

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Evolution of low-grade glioma through intratumoral heterogeneity of the genome and epigenome

Permalink

<https://escholarship.org/uc/item/28j577k5>

Author

Mazor, Tali

Publication Date

2016

Peer reviewed|Thesis/dissertation

Evolution of low-grade glioma through intratumoral heterogeneity of
the genome and epigenome

by

Tali Mazor

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Copyright 2016
by
Tali Mazor

DEDICATION AND ACKNOWLEDGMENTS

Parts of chapters one and six of this dissertation are a reprint of material in Mazor & Pankov et al. Intratumoral heterogeneity of the epigenome. *Cancer Cell* **29**, 440-451 (2016). The text of chapter two and parts of chapter five of this dissertation are a reprint of material in Johnson & Mazor et al. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* **343**, 189-193 (2014). The text of chapter three and parts of chapter five of this dissertation are a reprint of material in Mazor & Pankov et al. DNA methylation and somatic mutations converge on the cell cycle and define similar evolutionary histories in brain tumors. *Cancer Cell* **28**, 307-317 (2015).

The research described in this dissertation is the result of hard work and support from many people. First I want to thank my mentor, Joseph Costello. Joe has supported me and my research every step of the way. He has helped me grow as a scientist and as a person. I will be forever grateful for his unwavering confidence in me, which has given me the strength to bring these projects to fruition. Joe has built a lab full of smart and collaborative people, a group that I am proud to call friends as well as colleagues. I want to thank everyone who has been a part of the Costello Lab over the years. In particular, I want to acknowledge Brett Johnson and Alex Pankov for their significant contributions to this research. Brett graciously took me under his wing when I first joined lab – he brought me up to speed and shared his meticulous approach to research and writing with me. Alex came in with a different viewpoint and skillset that changed how I approached data analysis. I am a much better researcher as a result of working with Brett and Alex.

I would also like to thank the many other people that have mentored me throughout my scientific career, both officially and unofficially. I owe a great deal to Jared Shaw, who took a

chance on a young undergraduate and showed me how fun and exciting research can be, and Ernest Fraenkel, who gave me my start with data analysis and provided me with the experience that convinced me to apply to graduate school. During graduate school, there have been many people who have helped me grow, both as a scientist and a person, and I would like to particularly thank Barry Taylor, Adam Olshen, Annette Molinaro, Joanna Phillips, Andrew Bollen and Susan Chang for their mentorship and teaching.

I want to thank all of my friends who have supported and encouraged me from near and far. In particular, I want to acknowledge my Minty House family – Anna Gillespie, Bevin English, Lindsey Battenberg, Ally Girasole and Liz Krow-Lucal – thank you for always supporting me through the lows and celebrating the highs with me.

Finally, I want to thank my family, without whom I never could have done this. Thank you to all of my extended family, who have supported me and cheered me on, no matter the physical distance separating us. My brothers, Ofer and Yuval, have always been my role models. As a child, all of my teachers knew me as their little sister and expected me to live up to their successes. They set high standards and inspired me to achieve as much as I could. They also gave me two warm and supportive sisters-in-law – Melissa and Dori – and four perfect nieces and nephews – Ilan, Ayelet, Gili and Aviv. Lastly, I want to thank my parents. They have supported me and encouraged me in everything I've ever done, no matter how big or small, and in doing so they made it possible for me to follow my dreams. Thank you!

Evolution of low-grade glioma through intratumoral heterogeneity of the genome and epigenome

Tali Mazor

Low-grade glioma (LGG) is a diffuse and infiltrative adult brain tumor. Due to the invasive nature of the tumor, surgical resection is rarely curative. LGG often undergo malignant progression and recur as a high-grade glioblastoma (GBM). The genomic and epigenomic landscapes of these tumors, particularly at recurrence, are understudied yet clinically important. Recurrent tumors may be driven by a distinct set of genetic and epigenetic alterations than their initial tumors, yet therapeutic decisions are often made based on profiling of initial malignancies. Here we comprehensively profiled 33 LGGs and their patient-matched recurrences, including cases with 2-6 intratumoral samples, with exome sequencing to identify somatic mutations, with the Infinium 450K array to investigate DNA methylation changes, and with RNA sequencing to measure gene expression changes. We found a wide range in the degree of evolution from initial to recurrence, in terms of both somatic mutations and DNA methylation changes. Tumors treated with the alkylating chemotherapy temozolomide (TMZ) often recurred with a hypermutation phenotype that was suggestive of therapy-driven malignant progression to GBM. Therapy-associated mutational activation of the AKT-mTOR pathway was a consistent feature of these hypermutated recurrences, which promoted a new clinical trial combining the anti-tumor activity of TMZ with an mTOR inhibitor. Recurrence as GBM was associated with aberrations to cell cycle genes through convergence of both genetic and epigenetic mechanisms. Moreover, we found that the evolutionary history of a tumor is similar whether inferred from genetic or epigenetic data, suggesting co-evolution of the genome and epigenome. Finally, we identified

cases in which mutations in *IDH1*, which are the earliest known alteration in LGG and drive gliomagenesis, were either deleted or amplified at recurrence. Mutant IDH1 reprograms the epigenome and these recurrences showed partial reversion of these epigenomic alterations. Together, these findings highlight the heterogeneity and continual evolution of LGG and emphasize the importance of studying recurrent tumors for a more complete understanding of tumor evolution and to make more informed treatment decisions.

TABLE OF CONTENTS

Dedication and acknowledgements	iii
Abstract	v
Table of contents	vii
List of tables	xi
List of figures	xii

Chapter 1: Introduction

1.1 Low-grade glioma	2
1.2 Epigenetic modifications	4
1.3 Genetic and epigenetic drivers of cancer	5
1.4 Spatial and temporal tumor heterogeneity and the inference of tumor evolutionary history. .	7
1.5 Aims of this study	9

Chapter 2: Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma

2.1 Abstract	12
2.2 Main text	12

Chapter 3: DNA methylation and somatic mutations converge on cell cycle and define similar evolutionary histories in brain tumors

3.1 Abstract	52
3.2 Significance	52

3.3 Introduction	52
3.4. Results	54
3.4.i <i>Patterns of DNA methylation are patient-specific and evolve in a manner specific to the grade of recurrence</i>	55
3.4.ii <i>Identification of CpGs that lose methylation specifically during malignant progression to GBM</i>	56
3.4.iii <i>Cell cycle genes are specifically hypomethylated upon malignant progression</i>	58
3.4.iv <i>Reconstruction of tumor evolution from intratumoral and longitudinal DNA methylation patterns</i>	59
3.4.v <i>Enhanced model of tumor evolution derived from variation between phyloepigenetic and phylogenetic trees</i>	63
3.4.vi <i>Gene-level genetic and epigenetic convergence</i>	63
3.5 Discussion	64

Chapter 4: Extreme DNA hypomethylation and tumor progression following deletion or amplification of mutant *IDH1* in glioma

4.1 Abstract	87
4.2 Main text	87

Chapter 5: Materials and methods

5.1 Sample acquisition	108
5.2 DNA and RNA isolation	109
5.3 Exome sequencing	109

5.3.i	<i>Hybrid capture, sequencing and alignment</i>	109
5.3.ii	<i>Mutation identification</i>	109
5.3.iii	<i>Mutation validation: Sanger sequencing</i>	111
5.3.iv	<i>Mutation validation: PCR analysis of ATRX deletion</i>	112
5.3.v	<i>Mutation validation: droplet digital PCR</i>	113
5.3.vi	<i>Mutation rates</i>	114
5.3.vii	<i>Copy number analysis (Chapter 2)</i>	115
5.3.viii	<i>Copy number and decrease of heterozygosity analysis (Chapter 4)</i>	115
5.3.ix	<i>Tumor cell fraction analysis</i>	116
5.4	<i>DNA methylation</i>	117
5.4.i	<i>Illumina 450K array and preprocessing</i>	117
5.4.ii	<i>Differential methylation analysis (Chapter 3)</i>	117
5.4.iii	<i>Age-related methylation analysis</i>	118
5.4.iv	<i>Methylation analysis (Chapter 4)</i>	118
5.4.v	<i>Whole genome shotgun bisulfite sequencing</i>	119
5.4.vi	<i>MGMT bisulfite sequencing analysis</i>	120
5.4.vii	<i>MGMT methylation-specific PCR</i>	121
5.5	<i>Transcriptome sequencing</i>	121
5.5.i	<i>Library construction and sequencing</i>	121
5.5.ii	<i>Alignment and analysis (Chapter 2)</i>	122
5.5.iii	<i>Alignment and analysis (Chapter 3)</i>	122
5.6	<i>P value adjustment (Chapter 3)</i>	123
5.7	<i>Phylogenetic and phyloepigenetic trees</i>	124

5.7.i <i>Phylogenetic trees (Chapter 2)</i>	124
5.7.ii <i>Phylogenetic and phyloepigenetic trees (Chapters 3 and 4)</i>	125
5.7.iii <i>Identifying discriminative methylation probes by singular value decomposition</i>	125
5.8 Analysis of gene-level convergence	126
5.8.i <i>Analysis of gene-level convergence within each patient</i>	126
5.8.ii <i>Analysis of gene-level convergence among the patient cohort</i>	127
5.9 Gene ontology and enrichment analyses	127
5.10 Immunohistochemistry	127
5.11 Fluorescence in situ hybridization	128
5.12 Ex vivo HR-MAS spectroscopy	128
5.13 Chromatin immunoprecipitation sequencing	129
5.14 Expression of exogenous mTOR and western blots	130
5.15 Radiologic analysis	131
Chapter 6: Discussion	
6.1 Hypermutation	133
6.2 Co-evolution of genomics and epigenomics	135
6.3 Clinical implications	137
Bibliography	140

LIST OF TABLES

Table 2.1	Summary of the clinical and molecular features of each tumor in the cohort	45
Table 2.2	Quality control metrics for exome sequencing data	46
Table 2.3	Patterns of genetic evolution between initial and recurrent gliomas	48
Table 2.4	Mutation rates in each tumor at each dinucleotide context and overall	49
Table 2.5	Sanger sequencing of TMZ-associated mutations	50
Table 3.1	Summary of the data types, clinical and molecular features of each tumor in the cohort	83
Table 3.2	QC of all sequencing datasets	84
Table 3.3	Intra- and inter-patient gene-level convergence	85
Table 4.1	Summary of patients identified and samples profiled with copy number change at <i>IDH1</i>	104
Table 4.2	Summary of IHC and FISH results	105
Table 4.3	Results of Ki67 quantification in patients with regionally heterogeneous IDH1 R132H staining patterns	106

LIST OF FIGURES

Figure 1.1	Molecular classification of adult glioma	10
Figure 2.1	Genetic landscapes of low-grade gliomas and their patient-matched recurrences	21
Figure 2.2	Commonly mutated genes in grade II glioma	23
Figure 2.3	Tumor cell fraction of somatic mutations in paired initial and recurrent tumors	24
Figure 2.4	Sensitive detection of mutant alleles	25
Figure 2.5	The temporal and spatial patterns of clonal evolution in two glioma patients ...	27
Figure 2.6	Sensitive PCR analysis of an ATRX deletion in patient 17	28
Figure 2.7	Shared non-coding mutations in patient 17	29
Figure 2.8	Pre- and post-surgical MR imaging of patient 17	30
Figure 2.9	Pre- and post-surgical MR imaging of patient 04	32
Figure 2.10	The spectrum and context of somatic mutations in hypermutated and non-hypermutated gliomas	34
Figure 2.11	Recurrent tumors from patients treated with TMZ harbor genetic alterations in the RB and AKT-mTOR signaling pathways	35
Figure 2.12	The evolution of DNA methylation affecting <i>MGMT</i> during malignant progression	37
Figure 2.13	Functional assessment of <i>MTOR</i> mutations on mTORC1 signaling	38
Figure 2.14	Comparison of tumor samples from the first recurrence of patient 01	39
Figure 2.15	mTORC1 signaling in tumors at initial diagnosis and their GBM recurrences ..	40
Figure 2.16	An integrated timeline of the treatment, imaging, and clonal evolution of a low-grade astrocytoma that underwent TMZ-associated malignant progression	42
Figure 2.17	Pre- and post-surgical magnetic resonance imaging of patient 01	43

Figure 3.1	Evolutionary dynamics of the methylome and transcriptome in initial and recurrent glioma pairs	67
Figure 3.2	Plots supporting Figure 3.1	69
Figure 3.3	Cell cycle genes are hypomethylated and over-expressed specifically upon recurrence as GBM, coordinately with an increase in actively cycling cells	71
Figure 3.4	Plots supporting Figure 3.3	73
Figure 3.5	The spatial and temporal patterns of tumor evolution observed from DNA methylation dynamics and somatic mutations yield similar evolutionary histories	75
Figure 3.6	Plots supporting Figure 3.5	77
Figure 3.7	Phyloepigenetic trees coupled with phylogenetic trees from a low-grade glioma patient with three recurrences reveal an enhanced understanding of evolutionary relationships	79
Figure 3.8	Plots supporting Figure 3.7	80
Figure 3.9	A genomic and epigenomic co-dependency model of clonal evolution	81
Figure 4.1	Identification of two cases with deletion of mutant <i>IDH1</i>	95
Figure 4.2	DoH and TCN plots for all cases with allelic imbalance at <i>IDH1</i>	96
Figure 4.3	IHC and FISH from Patients 169 and 21	98
Figure 4.4	Representative images from H&E and IDH1 R132H IHC	99
Figure 4.5	Classification of <i>IDH1</i> copy number change tumors relative to <i>IDH</i> mutant and wild type gliomas	100
Figure 4.6	Average methylation levels following change in <i>IDH1</i> copy number	101
Figure 4.7	Similarity of methylation change following deletion or amplification of mutant <i>IDH1</i>	102

CHAPTER 1:
INTRODUCTION

1.1 Low-grade glioma

Gliomas are a class of brain tumors generally characterized by the diffuse and invasive nature of the tumor cells. Adult gliomas, which have an incidence of 5.15 per 100,000 in the United States (1), have historically been classified morphologically into astrocytic and oligodendroglial lineages, and further sub-classified into low- and high-grade disease based on histological characteristics including the presence of mitotic figures, microvascular proliferation and necrosis with pseudopalisading cells. Based on a classification schema from the World Health Organization (WHO), gliomas are graded based on increasing presence of malignant features as grade II (diffuse low-grade glioma, or LGG), grade III (anaplastic glioma), or grade IV (glioblastoma, or GBM) (2). Given the invasive nature of these tumors, they almost invariably recur (3). At recurrence, a tumor may be the same grade as the initial tumor or it may undergo malignant progression to a higher grade. Average survival time depends on initial diagnosis and can range from 13 years (grade II) (4) to just 14 months (grade IV) (5).

