1 is a potent 5' cap-dependent mRNA stabilizer." EMBO J **20**(19): 5491-5502.
- Evdokimova, V., C. Tognon, et al. (2009). "Translational activation of snail1 and other developmentally regulated transcription factors by YB-1 promotes an epithelial-mesenchymal transition." Cancer Cell **15**(5): 402-415.
- Evdokimova, V. M., E. A. Kovrigina, et al. (1998). "The major core protein of messenger ribonucleoprotein particles (p50) promotes initiation of protein biosynthesis in vitro." J Biol Chem **273**(6): 3574-3581.
- Evdokimova, V. M. and L. P. Ovchinnikov (1999). "Translational regulation by Y-box transcription factor: involvement of the major mRNA-associated protein, p50." Int J Biochem Cell Biol **31**(1): 139-149.
- Fanelli, M., S. Caprodossi, et al. (2008). "Loss of pericentromeric DNA methylation pattern in human glioblastoma is associated with altered DNA methyltransferases expression and involves the stem cell compartment." Oncogene **27**(3): 358-365.
- Faury, D., A. Nantel, et al. (2007). "Molecular profiling identifies prognostic subgroups of pediatric glioblastoma and shows increased YB-1 expression in tumors." J Clin Oncol **25**(10): 1196-1208.
- Figueroa, M. E., O. Abdel-Wahab, et al. (2010). "Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation." Cancer Cell **18**(6): 553-567.
- Finkbeiner, M. R., A. Astanehe, et al. (2009). "Profiling YB-1 target genes uncovers a new mechanism for MET receptor regulation in normal and malignant human mammary cells." Oncogene **28**(11): 1421-1431.
- Flaherty, K. T., I. Puzanov, et al. (2010). "Inhibition of mutated, activated BRAF in metastatic melanoma." N Engl J Med **363**(9): 809-819.
- Forsheaw, T., R. G. Tatevossian, et al. (2009). "Activation of the ERK/MAPK pathway: a signature genetic defect in posterior fossa pilocytic astrocytomas." J Pathol **218**(2): 172-181.

- Fujii, T., A. Kawahara, et al. (2008). "Expression of HER2 and estrogen receptor alpha depends upon nuclear localization of Y-box binding protein-1 in human breast cancers." Cancer Res **68**(5): 1504-1512.
- Fujisawa, H., M. Kurrer, et al. (1999). "Acquisition of the glioblastoma phenotype during astrocytoma progression is associated with loss of heterozygosity on 10q25-qter." Am J Pathol **155**(2): 387-394.
- Fujisawa, H., R. M. Reis, et al. (2000). "Loss of heterozygosity on chromosome 10 is more extensive in primary (de novo) than in secondary glioblastomas." Lab Invest **80**(1): 65-72.
- Fujita, T., K. Ito, et al. (2005). "Increased nuclear localization of transcription factor Y-box binding protein 1 accompanied by up-regulation of P-glycoprotein in breast cancer pretreated with paclitaxel." Clin Cancer Res **11**(24 Pt 1): 8837-8844.
- Fullgrabe, J., E. Kavanagh, et al. (2011). "Histone onco-modifications." Oncogene **30**(31): 3391-3403.
- Funakoshi, T., S. Kobayashi, et al. (2003). "Isolation and characterization of brain Y-box protein: developmentally regulated expression, polyribosomal association and dendritic localization." Brain Res Mol Brain Res **118**(1-2): 1-9.
- Gaudreault, I., D. Guay, et al. (2004). "YB-1 promotes strand separation in vitro of duplex DNA containing either mispaired bases or cisplatin modifications, exhibits endonucleolytic activities and binds several DNA repair proteins." Nucleic Acids Res **32**(1): 316-327.
- Gibbons, R. (2006). "Alpha thalassaemia-mental retardation, X linked." Orphanet J Rare Dis **1**: 15.
- Gibbons, R. J., A. Pellagatti, et al. (2003). "Identification of acquired somatic mutations in the gene encoding chromatin-remodeling factor ATRX in the alpha-thalassemia myelodysplasia syndrome (ATMDS)." Nat Genet **34**(4): 446-449.