Years of molecular characterization, particularly over the past decade of genome-wide sequencing, have identified common molecular features for each subtype of adult glioma, some of which are now incorporated into clinical diagnosis. Molecular classifications are based on the presence or absence of mutations in *IDH1* or *IDH2*, *TP53*, *ATRX*, and the promoter of *TERT*, as well as several genomic gains and losses (Figure 1.1) (6-8). Grade II astrocytomas canonically have mutations in *IDH1* or *IDH2*, *TP53*, and *ATRX*, whereas grade II oligodendrogliomas have mutations in *IDH1* or *IDH2*, the *TERT* promoter and codeletion of chromosomal arms 1p and 19q. Primary GBM lack mutations in *IDH1* or *IDH2*, but have *TERT* promoter mutations, gain of chromosome 7 and loss of chromosome 10. Recent studies suggest that incorporation of these molecular features stratifies patients better than a purely histological diagnosis (7, 8).

Standard of care for patients depends on the diagnosis. For GBM, standard of care is surgical resection followed by radiation and treatment with the alkylating chemotherapy temozolomide (TMZ) (5). Lower-grade tumors are also first treated with surgical resection, if possible, but the benefit of adjuvant therapies is only just becoming clear (3, 4). Radiation has been used historically, but the cognitive defects associated with this treatment are of great concern given the relatively younger age of patients with lower grade tumors and longer survival relative to GBM. The benefit of radiation and the combination chemotherapy PCV (procarbazine, CCNU and vincristine) has been shown for oligodendrogliomas, although the benefit in astrocytic tumors is less clear (4, 9, 10). TMZ has a similar mechanism of action as PCV, with reduced side effects and oral administration, and so is a common treatment option as well. However, there has not yet been a large-scale clinical trial to demonstrate the benefit of TMZ only in lower-grade patients.

The most common chemotherapeutics administered to patients with both low- and high-grade glioma are alkylating agents (TMZ, PCV, CCNU, BCNU). These drugs act by adding methyl or alkyl groups to DNA bases, causing DNA damage that should lead to cell death. As an example, TMZ adds a methyl group to DNA bases in several locations, including to the O6 position of guanine (11). This adduct is normally removed by O6-methyl-guanine-methyltransferase (MGMT), a suicide protein that removes that specific adduct and is then degraded. However, tumor cells often methylate the promoter of *MGMT*, silencing transcription of the gene. In the absence of MGMT, the adduct will persist and upon DNA replication, the O6-methyl group alters the base pairing of guanine such that O6-methyl-guanine base pairs with thymine rather than cytosine. This abnormal base pair is flagged by the mismatch repair (MMR) complex, which will excise the thymine. However, because the O6-methyl group remains,

thymine will again be incorporated at the same position. This leads to what is termed “futile cycling of mismatch repair”, and eventually to cell death. However, if a cell is deficient in MMR, through epigenetic silencing, mutation or deletion of an essential MMR protein, the cell will not die. Instead, these G:T mispairs will persist through DNA replication, leading to a G>A transition mutation in a daughter cell (12). This presumed mechanism of action is consistent with the finding that methylation at MGMT is predictive of response to TMZ (13). Moreover, several groups have found recurrent GBMs with an abnormally high rate of G>A mutations following treatment with TMZ (14-17). These hypermutated recurrences had MGMT hypermethylation and mutations in MMR genes (18).

1.2 Epigenetic modifications

Epigenetics is the study of mitotically heritable and reversible changes in gene expression or cellular phenotype that are not derived from genetic changes. Epigenetic changes can be thought about as alterations in DNA accessibility, whereas genetic changes alter the DNA sequence itself. The most commonly studied epigenetic changes are chromatin remodeling, histone modifications and DNA methylation; it is on the latter two mechanisms that this dissertation will focus, particularly on DNA methylation.

Within the nucleus of a cell, DNA is in complex with proteins, in a structure called chromatin that controls accessibility to the DNA. Linear DNA is wrapped around an octamer of histone proteins. One end of each histone protein, called the “tail”, can be covalently modified in a variety of ways that alter DNA-histone interactions and thus impact DNA accessibility. Certain histone modifications are associated with particular chromatin states: for example, addition of three methyl groups to the fourth lysine of histone three (H3K4me3) is associated with the

promoter of actively transcribed genes whereas tri-methylation of the ninth lysine of histone three (H3K9me3) is associated with inactive genes and a repressive chromatin state (19).

In mammalian cells, DNA methylation refers to the addition of a methyl group to the five position of the nucleotide cytosine (5meC). Specifically, when the linear sequence of DNA is a cytosine followed by a guanine (a CpG dinucleotide), the cytosine can be methylated. Within the genome there are regions with a high density of CpG sites, called CpG islands (CGIs), which are generally unmethylated (20). Outside of CGIs, most CpG sites are methylated (20). The effects of DNA methylation on gene expression are variable and depend upon the genomic context. CpG methylation in promoter regions is associated with gene silencing (21), whereas CpG methylation in gene bodies is associated with active transcription (22, 23). DNA methylation can have more subtle effects as well, such as CpG methylation near intron/exon boundaries which can alter RNA splicing (24, 25). Importantly, like all epigenetic modifications, DNA methylation is reversible and dynamically regulated. DNA methylation is laid down by DNA methyltransferases (DNMTs) (26) and removed by TET enzymes (27-29).

1.3 Genetic and epigenetic drivers of cancer

Historically, the drivers of cancer have been considered through the lens of genetic defects (30-33). Activation of oncogenes, through somatic mutation (34-36) or amplification (37-40) and inactivation of tumor suppressor genes, through somatic mutation (41-44) or biallelic deletion (45), as well as karyotype abnormalities (46, 47), provided the basis for early studies into the mechanisms of tumor growth.

Alteration to the epigenome is a fundamental characteristic of nearly all human cancers. Pioneering studies focused on DNA methylation and identified decreased 5-methylcytosine

content in tumors compared to normal tissue (48, 49), further loss of 5-methylcytosine during tumor progression (50) and increased methylation in normally unmethylated CpG islands and promoter regions. These promoter-based analyses were primarily focused on tumor suppressors that had been previously identified through genetic mechanisms, including *RB1* (51-53), *VHL* (54) and *p16* (55-57), as an alternative mechanism of eliminating functional protein.

Some early studies also identified genes that were rarely subject to mutation but often silenced through epigenetic mechanisms, including *MGMT* in glioma (58, 59). Indeed, subsequent studies revealed that DNA methylation primarily affected different sites than genetic alterations (60-63), suggesting independent but complementary and possibly interdependent roles for genetic and epigenetic alterations in tumorigenesis. Alteration to DNA methylation alone was shown to be sufficient to drive tumorigenesis in mouse models (64). Studies in both colorectal cancer and melanoma found that a common somatic mutation, *BRAF* V600E, is tightly associated with an altered epigenomic landscape, and may in fact require that aberrant epigenome (65-68).

More recently, the decreased cost of DNA sequencing and the resulting large-scale studies of somatic mutations have identified a common theme of somatic mutations in epigenome- and chromatin-related genes. Mutations in chromatin modifiers (69, 70) and in histone proteins themselves (71-74) are common across cancer types. Additionally, the highly recurrent mutations in *IDH1* and *IDH2*, which are common in glioma, AML, and several other cancer types (75), have a significant impact on the epigenome, although through a more indirect mechanism. Wild-type IDH enzymes catalyze the conversion of isocitrate to α -ketoglutarate (α KG), which is a cofactor for many cellular enzymes. Mutations in *IDH1* and *IDH2* lead to a neomorphic ability to convert α KG to 2-hydroxyglutarate (2HG) (76, 77). 2HG is a competitive

inhibitor of α KG-dependent enzymes, including chromatin modifiers and the TETs, leading to global changes in histone modifications and DNA methylation (in glioma, this is called the Glioma CpG Island Methylator Phenotype or G-CIMP) that are believed to drive tumorigenesis (78-83). 2HG is thus considered an onco-metabolite. Inhibition of mutant IDH enzymes, and subsequent depletion of 2HG, is an exciting new therapeutic approach (84, 85); there are currently ten open clinical trials testing five different inhibitors of IDH1 and IDH2 from three companies (86).

1.4 Spatial and temporal tumor heterogeneity and the inference of tumor evolutionary history

The theory underlying evolutionary biology is credited to Charles Darwin's seminal work, "On the Origin of Species by Means of Natural Selection" (87). In it, he proposes a branched evolutionary model for the development of species, in which species evolve from a common ancestor, diverging through natural selection to adapt to changing environments. The only figure in this work is a representation of branched evolution of species as a tree diagram, a representation still used today. However, Darwin's analyses were based strictly on morphological similarities and differences; he lacked any knowledge of Gregor Mendel's work on the inheritance of traits (88), or of DNA as the means of heredity (89). Today, as a result of the synthesis of these many fields, we can infer evolutionary relationships among species based on their DNA sequence (phylogenetics).

A similar framework can be applied to tumor evolution. First suggested in 1976 by Peter Nowell (90), clonal evolutionary theory provides a basis to infer the evolutionary history of a tumor, including the order in which molecular alterations were acquired (91, 92). Inference of

the evolutionary history of a tumor from somatic mutations relies on the pattern of shared mutations across multiple samples of a tumor: mutations present in all samples of a tumor are inferred to be acquired by early precursor cells which clonally expanded (clonal mutations); in contrast, mutations present in only a subset of samples are inferred to be later events, acquired at some point during or after the initial clonal expansion (subclonal mutations). A seminal publication (93) integrated the next-generation sequencing of tumors with principles from the clonal evolution theory of tumors. Gerlinger et al performed exome-sequencing of 14 and 10 spatially distinct biopsies from two individuals with metastatic renal cell carcinoma (RCC). Taking advantage of the genetic intratumoral heterogeneity delineated by the multiple samplings per tumor and analyzing the patterns of shared and unique mutations, early and late events were distinguished. Together the events revealed a branched evolutionary history with several instances of convergent evolution in which the same gene was mutated independently in multiple subclones within a single tumor. For each patient, these findings were presented with a phylogenetic tree, a graphical representation of the evolutionary history of a patient's tumor, as inferred from somatic mutations.

Recent work in solid tumors has extended genome-wide profiling of multiple intratumoral samples from genomics into epigenomics. Given that DNA methylation is reversible and more error prone than DNA replication, the evolutionary history of a tumor might appear different when inferred from genetic versus epigenetic data from the same intratumoral samples. Aryee et al (94) and Brocks et al (95) used arrays to simultaneously profile genome-wide DNA methylation and copy number alterations (CNA) in prostate cancer. Brocks et al examined multiple samples from the primary tumor site as well as premalignant lesions and metastases from five individuals, while Aryee et al examined metastases from thirteen subjects.

These studies revealed that while prostate-relevant enhancers frequently demonstrated methylation intratumoral heterogeneity (95), the sites of promoter methylation intratumoral heterogeneity and the expression of target genes did not correlate well (94). This may indicate that DNA methylation changes that alter gene expression are more likely to be selected for and become relatively homogeneously present across the tumor (94). In both studies, parallel analysis of the genome-wide DNA methylation and CNA produced highly similar tumor evolutionary histories.

1.5 Aims of this study

The genomic and epigenomic profiles of a recurrent tumor are clinically important. Following initial tumor resection, the tumor cells that remain are the target of adjuvant chemo- or radiation-therapy. It is these same residual tumor cells that must be the source of tumor recurrence; therefore profiling of recurrent tumors could allow inferences of drivers in residual disease and thereby inform better therapeutic decisions. However, these recurrences are rarely studied. The goal of this study was to understand the drivers of recurrent disease and the evolutionary history of tumors through comprehensive profiling of a unique cohort of patient-matched initial and recurrent tumors for both molecular and clinical features. We performed exome sequencing, transcriptome sequencing and DNA methylation arrays on these initial and recurrent tumors to understand the evolution of LGGs through recurrence and malignant progression, including the impact of chemotherapy on tumor evolution. In select patients, we were able to obtain multiple intratumoral samples, allowing for a comprehensive investigation of intratumoral heterogeneity and inference of the tumor's evolutionary history.

Figure 1.1

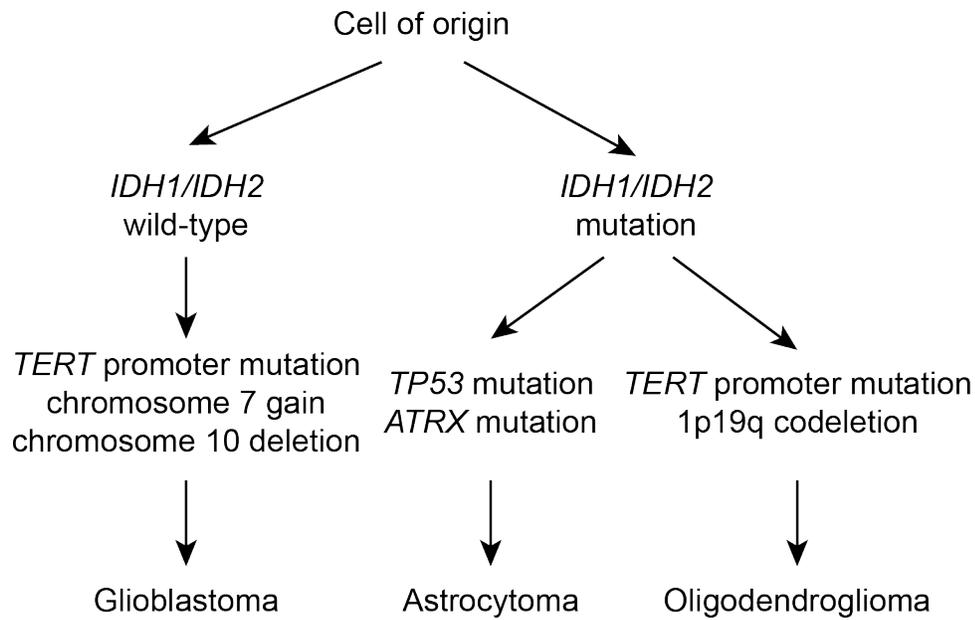


Figure 1.1. Molecular classification of adult glioma. Adult gliomas are classified primarily by the presence or absence of mutations in *IDH1* or *IDH2*. *IDH1/2*-mutant tumors are further classified by 1p19q codeletion status. Mutations in the *TERT* promoter, gain of chromosome 7 and loss of chromosome 10 are common in glioblastoma. *TERT* promoter mutations are also common in oligodendroglioma, while mutations in *TP53* and *ATRX* are common in astrocytoma.

CHAPTER 2:
MUTATIONAL ANALYSIS REVEALS THE ORIGIN AND THERAPY-DRIVEN
EVOLUTION OF RECURRENT GLIOMA

2.1 Abstract

Tumor recurrence is a leading cause of cancer mortality. Therapies for recurrent disease may fail, at least in part, because the genomic alterations driving the growth of recurrences are distinct from those in the initial tumor. To explore this hypothesis, we sequenced the exomes of 23 initial low-grade gliomas and recurrent tumors resected from the same patients. In 43% of cases, at least half of the mutations in the initial tumor were undetected at recurrence, including driver mutations in *TP53*, *ATRX*, *SMARCA4*, and *BRAF*, suggesting recurrent tumors are often seeded by cells derived from the initial tumor at a very early stage of their evolution. Notably, tumors from 6 of 10 patients treated with the chemotherapeutic drug temozolomide (TMZ) followed an alternative evolutionary path to high-grade glioma. At recurrence, these tumors were hypermutated and harbored driver mutations in the RB and AKT-mTOR pathways that bore the signature of TMZ-induced mutagenesis.

2.2 Main Text

The genetic landscape of tumors is continually evolving, which can be an impediment to the clinical management of cancer patients with recurrent disease (91, 96). In contrast to the clonal evolution of hematological malignancies (97, 98) and solid tumor metastases (93, 99, 100), the local regrowth of solid tumors after surgery occurs under a unique set of evolutionary pressures, which are further impacted by adjuvant therapies. Through acquisition of new mutations, residual tumor cells can progress to a more aggressive state. Grade II astrocytic gliomas are particularly troublesome from this perspective. While surgery is the standard of care, these invasive brain tumors typically recur (3). Many remain grade II at recurrence, while others

progress to a higher histological grade with a poor prognosis (101). The incidence and timing of malignant progression is variable and unpredictable (3).

We undertook genome sequence analysis of initial and recurrent human gliomas to address two questions: (i) what is the extent to which mutations in initial tumors differ from their subsequent recurrent tumors?; and (ii) how does chemotherapy with TMZ, a drug commonly used in the treatment of glioma, affect the mutational profile of recurrent tumors? We sequenced the exomes of 23 grade II gliomas at initial diagnosis and their recurrences resected from the same patients up to 11 years later (Table 2.1). We selected initial tumors of predominantly astrocytic histology that capture the full spectrum of glioma progression (histological grade II-IV at recurrence) and adjuvant treatment history. Tumor and matched normal DNA were sequenced to an average 125-fold coverage, enabling the sensitive detection of mutations down to a 10% variant frequency, small insertions/deletions, and DNA copy number alterations (CNAs) (Figure 2.1A and Table 2.2) (102).

We identified an average of 33 somatic coding mutations in each initial tumor, of which an average of 54% were also detected at recurrence (shared mutations) (Figure 2.1A). The shared mutations included those in *IDH1*, *TP53*, and *ATRX* in most but not all cases (Figure 2.2) (103-105). All other somatic mutations were identified only in the initial tumor or only in the recurrent tumor from a given patient (private mutations) and thus presumably arose later in tumor evolution. For example, mutations in *SMARCA4* were private to the initial or recurrent tumor in six of seven patients and therefore may confer a selective advantage in the context of pre-existing early driver events (106, 107). Overall, the initial and recurrent gliomas displayed a broad spectrum of genetic relatedness (Figure 2.3 and Table 2.3). At one end of this spectrum were four patients whose tumors showed a pattern of linear clonal evolution; we infer that the

recurrent tumors in these patients were seeded by cells bearing $\geq 75\%$ of the mutations detected in the initial tumors (as in patient 27, Figure 2.1B). At the other end of the spectrum, tumors from three patients showed branched clonal evolution; we infer that the recurrent tumors in these patients were seeded by cells derived from the initial tumor at an early stage of its evolution, as the recurrent tumors shared $\leq 25\%$ of mutations detected in the initial tumors. Patient 17 was an extreme example of branched clonal evolution, as the initial and recurrent tumors shared only the *IDH1* R132H mutation (Figure 2.1C). This further implicates *IDH1* mutations as an initiating event in low-grade gliomagenesis (104). Indeed, *IDH1* mutation was the only shared mutation in every patient, an observation that supports the current interest in *IDH1* as a therapeutic target (84). Paired tumors from the remaining 16 patients formed a continuum between linear and branched clonal evolution. Together, these data illustrate the extent to which genetically similar low-grade gliomas diverge after surgical resection, and that recurrences may emerge from early stages in the evolution of the initial tumor.

Many solid tumors, including glioblastoma (GBM) display intratumoral heterogeneity (108, 109). For example, geographically distinct parts of the tumor may have different mutations. Intratumoral heterogeneity could be a confounding factor in estimates of genetic divergence when only one relatively small fraction of a tumor is sampled. To explore the extent of intratumoral heterogeneity in our cases, we first analyzed the *BRAF* V600E mutation that was subclonal in the initial tumor of patient 18 and undetectable in the recurrent tumor by either exome sequencing or droplet digital PCR (Figures 2.1D and 2.4). *BRAF* V600E was present in three of six additional samples from geographically distinct regions of the initial tumor, while seven additional samples of the recurrence all lacked this mutation. This suggests the *BRAF*-mutant clone did not expand, despite the proliferative advantage typically conferred by this

mutation. This contrasts sharply with the selection and outgrowth of subclonal drivers during the evolution of chronic lymphocytic leukemias (97).

Beyond the actionable *BRAF* mutation, we sequenced the exomes of additional, geographically distinct samples from three cases to further determine the extent to which apparently private mutations might be misclassified due to intratumoral heterogeneity. For patient 17 in which all mutations except *IDHI* were private, intratumoral heterogeneity was observed in the initial and recurrent tumor. From the mutational profiles however, we inferred that three samples of the initial tumor and four samples of the recurrence all derived from a common tumor cell of origin that possessed only an *IDHI* R132H mutation (Figure 2.5A) (102). Moreover, the recurrent tumor contained driver mutations in *TP53* and *ATRX* distinct from those observed in the initial tumor. We found no evidence of these new *TP53* or *ATRX* mutations in the initial tumor at allele frequencies of ~0.1% (Figures 2.4 and 2.6), implying convergent phenotypic evolution (93) via a strong ongoing selection for loss of these genes. The initial and recurrent tumors likely did not arise independently, as they also shared three somatic non-coding mutations (Figure 2.7). Thus, the initial and recurrent tumors were only distantly related and, despite the local and relatively rapid recurrence (Figure 2.8), exonic mutations other than *IDHI* R132H were only transiently present during the course of this patient's disease. Finally, we sequenced the exomes of additional distinct samples of the initial and recurrent tumors from patients 26 and 27, broadening our assessment of the impact of intratumoral heterogeneity on the reported genetic divergence. We found only a small minority of private mutations were actually shared events (7%) (102). Intratumoral heterogeneity therefore could not explain the majority of the genetic divergence between the initial and recurrent tumors in our cohort, including the driver mutations in initial tumors that were undetected in their recurrence.

To investigate whether sequential recurrences from a single patient could each be traced to the same evolutionary stage of the initial tumor, we sequenced the exomes of the second and third recurrent tumors from patient 04 and constructed a disease phylogeny by clonal ordering (Figures 2.5B and 2.9) (92, 93, 102). The initial tumor and three sequential local recurrences were clonally related, as indicated by the shared phylogenetic branch containing early driver mutations in *IDH1* and *TP53*. We infer that the tumor cells seeding the second recurrence branched off from the initial tumor at a slightly earlier evolutionary stage than the cells seeding the first recurrence. In contrast, the third recurrent tumor was a direct outgrowth of the second recurrence. These results show that branched and linear patterns of clonal evolution occurred at differing times in the same patient and are therefore not intrinsic properties of the tumor.

Beyond maximal, safe, surgical resection, there is currently no standard of care for patients with low-grade glioma, and options include surveillance, adjuvant radiation alone, TMZ alone, or radiation and TMZ. TMZ is an alkylating agent that induces apoptosis in glioma cells and is sometimes used to defer or delay the use of radiation. However, there is currently no information on whether treatment of grade II astrocytomas with TMZ confers longer overall survival (3). As TMZ is also mutagenic (11), we sought to determine how adjuvant chemotherapy with TMZ affects the mutational profile of recurrent tumors by comparing the initial low-grade gliomas to their post treatment recurrence. While the initial tumors and most of the recurrent tumors in our cohort had 0.2-4.5 mutations per megabase (Mb) (17, 110), six of the ten patients treated with TMZ had recurrent tumors that were hypermutated; that is, they harbored 31.9-90.9 mutations per Mb (Table 2.4). Overall, 97% of these were C>T/G>A transitions predominantly occurring at CpC and CpT dinucleotides, a signature of TMZ-induced mutagenesis distinct from non-hypermutated tumors (Figure 2.10) (11, 15, 17). We classified

C>T/G>A transitions in each hypermutated tumor as TMZ-associated if they were undetected in the matched initial tumor, which was resected before TMZ treatment (Figure 2.11A). Although it is difficult to definitively attribute any single mutation to TMZ exposure, comparing the C>T/G>A mutation rates in each tumor pair suggested that >98.7% are due to TMZ-induced mutagenesis. To determine whether intratumoral heterogeneity in initial tumors resulted in the misclassification of some mutations as TMZ-associated, we sequenced the exomes of three additional geographically distinct samples of the untreated initial tumor from patient 18. For mutations classified as TMZ-associated, sequencing reads with the mutation were rare in the additional exomes and were found at rates no higher than expected by chance ($1.7 \pm 0.08\%$; p -value=0.5, Wilcoxon) further suggesting they are induced by TMZ.

Resistance to TMZ develops in part through the acquisition of mutations that inactivate the DNA mismatch repair (MMR) pathway. MMR pathway dysfunction and continued TMZ exposure can in turn result in hypermutation (14-17). Indeed, we found that hypermutated tumors acquired somatic mutations in MMR genes that were not detected in their initial tumors, as well as aberrant DNA methylation of O⁶-methylguanine-DNA methyltransferase (*MGMT*) (Figures 2.4, 2.12, and Table 2.1).

The introduction of thousands of *de novo* mutations may drive the evolution of TMZ-resistant glioma cells to higher states of malignant potential (15, 96). Indeed, all six recurrent tumors that showed evidence of TMZ-induced hypermutation underwent malignant progression to GBM, a high-grade tumor with a worse prognosis (3, 101). To investigate this hypothesis and identify TMZ-associated mutations that may drive the outgrowth of GBM from low-grade glioma, we focused on the RB and AKT-mTOR signaling pathways which are associated with high-grade gliomas (Figure 2.11B) (17, 111-113). In each hypermutated recurrence, TMZ-

associated mutations affected genes coding for essential signaling molecules in these two pathways. For example, in the RB pathway we identified a TMZ-associated RB1 c.2520+1G>A splice site mutation found previously in the germline of patients with hereditary retinoblastoma (114, 115). Transcriptome sequencing confirmed this mutation triggered aberrant splicing, premature termination, and loss of the RB1 C-terminal domain necessary for growth suppression (Figure 2.11C) (116). Recurrent tumors from patient 05 and patient 10 each had a TMZ-associated *CDKN2A* P114L mutation that prevents it from inhibiting CDK4 or inducing cell cycle arrest (117). The same mutation has been reported in other tumor types (118) and in the germline of patients with familial melanoma (119). Gene set enrichment analysis further confirmed deregulation of RB1-mediated cell cycle control upon tumor recurrence (Figure 2.11D), suggesting that TMZ-associated mutations compromise the function of the RB tumor suppressor pathway.