- Gibbons, R. J., D. J. Picketts, et al. (1995). "Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome)." Cell **80**(6): 837-845.
- Gibbons, R. J., T. Wada, et al. (2008). "Mutations in the chromatin-associated protein ATRX." Hum Mutat **29**(6): 796-802.
- Goldberg, A. D., L. A. Banaszynski, et al. (2010). "Distinct factors control histone variant H3.3 localization at specific genomic regions." Cell **140**(5): 678-691.
- Grant, C. E. and R. G. Deeley (1993). "Cloning and characterization of chicken YB-1: regulation of expression in the liver." Mol Cell Biol **13**(7): 4186-4196.
- Grassian, A. R., F. Lin, et al. (2012). "Isocitrate Dehydrogenase (IDH) Mutations Promote a Reversible ZEB1/mir-200-Dependent Epithelial-Mesenchymal (EMT) Transition." J Biol Chem.
- Gravendeel, L. A., M. C. Kouwenhoven, et al. (2009). "Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology." Cancer Res **69**(23): 9065-9072.
- Guay, D., A. A. Evoy, et al. (2008). "The strand separation and nuclease activities associated with YB-1 are dispensable for cisplatin resistance but overexpression of YB-1 in MCF7 and MDA-MB-231 breast tumor cells generates several chemoresistance signatures." Int J Biochem Cell Biol **40**(11): 2492-2507.
- Gui, Y., G. Guo, et al. (2011). "Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder." Nat Genet **43**(9): 875-878.
- Haas, J., H. A. Katus, et al. (2011). "Next-generation sequencing entering the clinical arena." Mol Cell Probes **25**(5-6): 206-211.
- Hakin-Smith, V., D. A. Jellinek, et al. (2003). "Alternative lengthening of telomeres and survival in patients with glioblastoma multiforme." Lancet **361**(9360): 836-838.

- Haque, T., D. Faury, et al. (2007). "Gene expression profiling from formalin-fixed paraffin-embedded tumors of pediatric glioblastoma." Clin Cancer Res **13**(21): 6284-6292.
- Hashimoto, H., P. M. Vertino, et al. (2010). "Molecular coupling of DNA methylation and histone methylation." Epigenomics **2**(5): 657-669.
- Heaphy, C. M., R. F. de Wilde, et al. (2011). "Altered telomeres in tumors with ATRX and DAXX mutations." Science **333**(6041): 425.
- Hegi, M. E., A. C. Diserens, et al. (2005). "MGMT gene silencing and benefit from temozolomide in glioblastoma." N Engl J Med **352**(10): 997-1003.
- Homer, C., D. A. Knight, et al. (2005). "Y-box factor YB1 controls p53 apoptotic function." Oncogene **24**(56): 8314-8325.
- Huang, H. S., M. Nagane, et al. (1997). "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 **272**(5): 2927-2935.
- Ibanez, M., E. Such, et al. (2012). "Rapid Screening of ASXL1, IDH1, IDH2, and c-CBL Mutations in de Novo Acute Myeloid Leukemia by High-Resolution Melting." J Mol Diagn **14**(6): 594-601.
- Ise, T., G. Nagatani, et al. (1999). "Transcription factor Y-box binding protein 1 binds preferentially to cisplatin-modified DNA and interacts with proliferating cell nuclear antigen." Cancer Res **59**(2): 342-346.
- Ito, K., K. Tsutsumi, et al. (1994). "A novel growth-inducible gene that encodes a protein with a conserved cold-shock domain." Nucleic Acids Res **22**(11): 2036-2041.
- Ito, Y., H. Yoshida, et al. (2003). "Y-box binding protein expression in thyroid neoplasms: its linkage with anaplastic transformation." Pathol Int **53**(7): 429-433.
- Iwase, S., B. Xiang, et al. (2011). "ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome." Nat Struct Mol Biol **18**(7): 769-776.