We also investigated TMZ-associated mutations that may activate the AKT-mTOR signaling pathway. We identified a TMZ-associated mutation *PIK3CA* E542K in the recurrent tumor of patient 18 that drives Akt hyperactivation and induces mTOR-dependent oncogenic transformation (120). Similarly, the TMZ-treated second recurrence of patient 24 had TMZ-associated mutations in *PTEN* (A121T and G165R) at residues critical to its phosphatase activity (121) that are recurrently mutated in GBM (118). Finally, we validated *in vitro* that a TMZ-associated *MTOR* S2215F mutation in the recurrent tumor of patient 01 was constitutively activating (Figure 2.13), similar to the previously identified *MTOR* S2215Y (122). Moreover, adjacent regions of this recurrence showed heterogeneous mTORC1 activity (Figures 2.11E and 2.14). Microdissection revealed that while these adjacent regions shared a subset of the mutations found in the initial tumor, *MTOR* S2215F and other TMZ-associated mutations were

present only in the region that stained strongly for mTORC1 activation, which also had higher Ki-67, implying that the TMZ-associated mutations conferred a proliferative advantage. A distal second recurrence harbored the same TMZ-associated mutations and stained strongly and homogeneously for mTORC1 targets (Figure 2.15). Although both regions of the first recurrence were GBM, the hypermutated subclone underwent *in vivo* selection, invaded distally, and seeded the second recurrence (Figures 2.16 and 2.17). Across our cohort, AKT-mTOR pathway mutations corresponded with elevated phospho-4E-BP1 and RPS6 *in vivo*, indicating hyper-activated mTORC1 in recurrent GBMs relative to their initial tumors (Figure 2.15).

There was no evidence that the mutations in the RB and AKT-mTOR signaling pathways preceded TMZ treatment, based on analysis of additional geographically distinct samples of initial tumors from four of the six patients with hypermutated recurrent tumors (Table 2.5). Non-hypermutated recurrent tumors that progressed to GBM also acquired genetic changes in these signaling pathways, but through alternative mechanisms. In contrast, none of the grade II-III recurrences acquired mutations in these pathways. These data suggest a connection between TMZ treatment, driver mutations in oncogenic signaling pathways, and malignant progression.

In summary, through direct comparison of the genomic landscape of gliomas at initial diagnosis and recurrence, we were able to infer the mutational character of the infiltrating tumor cells that give rise to recurrence and that adjuvant therapy with TMZ is intended to eliminate. Recurrences did not typically arise from cells bearing the full set of mutations found in the initial tumor, as would be expected from a local recurrence in the absence of selective pressure from adjuvant chemotherapy. This finding complicates the use of tumor genomics to design precision therapies targeting residual disease. We also demonstrated an alternative evolutionary path of low-grade glioma that is largely determined by adjuvant chemotherapy with TMZ. This extends

earlier studies of primary GBMs (14, 15), unpaired recurrent tumors (17), and a cell culture model (11). Future basic and clinical studies must weigh the initial antitumor effects of TMZ against the potential risk of inducing new driver mutations and malignant progression. Ultimately, a better understanding of the invading cells that give rise to recurrent tumors and the effect of adjuvant therapeutics on their evolution will facilitate the development of new strategies to delay or prevent recurrence and malignant progression.

Figure 2.1

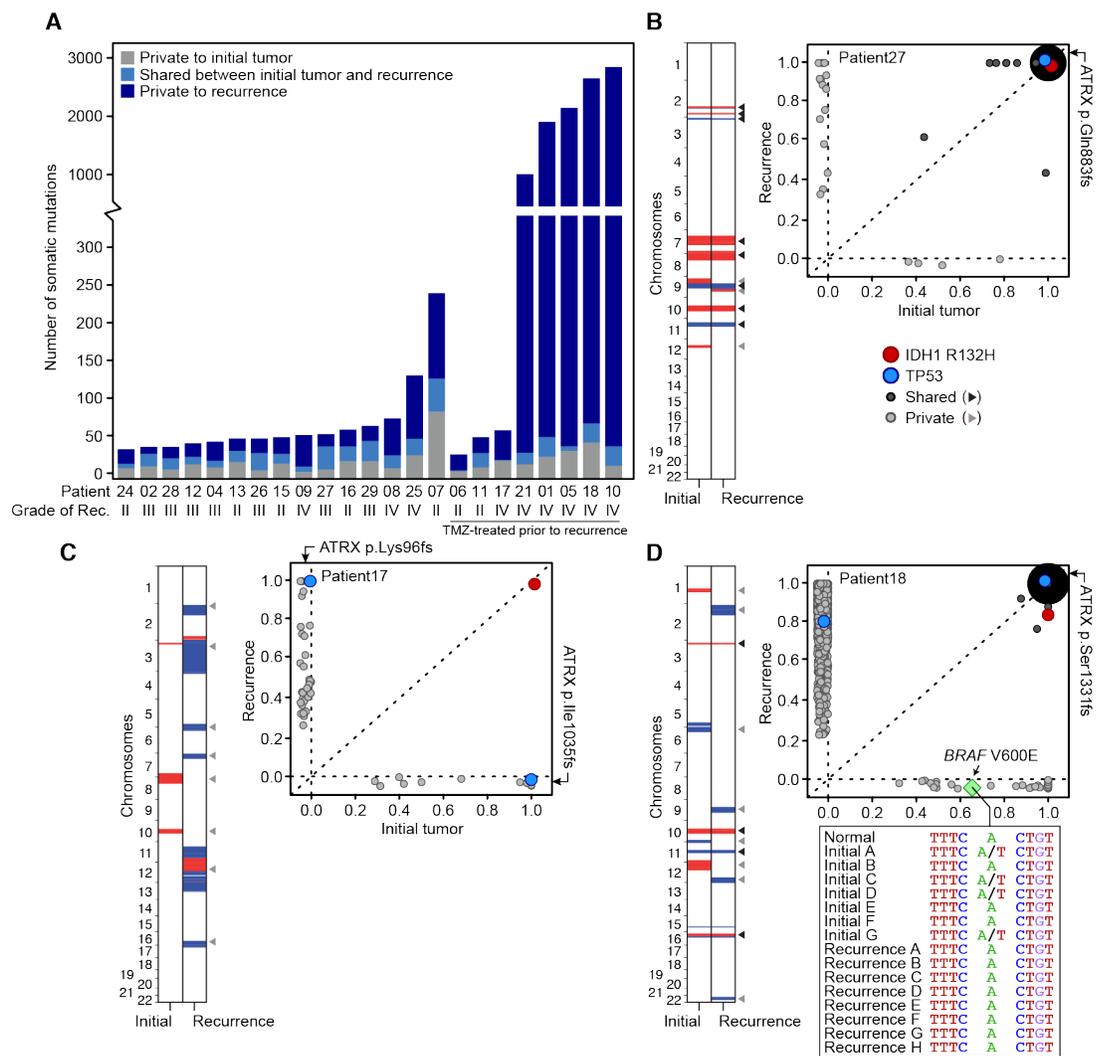


Figure 2.1. Genetic landscapes of low-grade gliomas and their patient-matched recurrences. (A) Total number of mutations private to or shared between the initial and first recurrent glioma of 23 patients. (B to D) Shared and private somatic mutations in paired initial and recurrent tumors (x and y axes respectively) as a function of the estimated fraction of tumor cells carrying the mutant allele. Mutations present in all the cells in both tumors are represented by a single point whose radius is scaled by the log count of such mutations. Shared and private CNAs are indicated (red and blue are gains and losses respectively, white is copy-neutral). In panel C, clonal *TP53* and *ATRX* mutations in the initial tumor were not identified in the recurrent

tumor, but different clonal mutations in these two genes were acquired. **(D)** Inset shows the DNA sequence encompassing *BRAF* V600E in the normal tissue and in 15 geographically distinct samples of the initial and recurrent tumors.

Figure 2.2

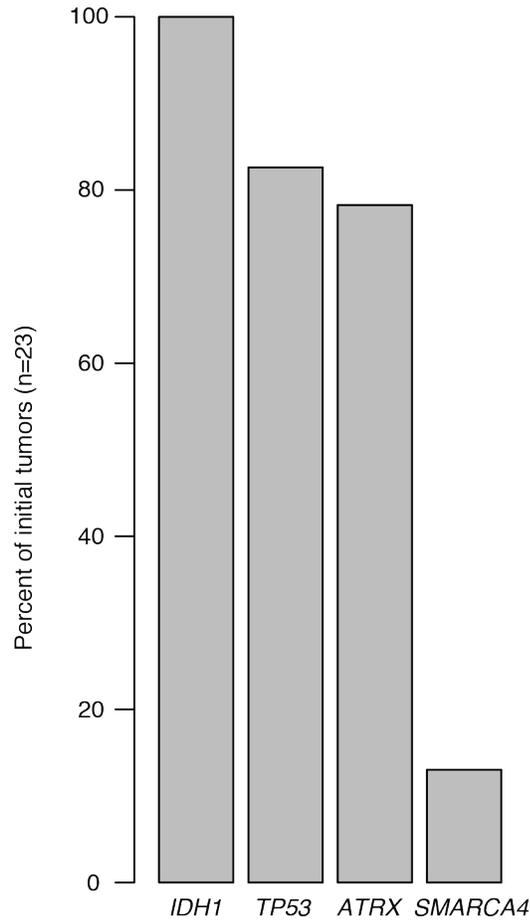


Figure 2.2. Commonly mutated genes in grade II glioma. The three genes most commonly mutated in grade II glioma at initial diagnosis are each identified in >75% (23/23, 19/23, 18/23) of this cohort. The next most commonly mutated gene, *SMARCA4*, is identified in 13% (3/23) of the initial tumors in this cohort. 13 additional genes are identified in 9% (2/23) of the cohort.

Figure 2.3

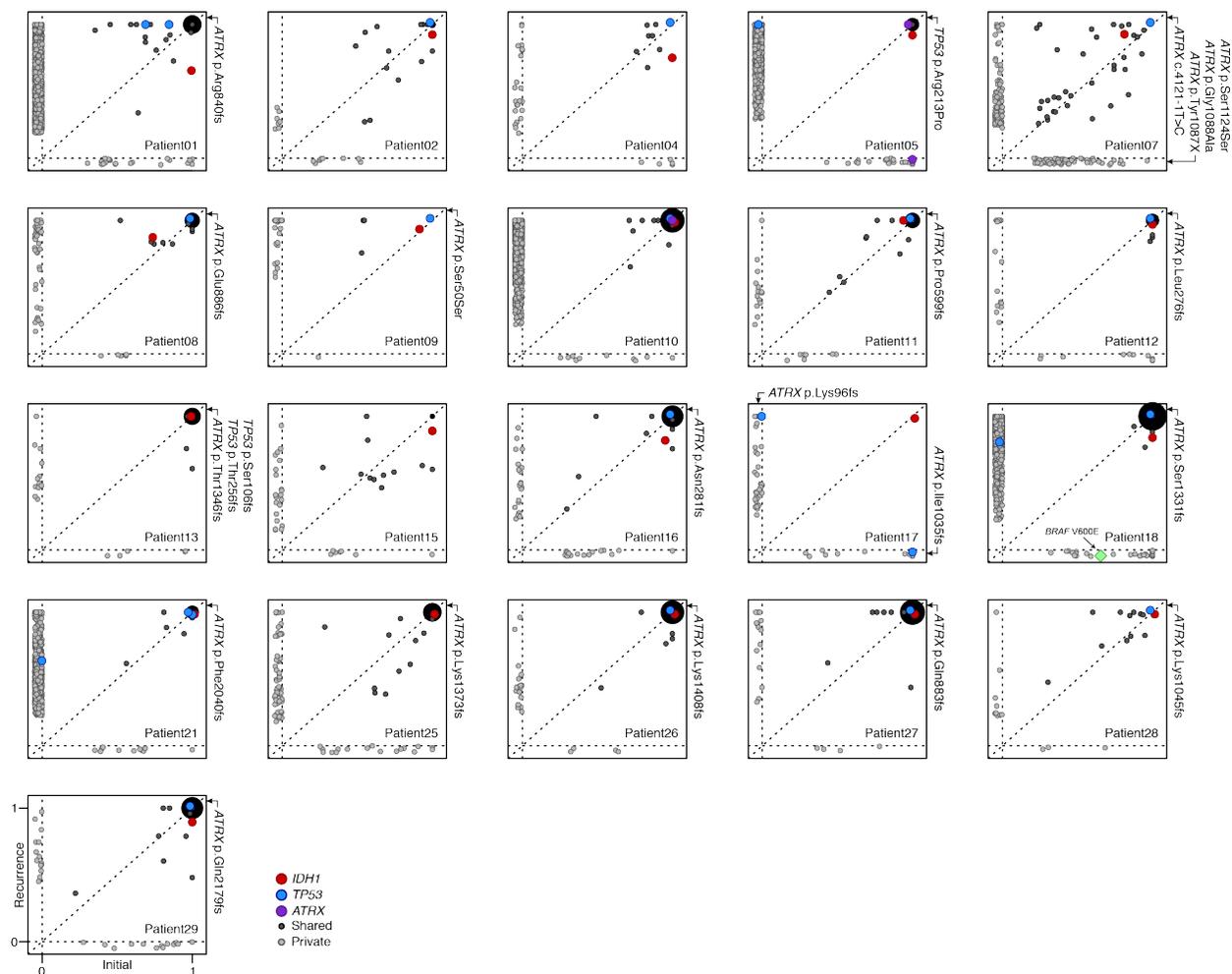


Figure 2.3. Tumor cell fraction of somatic mutations in paired initial and recurrent tumors.

Mutations private to tumors at initial diagnosis (x-axis), private to recurrent tumors (y-axis) or shared between the two tumors are shown as a function of the fraction of tumor cells containing the mutation. Those mutations clonal in both tumors are represented by a single point whose radius is scaled by the log count of such mutations. Key mutations are colored as indicated. Data from patients 06 and 24 were not available.

Figure 2.4

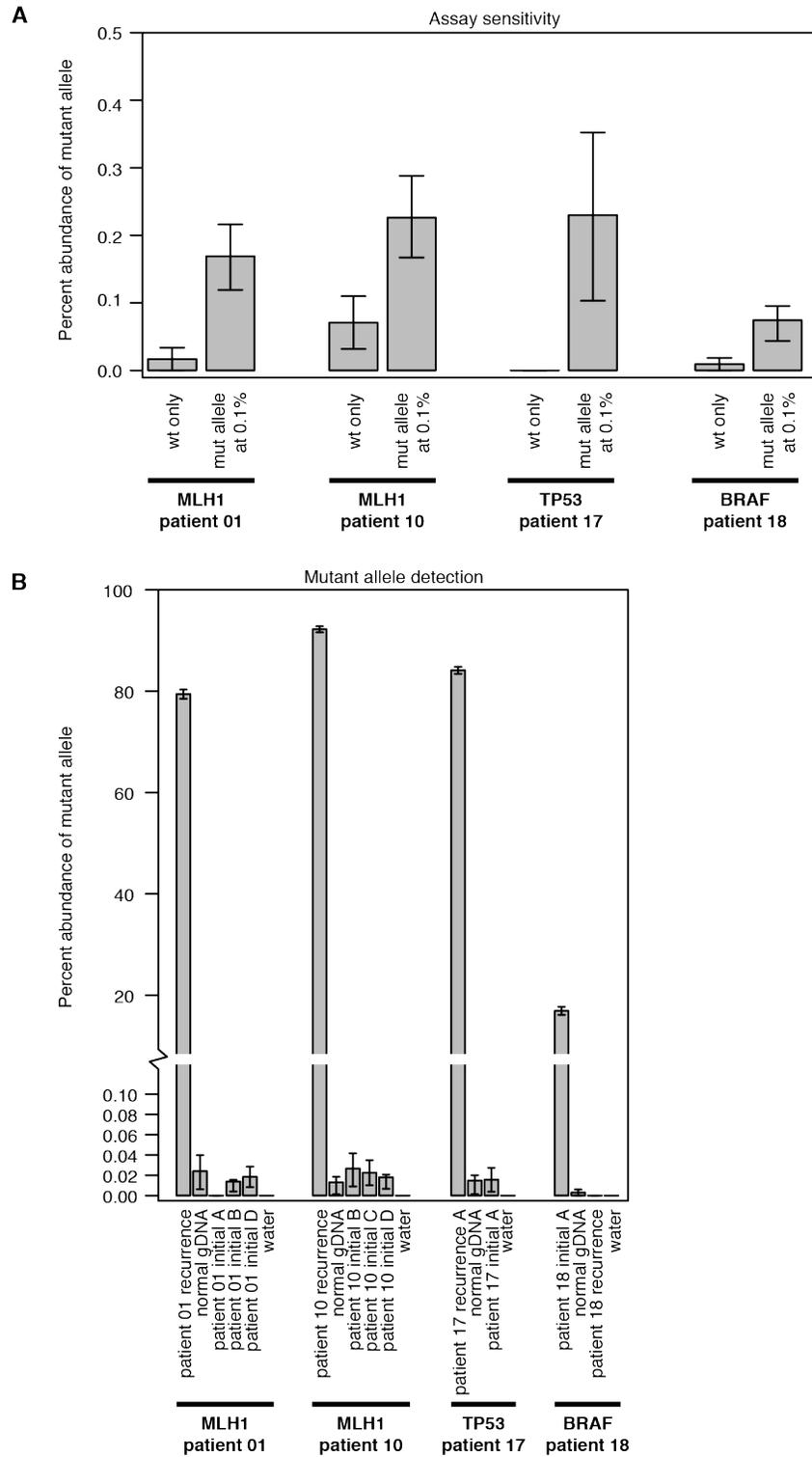


Figure 2.4. Sensitive detection of mutant alleles. (A) Droplet digital PCR assays against 4

mutations identified as private to either an initial (BRAF in patient 18) or recurrent (MLH1 patient 01, MLH1 patient 10, TP53 patient 17) tumor can detect the mutant allele down to a frequency of 0.1% (0.1% samples run in triplicate). The background positive level ranges from 0.01% to 0, allowing for sensitive detection of the mutant allele at 0.1%. **(B)** Geographically distinct samples of patient-matched tumor in which the mutation was not originally identified show no evidence of the mutant allele. Three distinct samples of the initial tumor from patients 01 and 10, one sample of the initial tumor from patient 17 and one sample of the recurrent tumor from patient 18 (all test samples run in quadruplicate), show background-level signals indicating that the mutant alleles are not present at a sensitivity of 0.1%. All error bars indicate the standard deviation from the Poisson calculation of allele concentrations.

Figure 2.5

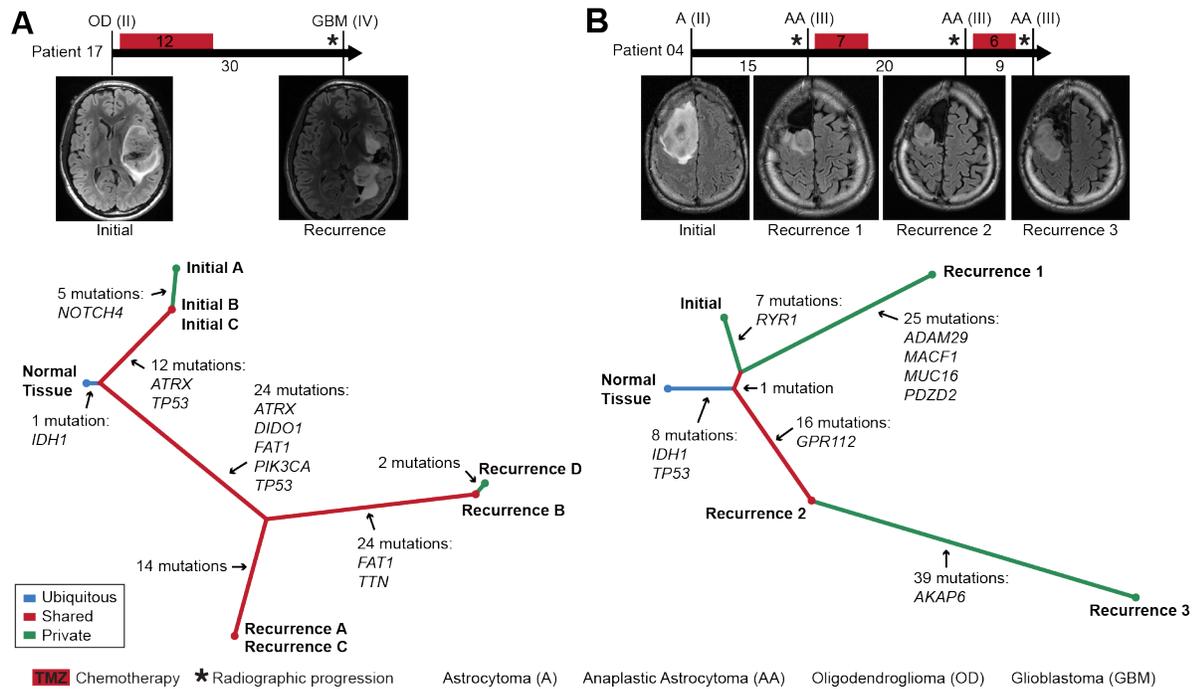


Figure 2.5. The temporal and spatial patterns of clonal evolution in two glioma patients. (A and B). A timeline of treatment histories for patient 17 (A) and patient 04 (B) (top, intervals labeled in months). Vertical bars correspond to the time of tumor resection and are labeled with the tumor diagnosis and grade. Representative MRIs are also shown. A phylogenetic tree (bottom) depicts the patterns of clonal evolution of these tumors inferred from the pattern and frequency of somatic mutations, highlighting genes frequently mutated in cancer.

Figure 2.6

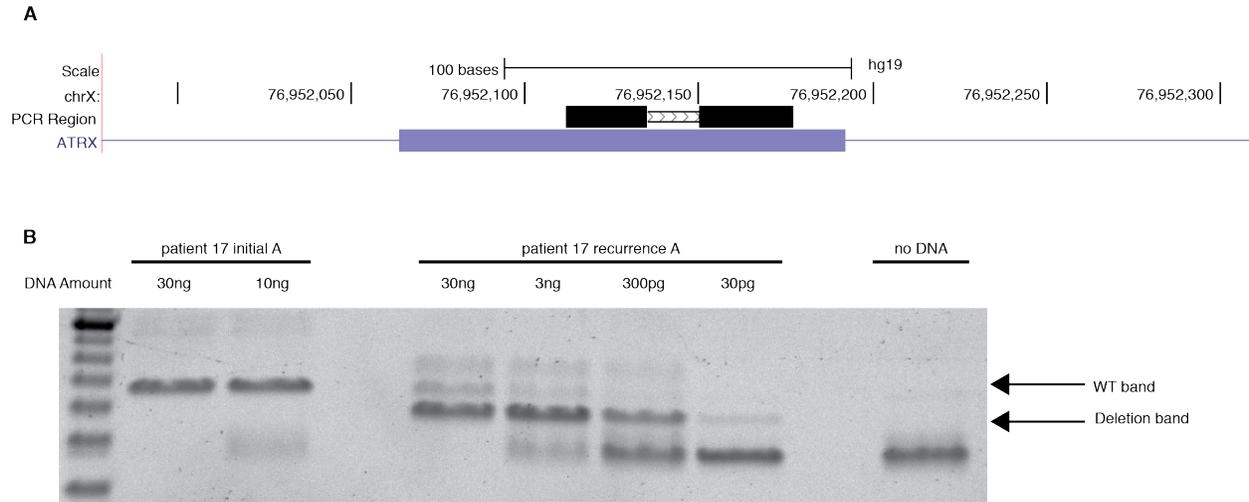


Figure 2.6. Sensitive PCR analysis of an ATRX deletion in patient 17. (A) The position of the amplicon used to assess the 8bp deletion starting at chrX:76952141 identified by exome sequencing as private to the recurrence of patient 17. (B) A PCR analysis of this region using primers that flank the deletion and allow for amplification of both the wild-type (65bp) and deletion (57bp) alleles. The PCR product corresponding to the allele with a deletion was observed in patient 17 recurrent DNA with 30ng down to 30pg of template DNA. However, no such deletion-specific PCR product was observed with 30ng or 10ng of input from the initial tumor sample.

Figure 2.7

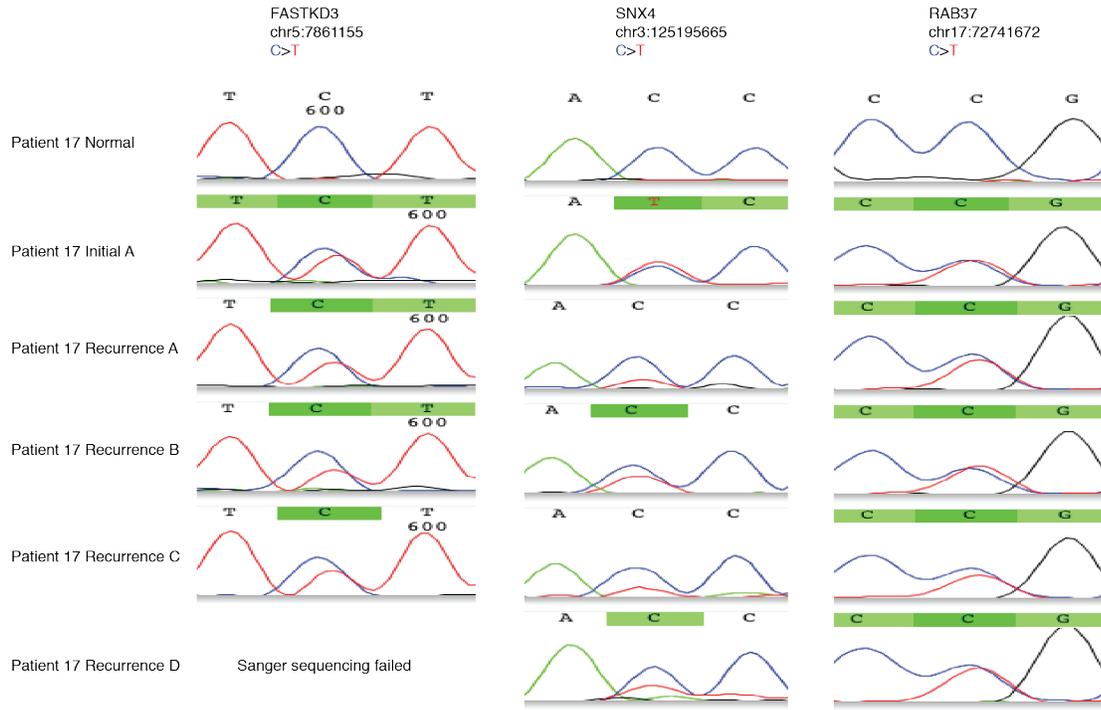


Figure 2.7. Shared non-coding mutations in patient 17. Sanger sequencing confirms that at least three non-coding mutations are shared between the initial and recurrent tumors of patient 17. This indicates that these tumors diverged early in their evolutionary history from a nearest common ancestor that includes these three non-coding mutations and the shared *IDH1* R132H coding mutation.