- Izumi, H., T. Imamura, et al. (2001). "Y box-binding protein-1 binds preferentially to single-stranded nucleic acids and exhibits 3'-->5' exonuclease activity." Nucleic Acids Res **29**(5): 1200-1207.
- Jabado, N., F. Le Deist, et al. (1994). "Interaction of HIV gp120 and anti-CD4 antibodies with the CD4 molecule on human CD4+ T cells inhibits the binding activity of NF-AT, NF-kappa B and AP-1, three nuclear factors regulating interleukin-2 gene enhancer activity." Eur J Immunol **24**(11): 2646-2652.
- Jacob, K., S. Albrecht, et al. (2009). "Duplication of 7q34 is specific to juvenile pilocytic astrocytomas and a hallmark of cerebellar and optic pathway tumours." Br J Cancer **101**(4): 722-733.
- Janzarik, W. G., C. P. Kratz, et al. (2007). "Further evidence for a somatic KRAS mutation in a pilocytic astrocytoma." Neuropediatrics **38**(2): 61-63.
- Jiao, Y., P. J. Killela, et al. (2012). "Frequent ATRX, CIC, and FUBP1 mutations refine the classification of malignant gliomas." Oncotarget **3**(7): 709-722.
- Jiao, Y., C. Shi, et al. (2011). "DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors." Science **331**(6021): 1199-1203.
- Jones, C., L. Perryman, et al. (2012). "Paediatric and adult malignant glioma: close relatives or distant cousins?" Nat Rev Clin Oncol **9**(7): 400-413.
- Jones, D. T., J. Gronych, et al. (2011). "MAPK pathway activation in pilocytic astrocytoma." Cell Mol Life Sci **69**(11): 1799-1811.
- Jones, D. T., J. Gronych, et al. (2012). "MAPK pathway activation in pilocytic astrocytoma." Cell Mol Life Sci **69**(11): 1799-1811.
- Jones, D. T., S. Kocialkowski, et al. (2009). "Oncogenic RAF1 rearrangement and a novel BRAF mutation as alternatives to KIAA1549:BRAF fusion in activating the MAPK pathway in pilocytic astrocytoma." Oncogene **28**(20): 2119-2123.
- Jones, P. A. and S. B. Baylin (2007). "The epigenomics of cancer." Cell **128**(4): 683-692.

- Jones, S., R. H. Hruban, et al. (2009). "Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene." Science **324**(5924): 217.
- Jurchott, K., S. Bergmann, et al. (2003). "YB-1 as a cell cycle-regulated transcription factor facilitating cyclin A and cyclin B1 gene expression." J Biol Chem **278**(30): 27988-27996.
- Kamnasaran, D., B. Qian, et al. (2007). "GATA6 is an astrocytoma tumor suppressor gene identified by gene trapping of mouse glioma model." Proc Natl Acad Sci U S A **104**(19): 8053-8058.
- Kamura, T., H. Yahata, et al. (1999). "Is nuclear expression of Y box-binding protein-1 a new prognostic factor in ovarian serous adenocarcinoma?" Cancer **85**(11): 2450-2454.
- Kang, M. R., M. S. Kim, et al. (2009). "Mutational analysis of IDH1 codon 132 in glioblastomas and other common cancers." Int J Cancer **125**(2): 353-355.
- Kannan, K., A. Inagaki, et al. (2012). "Whole-exome sequencing identifies ATRX mutation as a key molecular determinant in lower-grade glioma." Oncotarget **3**(10): 1194-1203.
- Kashihara, M., K. Azuma, et al. (2009). "Nuclear Y-box binding protein-1, a predictive marker of prognosis, is correlated with expression of HER2/ErbB2 and HER3/ErbB3 in non-small cell lung cancer." J Thorac Oncol **4**(9): 1066-1074.
- Khuong-Quang, D. A., P. Buczkowicz, et al. (2012). "K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas." Acta Neuropathol **124**(3): 439-447.
- Kieran, M. W., D. Walker, et al. (2010). "Brain tumors: from childhood through adolescence into adulthood." J Clin Oncol **28**(32): 4783-4789.
- Kim, K., J. Choi, et al. (2008). "Isolation and characterization of a novel H1.2 complex that acts as a repressor of p53-mediated transcription." J Biol Chem **283**(14): 9113-9126.
- Kim, W. and L. M. Liau (2012). "IDH mutations in human glioma." Neurosurg Clin N Am **23**(3): 471-480.

- Ko, M., Y. Huang, et al. (2010). "Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2." Nature **468**(7325): 839-843.
- Koboldt, D. C., L. Ding, et al. (2010). "Challenges of sequencing human genomes." Brief Bioinform **11**(5): 484-498.
- Kohno, K., H. Izumi, et al. (2003). "The pleiotropic functions of the Y-box-binding protein, YB-1." Bioessays **25**(7): 691-698.
- Koike, K., T. Uchiumi, et al. (1997). "Nuclear translocation of the Y-box binding protein by ultraviolet irradiation." FEBS Lett **417**(3): 390-394.
- Kolasinska-Zwierz, P., T. Down, et al. (2009). "Differential chromatin marking of introns and expressed exons by H3K36me3." Nat Genet **41**(3): 376-381.
- Kramm, C. M., S. Butenhoff, et al. (2011). "Thalamic high-grade gliomas in children: a distinct clinical subset?" Neuro Oncol **13**(6): 680-689.
- Ku, C. S., D. N. Cooper, et al. (2012). "Exome sequencing: dual role as a discovery and diagnostic tool." Ann Neurol **71**(1): 5-14.
- Laperriere, N., L. Zuraw, et al. (2002). "Radiotherapy for newly diagnosed malignant glioma in adults: a systematic review." Radiother Oncol **64**(3): 259-273.
- Lasham, A., S. Moloney, et al. (2003). "The Y-box-binding protein, YB1, is a potential negative regulator of the p53 tumor suppressor." J Biol Chem **278**(37): 35516-35523.
- Lee, J. C., I. Vivanco, et al. (2006). "Epidermal growth factor receptor activation in glioblastoma through novel missense mutations in the extracellular domain." PLoS Med **3**(12): e485.
- Lewis, P. W., S. J. Elsaesser, et al. (2010). "Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres." Proc Natl Acad Sci U S A **107**(32): 14075-14080.
- Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics **25**(14): 1754-1760.

- Li, H., B. Handsaker, et al. (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.
- Liu, J., W. Lee, et al. (2012). "Genome and transcriptome sequencing of lung cancers reveal diverse mutational and splicing events." Genome Res.
- Liu, X. Y., N. Gerges, et al. (2012). "Frequent ATRX mutations and loss of expression in adult diffuse astrocytic tumors carrying IDH1/IDH2 and TP53 mutations." Acta Neuropathol **124**(5): 615-625.
- Louis, D. N., H. Ohgaki, et al. (2007). "The 2007 WHO classification of tumours of the central nervous system." Acta Neuropathol **114**(2): 97-109.
- Louis, D. N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K. (2007). WHO Classification of Tumours of the Central Nervous System. Lyon, IARC.
- Lu, Z. H., J. T. Books, et al. (2005). "YB-1 is important for late-stage embryonic development, optimal cellular stress responses, and the prevention of premature senescence." Mol Cell Biol **25**(11): 4625-4637.
- Majewski, J., J. Schwartzentruber, et al. (2011). "What can exome sequencing do for you?" J Med Genet **48**(9): 580-589.
- Manival, X., L. Ghisolfi-Nieto, et al. (2001). "RNA-binding strategies common to cold-shock domain- and RNA recognition motif-containing proteins." Nucleic Acids Res **29**(11): 2223-2233.
- Matsumoto, K., F. Meric, et al. (1996). "Translational repression dependent on the interaction of the Xenopus Y-box protein FRGY2 with mRNA. Role of the cold shock domain, tail domain, and selective RNA sequence recognition." J Biol Chem **271**(37): 22706-22712.
- Meissner, A., T. S. Mikkelsen, et al. (2008). "Genome-scale DNA methylation maps of pluripotent and differentiated cells." Nature **454**(7205): 766-770.
- Mertens, P. R., K. Steinmann, et al. (2002). "Combinatorial interactions of p53, activating protein-2, and YB-1 with a single enhancer element regulate gelatinase A expression in neoplastic cells." J Biol Chem **277**(28): 24875-24882.