Figure 2.8

Patient 17

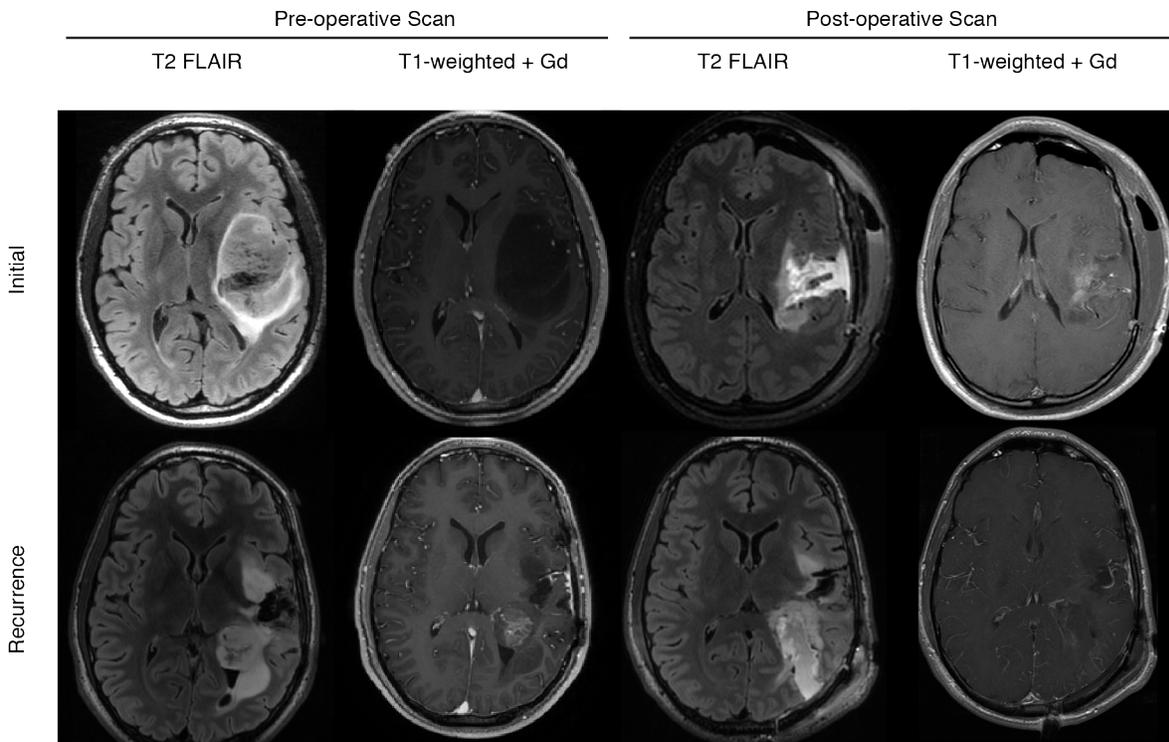


Figure 2.8. Pre- and post-surgical MR imaging of patient 17. T1- and T2-weighted images representative of the tumor region before and after the initial resection and recurrence. Imaging for the initial resection demonstrated a large T2 hyperintense non-enhancing mass situated in the left insula with significant mass effect on the left lateral ventricle. After a near gross total resection the pathological diagnosis was a WHO grade II oligodendroglioma. Significant T2-hyperintensity and tissue shifts were observed post resection. At the time of recurrence, an enhancing mass centered in the left posterior temporal and parietal white matter was noted with interval growth, avid heterogeneous enhancement, and MR features consistent with upgrade to a high-grade neoplasm. Abnormal susceptibility and T1-hyperintensity with focal reduced diffusion (not shown) in the center of the mass was consistent with central necrosis. Surrounding masslike T2 hyperintensities without contrast enhancement were seen in the left insular white

matter, extending into the left temporal lobe, left parietal lobe, and left corona radiata and were consistent with residual low-grade neoplasm. Gross total resection was performed on the enhancing portion of the left parietal mass and determined pathologically to be grade IV glioblastoma multiforme.

Figure 2.9

Patient 04

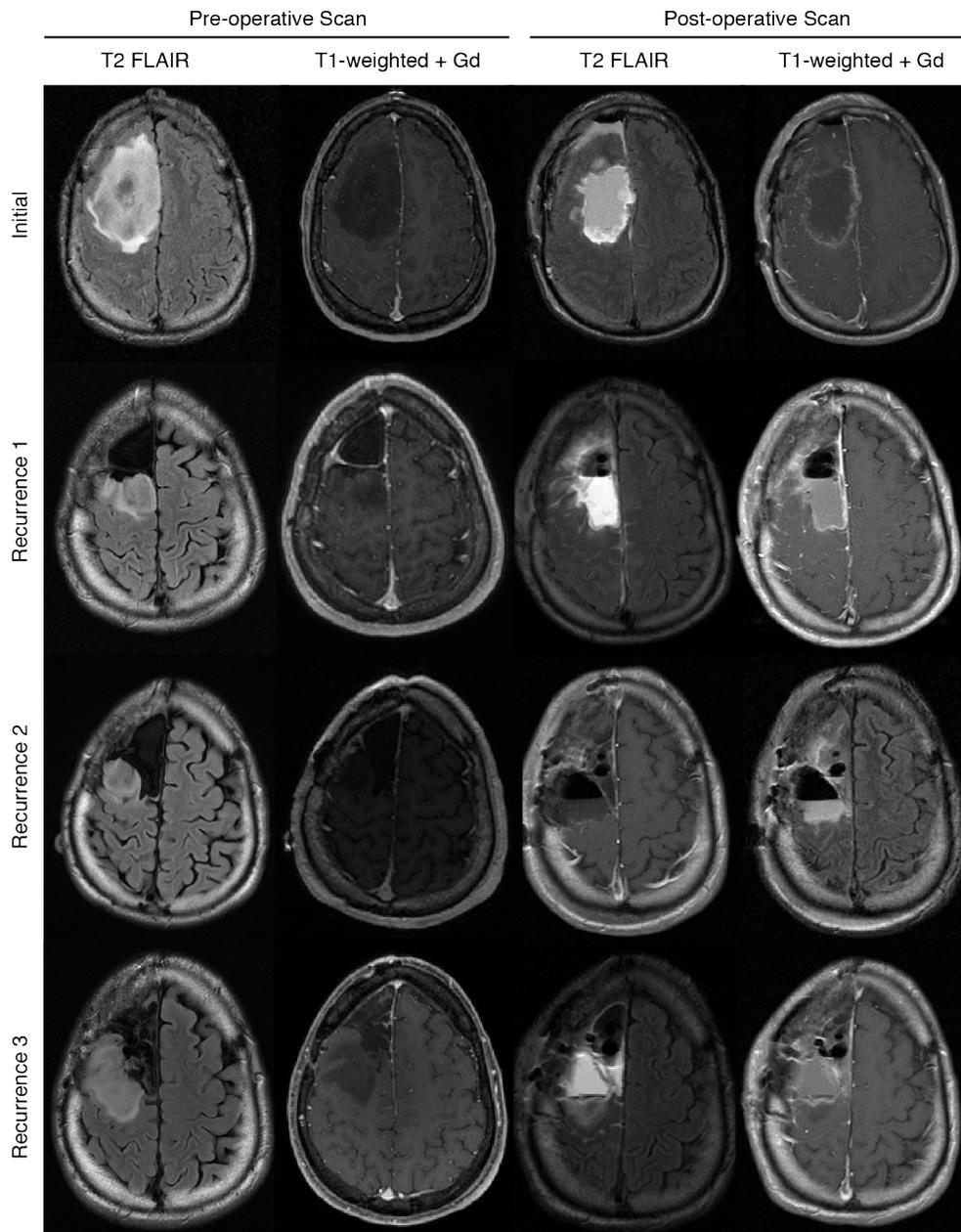


Figure 2.9. Pre- and post-surgical MR imaging of patient 04. T1- and T2-weighted images representative of the tumor region before and after the initial resection and subsequent three recurrences. Imaging for the initial resection demonstrated a large T2 hyperintense non-enhancing mass situated in the right frontal lobe with significant mass effect. Perfusion imaging

and spectroscopic imaging (not shown) were suggestive of low-grade neoplasm, confirmed after gross total resection to be a grade II astrocytoma. Upon first recurrence, an abnormal T2 lesion was observed along the posterior resection cavity, extending superiorly and inferiorly to the level of the lateral ventricles, with corresponding T1 hypointensity, characteristic of highly cellular recurrent tumor. After a gross total resection, pathology analysis indicated the tumor had upgraded to a grade III anaplastic astrocytoma. At the time of the second recurrence, a mass-like non-enhancing lesion was identified in the posterior superior lateral aspect of the resection cavity with similar imaging characteristics as the previous recurrence. A gross total resection was obtained surgically, but with limited posterior margins due to proximity with the motor tracts. The lesion remained pathologically grade III. At the time of the third recurrence, there was continued interval progression of a mass-like T2 hyperintensity within the right middle frontal and precentral gyri posterior to the surgical cavity. Signal abnormality further involved the anteromedial margins of the cavity infiltrating inferiorly into the ipsilateral corona radiata. Surgical resection was limited to 80-85% due to infiltration of the motor tracts, and the tumor tissue was confirmed to remain grade III.

Figure 2.10

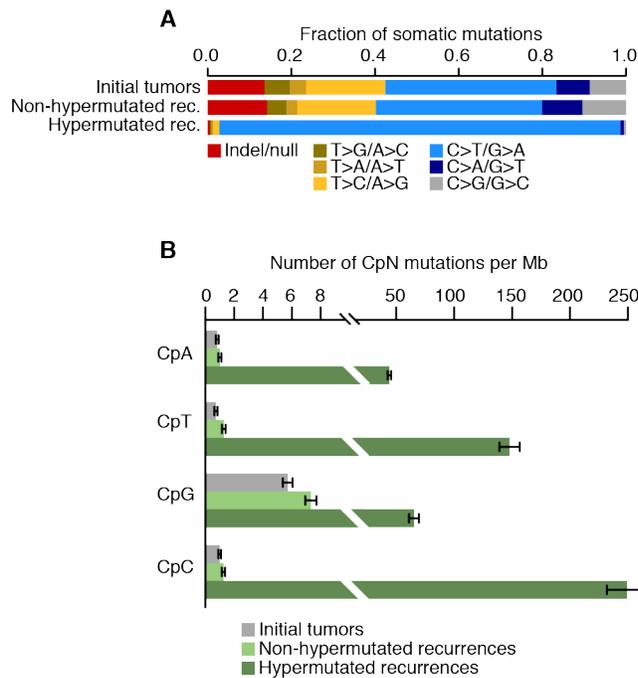


Figure 2.10. The spectrum and context of somatic mutations in hypermutated and non-hypermutated gliomas. (A) The spectrum of somatic mutation types observed in initial tumors as well as both non-hypermutated and hypermutated recurrences indicates a massive increase in the C>T/G>A mutation rate in only the latter. (B) Somatic mutation rates for each CpN dinucleotide context indicates a propensity for C>T/G>A mutations to arise outside the CpG and CpA dinucleotide contexts.

Figure 2.11

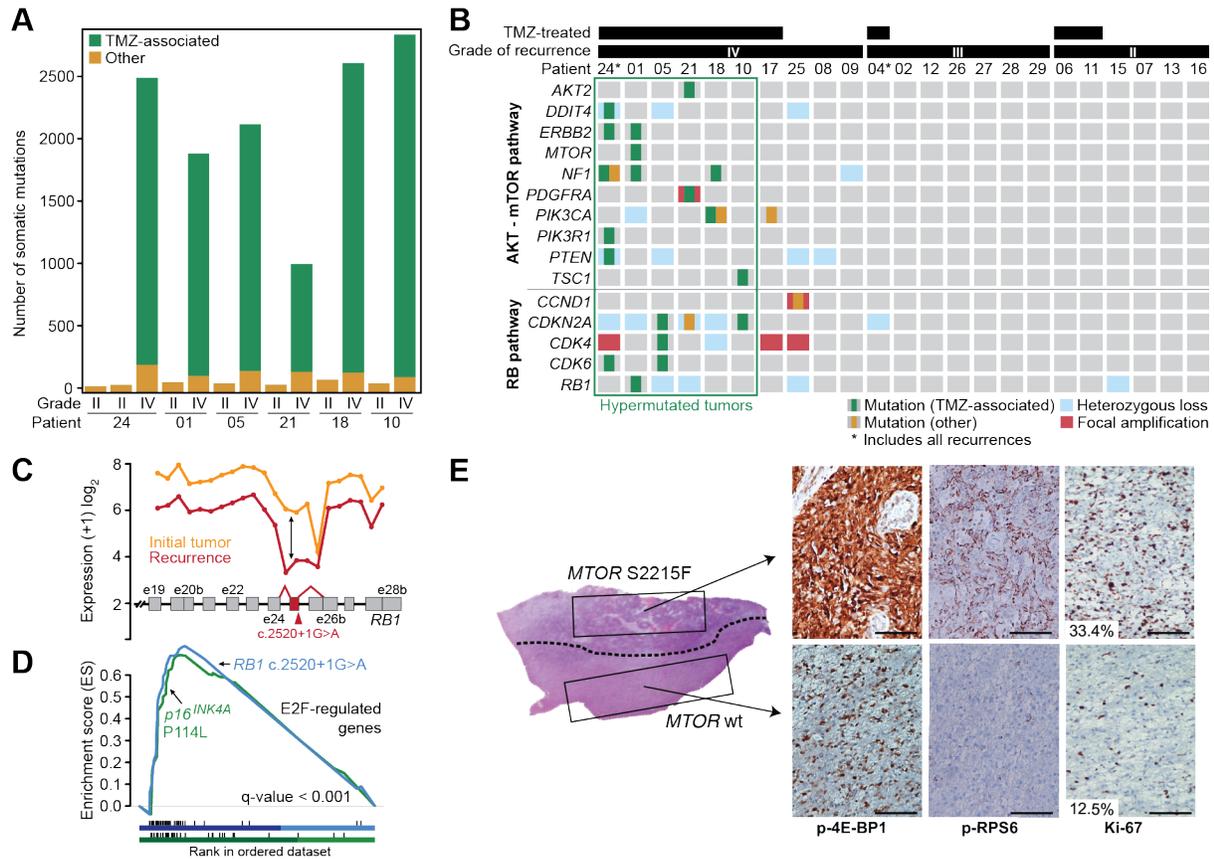


Figure 2.11. Recurrent tumors from patients treated with TMZ harbor genetic alterations in the RB and AKT-mTOR signaling pathways. (A) The number of TMZ-associated mutations and other mutations identified in the six patients with hypermutated recurrent tumors. (B) Somatic mutations and CNAs acquired upon recurrence in key genes of pathways associated with GBM. (C) Expression level of *RB1* at each exon and exon-exon junction in the initial and recurrent tumor of patient 01 showing aberrant splicing of the *RB1* transcript in the recurrent tumor harboring the *RB1* c.2520+1G>A splice-site mutation. The *RB1* exon and exon junctions with significant differential usage (red) and the location of the splice-site mutation are shown. (D) Gene set enrichment analysis shows significant enrichment of genes down-regulated by *RB1* and up-regulated by *E2F* in the recurrent tumors of patients 01 (blue) and 10 (green),