- Meyerson, M., S. Gabriel, et al. (2010). "Advances in understanding cancer genomes through second-generation sequencing." Nat Rev Genet **11**(10): 685-696.
- Miwa, A., T. Higuchi, et al. (2006). "Expression and polysome association of YB-1 in various tissues at different stages in the lifespan of mice." Biochim Biophys Acta **1760**(11): 1675-1681.
- Molenaar, J. J., J. Koster, et al. (2012). "Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes." Nature **483**(7391): 589-593.
- Morin, R. D., N. A. Johnson, et al. (2010). "Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin." Nat Genet **42**(2): 181-185.
- Nagarajan, R. P. and J. F. Costello (2009). "Epigenetic mechanisms in glioblastoma multiforme." Semin Cancer Biol **19**(3): 188-197.
- Nakamura, K., T. Oshima, et al. (2011). "Sequence-specific error profile of Illumina sequencers." Nucleic Acids Res **39**(13): e90.
- Nakamura, M., E. Ishida, et al. (2005). "Frequent LOH on 22q12.3 and TIMP-3 inactivation occur in the progression to secondary glioblastomas." Lab Invest **85**(2): 165-175.
- Nekrasov, M. P., M. P. Ivshina, et al. (2003). "The mRNA-binding protein YB-1 (p50) prevents association of the eukaryotic initiation factor eIF4G with mRNA and inhibits protein synthesis at the initiation stage." J Biol Chem **278**(16): 13936-13943.
- Ng, S. B., E. H. Turner, et al. (2009). "Targeted capture and massively parallel sequencing of 12 human exomes." Nature **461**(7261): 272-276.
- Nishikawa, R., X. D. Ji, et al. (1994). "A mutant epidermal growth factor receptor common in human glioma confers enhanced tumorigenicity." Proc Natl Acad Sci U S A **91**(16): 7727-7731.

- Noushmehr, H., D. J. Weisenberger, et al. (2010). "Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma." Cancer Cell **17**(5): 510-522.
- Oda, Y., K. Kohashi, et al. (2008). "Different expression profiles of Y-box-binding protein-1 and multidrug resistance-associated proteins between alveolar and embryonal rhabdomyosarcoma." Cancer Sci **99**(4): 726-732.
- Oda, Y., Y. Ohishi, et al. (2007). "Prognostic implications of the nuclear localization of Y-box-binding protein-1 and CXCR4 expression in ovarian cancer: their correlation with activated Akt, LRP/MVP and P-glycoprotein expression." Cancer Sci **98**(7): 1020-1026.
- Oda, Y., A. Sakamoto, et al. (1998). "Nuclear expression of YB-1 protein correlates with P-glycoprotein expression in human osteosarcoma." Clin Cancer Res **4**(9): 2273-2277.
- Ohga, T., K. Koike, et al. (1996). "Role of the human Y box-binding protein YB-1 in cellular sensitivity to the DNA-damaging agents cisplatin, mitomycin C, and ultraviolet light." Cancer Res **56**(18): 4224-4228.
- Ohga, T., T. Uchiumi, et al. (1998). "Direct involvement of the Y-box binding protein YB-1 in genotoxic stress-induced activation of the human multidrug resistance 1 gene." J Biol Chem **273**(11): 5997-6000.
- Ohgaki, H., P. Dessen, et al. (2004). "Genetic pathways to glioblastoma: a population-based study." Cancer Res **64**(19): 6892-6899.
- Ohgaki, H. and P. Kleihues (2005). "Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas." J Neuropathol Exp Neurol **64**(6): 479-489.
- Ohgaki, H. and P. Kleihues (2007). "Genetic pathways to primary and secondary glioblastoma." Am J Pathol **170**(5): 1445-1453.
- Ohgaki, H. and P. Kleihues (2011). "Genetic profile of astrocytic and oligodendroglial gliomas." Brain Tumor Pathol **28**(3): 177-183.

- Okamoto, T., H. Izumi, et al. (2000). "Direct interaction of p53 with the Y-box binding protein, YB-1: a mechanism for regulation of human gene expression." Oncogene **19**(54): 6194-6202.
- Parsons, D. W., S. Jones, et al. (2008). "An integrated genomic analysis of human glioblastoma multiforme." Science **321**(5897): 1807-1812.
- Parsons, D. W., M. Li, et al. (2011). "The genetic landscape of the childhood cancer medulloblastoma." Science **331**(6016): 435-439.
- Paugh, B. S., C. Qu, et al. (2010). "Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease." J Clin Oncol **28**(18): 3061-3068.
- Peiffer, D. A., J. M. Le, et al. (2006). "High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping." Genome Res **16**(9): 1136-1148.
- Pfister, S., W. G. Janzarik, et al. (2008). "BRAF gene duplication constitutes a mechanism of MAPK pathway activation in low-grade astrocytomas." J Clin Invest **118**(5): 1739-1749.
- Phillips, H. S., S. Kharbanda, 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**(3): 157-173.
- Pisarev, A. V., M. A. Skabkin, et al. (2002). "Positive and negative effects of the major mammalian messenger ribonucleoprotein p50 on binding of 40 S ribosomal subunits to the initiation codon of beta-globin mRNA." J Biol Chem **277**(18): 15445-15451.
- Pollack, I. F. (1999). "Pediatric brain tumors." Semin Surg Oncol **16**(2): 73-90.
- Pollack, I. F., R. L. Hamilton, et al. (2010). "IDH1 mutations are common in malignant gliomas arising in adolescents: a report from the Children's Oncology Group." Childs Nerv Syst **27**(1): 87-94.
- Qu, H. Q., K. Jacob, et al. (2010). "Genome-wide profiling using single-nucleotide polymorphism arrays identifies novel chromosomal imbalances in pediatric glioblastomas." Neuro Oncol **12**(2): 153-163.

- Rakheja, D., S. Konoplev, et al. (2012). "IDH mutations in acute myeloid leukemia." Hum Pathol **43**(10): 1541-1551.
- Rineer, J., D. Schreiber, et al. (2010). "Characterization and outcomes of infratentorial malignant glioma: a population-based study using the Surveillance Epidemiology and End-Results database." Radiother Oncol **95**(3): 321-326.
- Rodriguez, F. J., C. Giannini, et al. (2008). "Gene expression profiling of NF-1-associated and sporadic pilocytic astrocytoma identifies aldehyde dehydrogenase 1 family member L1 (ALDH1L1) as an underexpressed candidate biomarker in aggressive subtypes." J Neuropathol Exp Neurol **67**(12): 1194-1204.
- Sabath, D. E., P. L. Podolin, et al. (1990). "cDNA cloning and characterization of interleukin 2-induced genes in a cloned T helper lymphocyte." J Biol Chem **265**(21): 12671-12678.
- Sakura, H., T. Maekawa, et al. (1988). "Two human genes isolated by a novel method encode DNA-binding proteins containing a common region of homology." Gene **73**(2): 499-507.
- Sanger, F., G. M. Air, et al. (1977). "Nucleotide sequence of bacteriophage phi X174 DNA." Nature **265**(5596): 687-695.
- Sanger, F., S. Nicklen, et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-5467.
- Sarkies, P. and J. E. Sale (2012). "Cellular epigenetic stability and cancer." Trends Genet **28**(3): 118-127.
- Schiffman, J. D., J. G. Hodgson, et al. (2010). "Oncogenic BRAF mutation with CDKN2A inactivation is characteristic of a subset of pediatric malignant astrocytomas." Cancer Res **70**(2): 512-519.
- Schneiderman, J. I., A. Sakai, et al. (2009). "The XNP remodeler targets dynamic chromatin in *Drosophila*." Proceedings of the National Academy of Sciences of the United States of America **106**(34): 14472-14477.