coincident with the acquisition of TMZ-associated mutations in the RB pathway. **(E)** Hematoxylin and eosin (H&E)-stained tumor sample from the first recurrent tumor of patient 01. A dotted line separates the two morphologically distinct regions. IHC for phospho-RPS6, phospho-4E-BP1 and Ki-67 show differential activation of mTORC1 targets and proliferation rates in the two adjacent regions. Bars represent 100 microns.

Figure 2.12

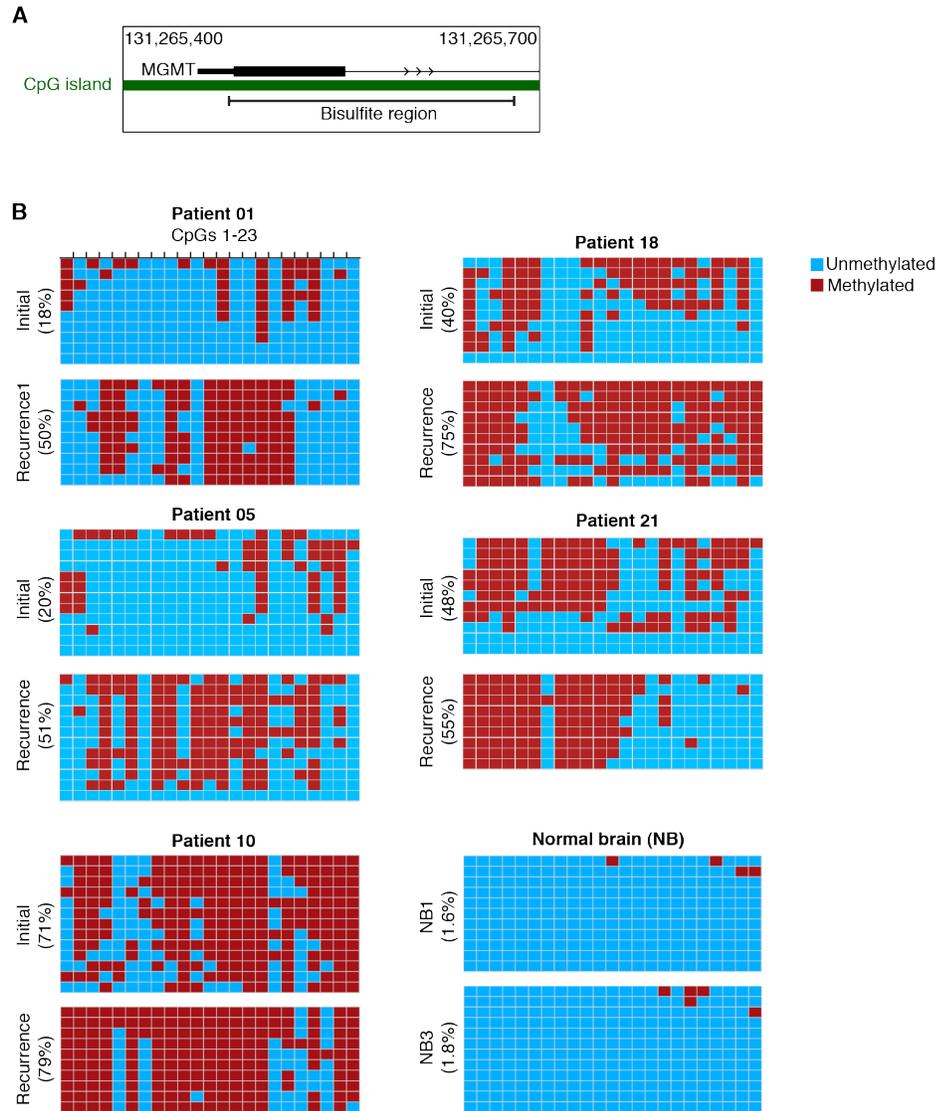


Figure 2.12. The evolution of DNA methylation affecting *MGMT* during malignant progression. (A) The position of the amplicon used to assess *MGMT* methylation levels. **(B)** Methylation status of CpG sites in individual clones of the PCR product from bisulfite-treated DNA from the initial and recurrent tumors of patients 01, 05, 10, 18, and 21 as well as two normal brain samples. Each row represents a single clone with each CpG site marked as either methylated (red) or unmethylated (blue). The total methylation percentage of all clones is presented to the left of each panel.

Figure 2.13

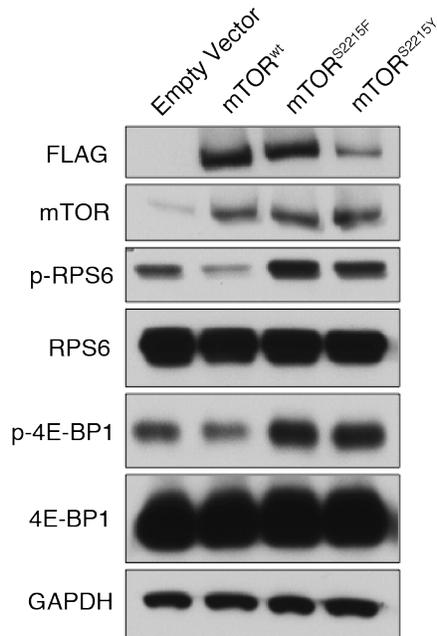


Figure 2.13. Functional assessment of *MTOR* mutations on mTORC1 signaling. Western blot on protein from HEK293 cells expressing wild-type or one of two mutant mTOR vectors. Constitutive phosphorylation of RPS6 and 4E-BP1 indicates that the *MTOR* S2215F mutation constitutively activates mTORC1 signaling, much like the previously characterized *MTOR* S2215Y mutation.

Figure 2.14

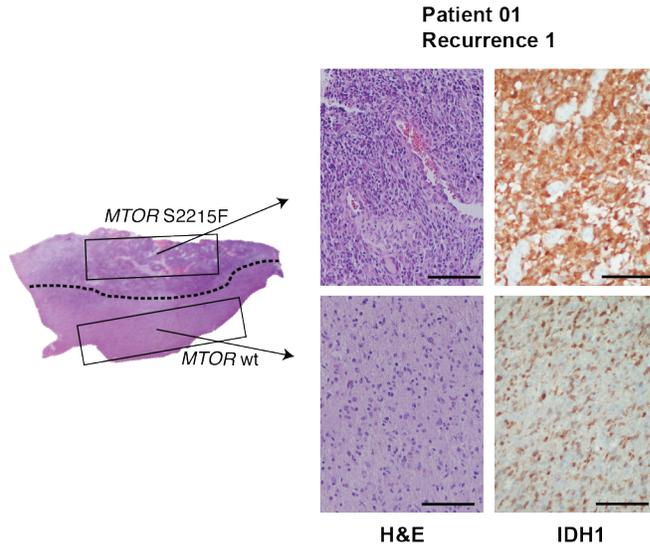


Figure 2.14. Comparison of tumor samples from the first recurrence of patient 01.

Hematoxylin and eosin (H&E)-stained tumor sample from the first recurrence of patient 01. A dotted line separates the two morphologically distinct regions. H&E-stained sections indicate that both regions are histologically GBM. IHC for mutant IDH1 in these same regions indicates comparable tumor content. Bars represent 100 microns.

Figure 2.15

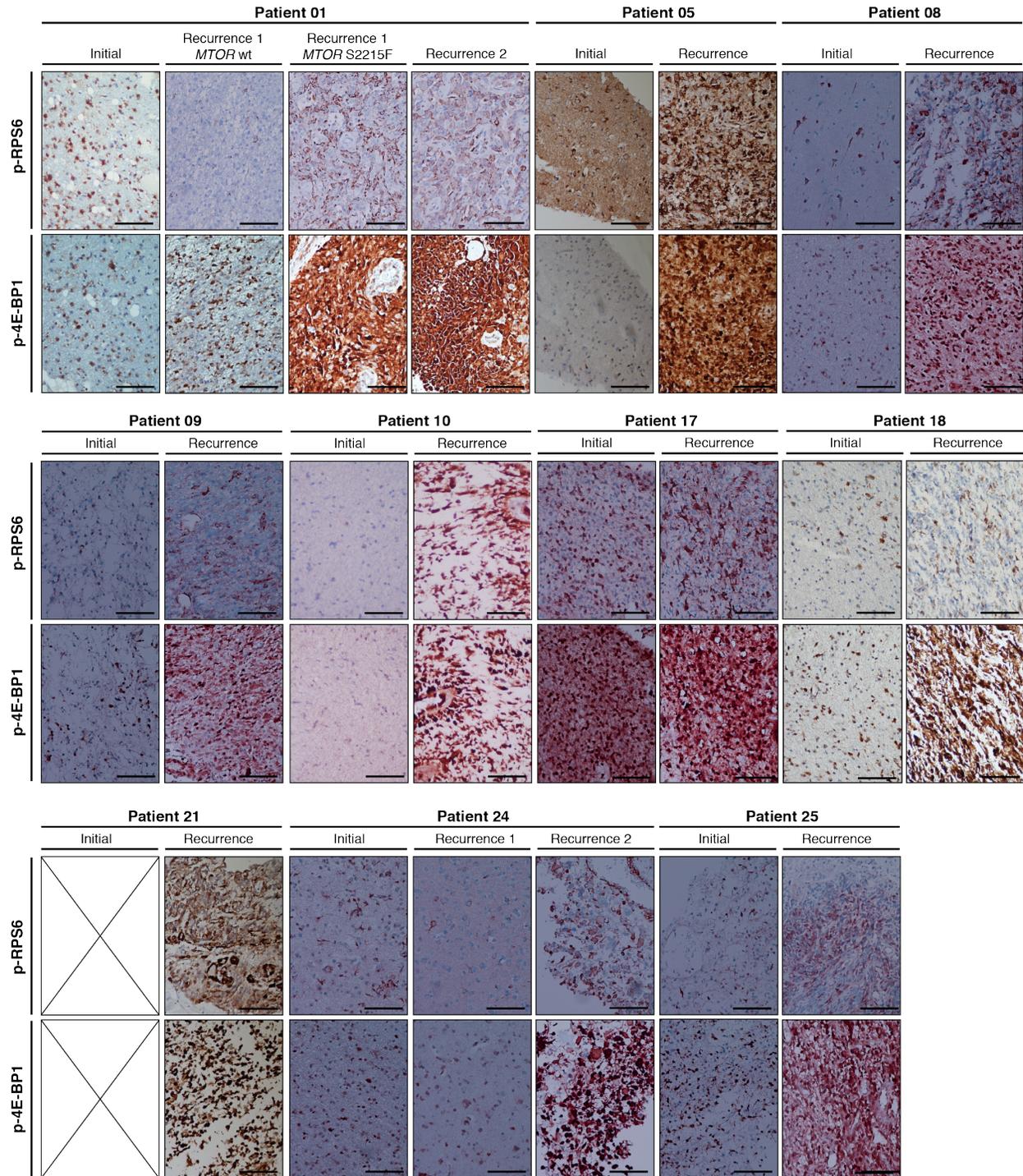


Figure 2.15. mTORC1 signaling in tumors at initial diagnosis and their GBM recurrences.

Immunohistochemistry (IHC) for phospho-RPS6 and phospho-4E-BP1 in the initial and

recurrent tumors of all patients that undergo malignant progression to GBM indicates an increase in mTORC1 signaling across GBMs relative to the patient-matched initial tumors. Slides for the initial tumor from patient 21 were not available. Bars in all panels represent 100 microns.