- Schuster, S. C. (2008). "Next-generation sequencing transforms today's biology." Nat Methods **5**(1): 16-18.
- Schwartzentruber, J., A. Korshunov, et al. (2012). "Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma." Nature **482**(7384): 226-231.
- Sengupta, S., A. K. Mantha, et al. (2011). "Human AP endonuclease (APE1/Ref-1) and its acetylation regulate YB-1-p300 recruitment and RNA polymerase II loading in the drug-induced activation of multidrug resistance gene MDR1." Oncogene **30**(4): 482-493.
- Shibahara, K., K. Sugio, et al. (2001). "Nuclear expression of the Y-box binding protein, YB-1, as a novel marker of disease progression in non-small cell lung cancer." Clin Cancer Res **7**(10): 3151-3155.
- Shibahara, K., T. Uchiumi, et al. (2004). "Targeted disruption of one allele of the Y-box binding protein-1 (YB-1) gene in mouse embryonic stem cells and increased sensitivity to cisplatin and mitomycin C." Cancer Sci **95**(4): 348-353.
- Shibao, K., H. Takano, et al. (1999). "Enhanced coexpression of YB-1 and DNA topoisomerase II alpha genes in human colorectal carcinomas." Int J Cancer **83**(6): 732-737.
- Shiota, M., H. Izumi, et al. (2008). "Twist promotes tumor cell growth through YB-1 expression." Cancer Res **68**(1): 98-105.
- Sinnberg, T., B. Sauer, et al. (2012). "MAPK and PI3K/AKT mediated YB-1 activation promotes melanoma cell proliferation which is counteracted by an autoregulatory loop." Exp Dermatol **21**(4): 265-270.
- Skabkin, M. A., O. I. Kiselyova, et al. (2004). "Structural organization of mRNA complexes with major core mRNP protein YB-1." Nucleic Acids Res **32**(18): 5621-5635.
- Sorokin, A. V., A. A. Selyutina, et al. (2005). "Proteasome-mediated cleavage of the Y-box-binding protein 1 is linked to DNA-damage stress response." EMBO J **24**(20): 3602-3612.

- Stein, U., K. Jurchott, et al. (2001). "Hyperthermia-induced nuclear translocation of transcription factor YB-1 leads to enhanced expression of multidrug resistance-related ABC transporters." J Biol Chem **276**(30): 28562-28569.
- Stenina, O. I., K. M. Shaneyfelt, et al. (2001). "Thrombin induces the release of the Y-box protein dbpB from mRNA: a mechanism of transcriptional activation." Proc Natl Acad Sci U S A **98**(13): 7277-7282.
- Stratford, A. L., C. J. Fry, et al. (2008). "Y-box binding protein-1 serine 102 is a downstream target of p90 ribosomal S6 kinase in basal-like breast cancer cells." Breast Cancer Res **10**(6): R99.
- Stratford, A. L., G. Habibi, et al. (2007). "Epidermal growth factor receptor (EGFR) is transcriptionally induced by the Y-box binding protein-1 (YB-1) and can be inhibited with Iressa in basal-like breast cancer, providing a potential target for therapy." Breast Cancer Res **9**(5): R61.
- Stupp, R., M. E. Hegi, et al. (2009). "Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial." Lancet Oncol **10**(5): 459-466.
- Stupp, R., W. P. Mason, et al. (2005). "Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma." N Engl J Med **352**(10): 987-996.
- Sturm, D., H. Witt, et al. (2012). "Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma." Cancer Cell **22**(4): 425-437.
- Sutherland, B. W., J. Kucab, et al. (2005). "Akt phosphorylates the Y-box binding protein 1 at Ser102 located in the cold shock domain and affects the anchorage-independent growth of breast cancer cells." Oncogene **24**(26): 4281-4292.
- Tabori, U., B. Baskin, et al. (2010). "Universal poor survival in children with medulloblastoma harboring somatic TP53 mutations." J Clin Oncol **28**(8): 1345-1350.

- Talbert, P. B. and S. Henikoff (2010). "Histone variants--ancient wrap artists of the epigenome." Nat Rev Mol Cell Biol **11**(4): 264-275.
- TCGA (2008). "Comprehensive genomic characterization defines human glioblastoma genes and core pathways." Nature **455**(7216): 1061-1068.
- Teo, S. M., Y. Pawitan, et al. (2012). "Statistical challenges associated with detecting copy number variations with next-generation sequencing." Bioinformatics **28**(21): 2711-2718.
- Theeler, B. J., W. K. Yung, et al. (2012). "Moving toward molecular classification of diffuse gliomas in adults." Neurology **79**(18): 1917-1926.
- Toyota, M., N. Ahuja, et al. (1999). "CpG island methylator phenotype in colorectal cancer." Proc Natl Acad Sci U S A **96**(15): 8681-8686.
- Turcan, S., D. Rohle, et al. (2012). "IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype." Nature **483**(7390): 479-483.
- Uchiumi, T., A. Fotovati, et al. (2006). "YB-1 is important for an early stage embryonic development: neural tube formation and cell proliferation." J Biol Chem **281**(52): 40440-40449.
- Uramoto, H., H. Izumi, et al. (2002). "p73 Interacts with c-Myc to regulate Y-box-binding protein-1 expression." J Biol Chem **277**(35): 31694-31702.
- Verhaak, R. G., K. A. Hoadley, et al. (2010). "Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1." Cancer Cell **17**(1): 98-110.
- Villard, L., J. Gecz, et al. (1996). "XNP mutation in a large family with Juberg-Marsidi syndrome." Nat Genet **12**(4): 359-360.
- Vinay Kumar, A. K. A., Nelson Fausto (2005). Robbins and Cotran Pathologic Basis of Disease. Philadelphia, Elsevier.
- Wang, K., M. Li, et al. (2010). "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data." Nucleic Acids Res **38**(16): e164.

- Weinberg, R. A. (2007). The Biology of Cancer. New York, USA, Garland Science, Taylor & Francis Group, LLC.
- Wetterstrand, K. A. (May 21, 2012). "DNA Sequencing costs: Data from the NHGRI Large-Scale Genome Sequencing Program." Retrieved November 26, 2012, from [www.genome.gov/sequencingcosts](http://www.genome.gov/sequencingcosts).
- Wong, K. K., Y. T. Tsang, et al. (2006). "Genome-wide allelic imbalance analysis of pediatric gliomas by single nucleotide polymorphic allele array." Cancer Res **66**(23): 11172-11178.
- Wong, L. H., J. D. McGhie, et al. (2010). "ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells." Genome Res **20**(3): 351-360.
- Wu, G., A. Broniscer, et al. (2012). "Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas." Nat Genet **44**(3): 251-253.
- Wu, H. and Y. Zhang (2011). "Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation." Genes Dev **25**(23): 2436-2452.
- Wu, J., C. Lee, et al. (2006). "Disruption of the Y-box binding protein-1 results in suppression of the epidermal growth factor receptor and HER-2." Cancer Res **66**(9): 4872-4879.
- Xu, W., H. Yang, et al. (2011). "Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases." Cancer Cell **19**(1): 17-30.
- Xu, W., L. Zhou, et al. (2009). "Nuclear expression of YB-1 in diffuse large B-cell lymphoma: correlation with disease activity and patient outcome." Eur J Haematol **83**(4): 313-319.
- Yan, H., D. W. Parsons, et al. (2009). "IDH1 and IDH2 mutations in gliomas." N Engl J Med **360**(8): 765-773.
- Yan, X. J., J. Xu, et al. (2011). "Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia." Nat Genet **43**(4): 309-315.

- Yasen, M., K. Kajino, et al. (2005). "The up-regulation of Y-box binding proteins (DNA binding protein A and Y-box binding protein-1) as prognostic markers of hepatocellular carcinoma." Clin Cancer Res **11**(20): 7354-7361.
- Yen, K. E., M. A. Bittinger, et al. (2010). "Cancer-associated IDH mutations: biomarker and therapeutic opportunities." Oncogene **29**(49): 6409-6417.
- Zhang, Y. F., C. Homer, et al. (2003). "Nuclear localization of Y-box factor YB1 requires wild-type p53." Oncogene **22**(18): 2782-2794.