Figure 2.16

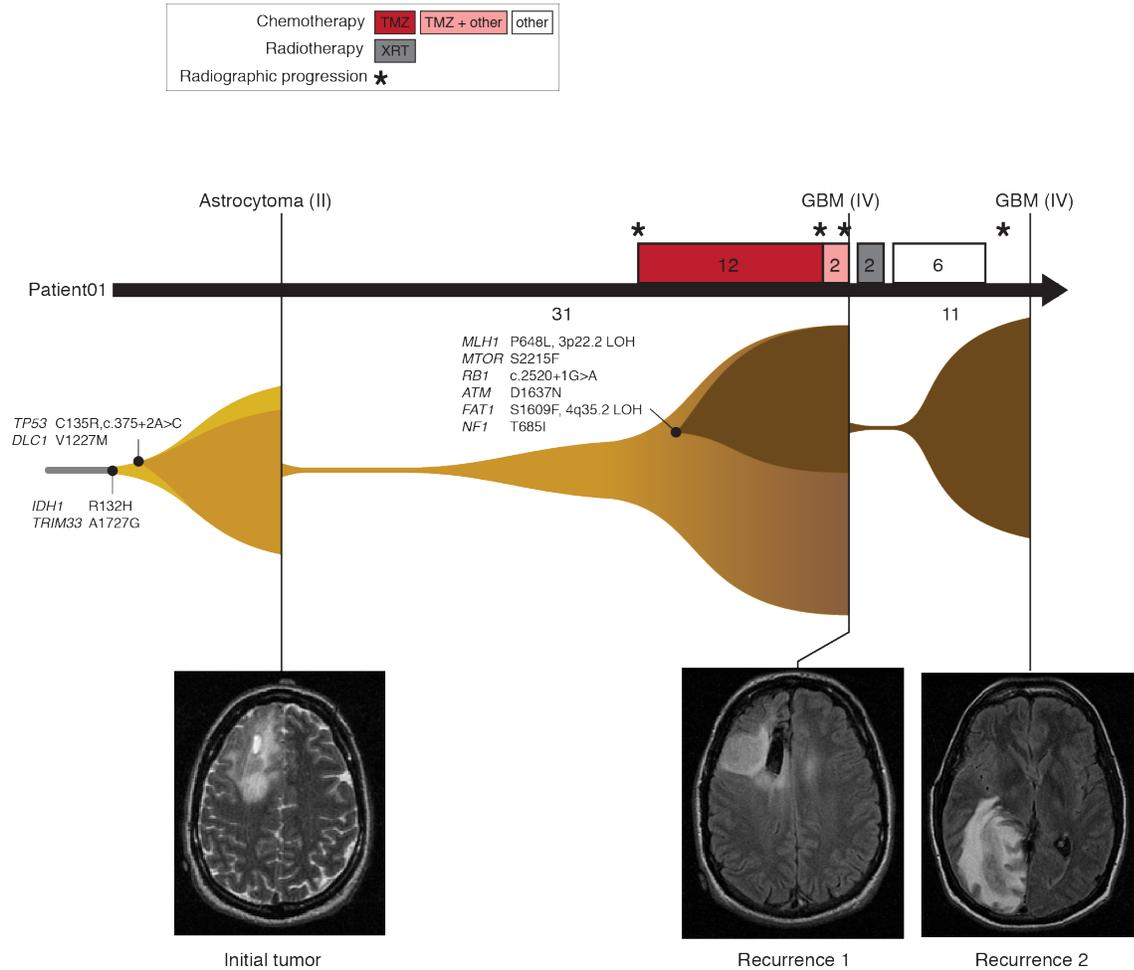


Figure 2.16. An integrated timeline of the treatment, imaging, and clonal evolution of a low-grade astrocytoma that underwent TMZ-associated malignant progression. A timeline of the treatment received by patient 01 (all intervals labeled in months). Vertical bars represent tumor resections and are labeled with their diagnosis and grade (top), as well as representative MRIs (bottom). A graphical representation of one model of the clonal evolution of these tumors (middle) begins with a founding clone with early mutations in *IDH1*, *TRIM33*, *TP53* and *DLC1*. Upon first recurrence, two morphologically distinct regions of GBM are present, with only one region harboring TMZ-associated mutations in key functional cancer genes. This hypermutated clone then seeds the distal second recurrence.

Figure 2.17

Patient 01

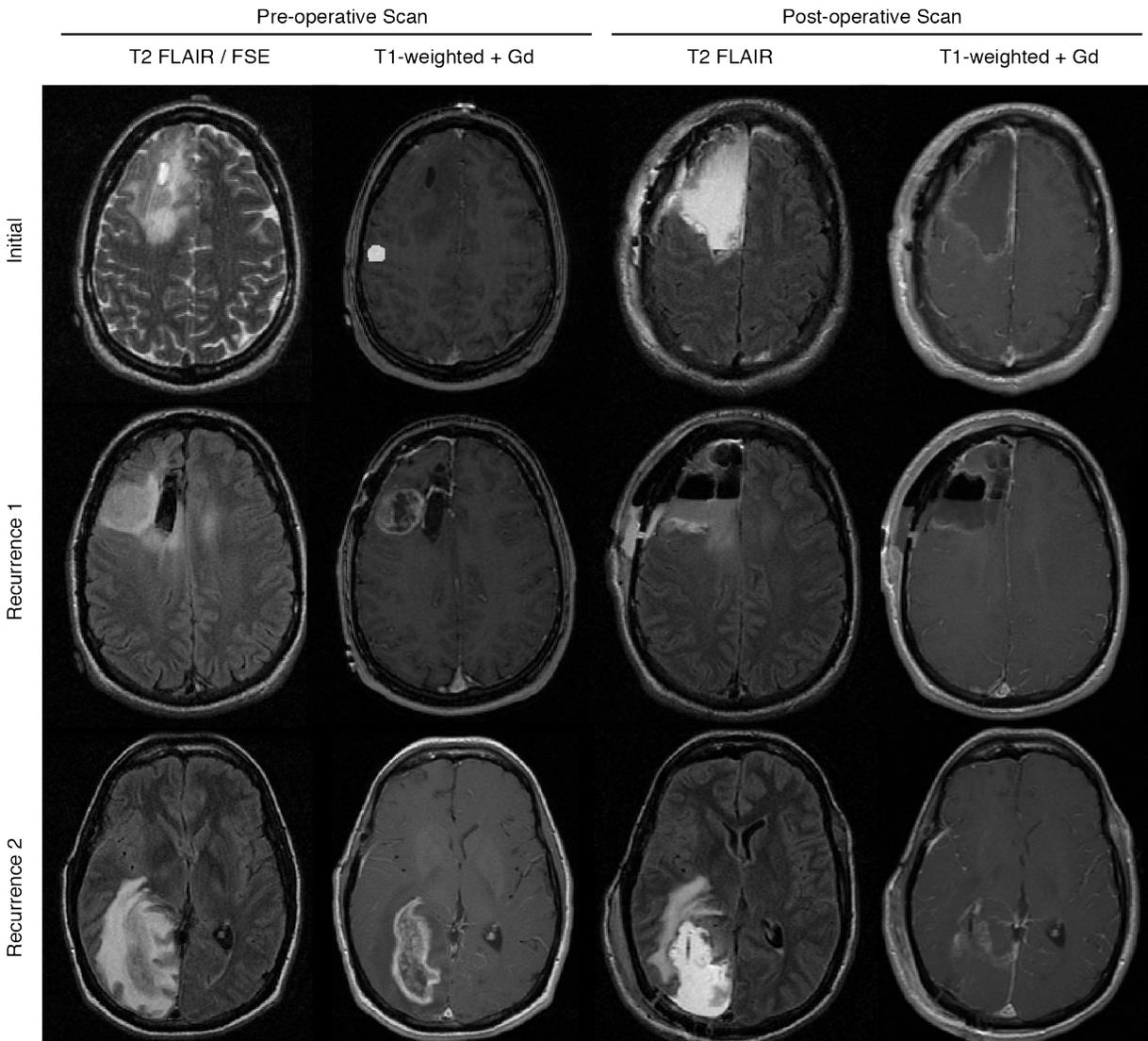


Figure 2.17. Pre- and post-surgical magnetic resonance imaging of patient 01. T1- and T2-weighted images representative of the tumor region throughout the course of treatment. At the time of initial resection, the tumor lesion appeared localized to the right frontal lobe in the pre-operative scan. The first tumor recurrence was at the posterior aspect of the initial resection site, and the second recurrence was found distally in the right temporal occipital region in a lower

horizontal plane. The hyper-intense region in the initial resection, pre-surgical spoiled gradient echo (SPGR), is a diffusion tensor imaging mask used for white matter tracking during surgery.

Table 2.1

Patient	Gender	Age at diagnosis	Tumor sample	Diagnosis (WHO grade)	Surgical interval (months)	Non-surgical treatment (months)	Overall survival (months)	MGMT methylation status	IDH1 status	1p/19q status	TP53, ATRX, CIC, FUBP1*
01	Male	28	Initial tumor	Astrocytoma (II)	31	TMZ (14)†	58	Unmethylated	R132H	Intact	TP53, ATRX
			Recurrence	Glioblastoma (IV)				Methylated	R132H	19q deletion	TP53, ATRX
02	Female	26	Initial tumor	Oligoastrocytoma (II)	74	None	5	Methylated	R132H	19q deletion	TP53
			Recurrence	Anaplastic astrocytoma (III)				Unmethylated	R132H	19q deletion	TP53
			Initial tumor	Astrocytoma (II)	15	None		Unknown	R132C	Intact	TP53
04	Male	22	Recurrence 1	Anaplastic astrocytoma (III)	20	TMZ (7)	61	Unknown	R132C	Intact	TP53
			Recurrence 2	Anaplastic astrocytoma (III)	9	TMZ (6)		Unknown	R132C	Intact	TP53
			Recurrence 3	Anaplastic astrocytoma (III)				Unknown	R132C	Intact	TP53
05	Female	39	Initial tumor	Astrocytoma (II)	56	TMZ (12)	74	Methylated	R132H	Intact	TP53, ATRX
			Recurrence	Glioblastoma (IV)				Methylated	R132H	Intact	TP53, ATRX
06	Male	22	Initial tumor	Astrocytoma (II)	32	TMZ (6) §	87	Unmethylated	R132H	Intact	TP53, ATRX
			Recurrence	Astrocytoma (II)				Methylated	R132H	Intact	TP53, ATRX
07	Male	30	Initial tumor	Astrocytoma (II)	105	XRT (1)	148	Unknown	R132H	Intact	TP53, ATRX
			Recurrence	Astrocytoma (II)				Unknown	R132H	Intact	TP53, ATRX
08	Male	44	Initial tumor	Oligoastrocytoma (II)	40	None	103	Unknown	R132H	Intact	TP53, ATRX
			Recurrence	Glioblastoma (IV)				Unknown	R132H	19q deletion	TP53, ATRX
09	Male	28	Initial tumor	Astrocytoma (II)	35	XRT (1)	45	Methylated	R132H	Intact	TP53
			Recurrence	Glioblastoma (IV)				Methylated	R132H	Intact	TP53
10	Female	41	Initial tumor	Astrocytoma (II)	25	TMZ (9)	44	Methylated	R132H	Intact	TP53, ATRX
			Recurrence	Glioblastoma (IV)				Methylated	R132H	Intact	TP53, ATRX
11	Female	30	Initial tumor	Oligoastrocytoma (II)	132	XRT (1), TMZ (26)	186	Methylated	R132H	Intact	TP53, ATRX
			Recurrence	Oligoastrocytoma (II)				Methylated	R132H	Intact	TP53, ATRX
12	Male	35	Initial tumor	Astrocytoma (II)	17	None	82 ¶	Unknown	R132H	Intact	TP53, ATRX
			Recurrence	Anaplastic astrocytoma (III)				Unknown	R132H	Intact	TP53, ATRX
13	Male	24	Initial tumor	Oligoastrocytoma (II)	21	None	106	Unmethylated	R132G	Intact	TP53, ATRX
			Recurrence	Oligoastrocytoma (II)				Unmethylated	R132G	Intact	TP53, ATRX
15	Female	38	Initial tumor	Oligoastrocytoma (II)	119	None	143 ¶	Unknown	R132H	Co-deletion	CIC
			Recurrence	Oligoastrocytoma (II)				Unknown	R132H	Co-deletion	CIC
16	Female	35	Initial tumor	Astrocytoma (II)	5 †	None	38	Methylated	R132H	Intact	TP53, ATRX
			Recurrence	Astrocytoma (II)				Methylated	R132H	Intact	TP53, ATRX
17	Male	27	Initial tumor	Oligodendroglioma (II)	30	TMZ (12)	59 ¶	Methylated	R132H	Intact	TP53, ATRX
			Recurrence	Glioblastoma (IV)				Unmethylated	R132H	Intact	TP53, ATRX
18	Male	49	Initial tumor	Oligoastrocytoma (II)	94	TMZ (11)	106 ¶	Methylated	R132H	Intact	TP53, ATRX
			Recurrence	Glioblastoma (IV)				Methylated	R132H	Intact	TP53, ATRX
21	Male	27	Initial tumor	Oligoastrocytoma (II)	21	TMZ (12)	35	Methylated	R132H	19q deletion	TP53, ATRX
			Recurrence	Glioblastoma (IV)				Methylated	R132H	Intact	TP53, ATRX
24	Male	34	Initial tumor	Astrocytoma (II)	88	XRT (1)	109	Unmethylated	R132H	Intact	TP53, ATRX
			Recurrence 1	Astrocytoma (II)	14	TMZ (12)		Unmethylated	R132H	Intact	TP53, ATRX
			Recurrence 2	Glioblastoma (IV)				Methylated	R132H	Intact	TP53, ATRX
25	Male	39	Initial tumor	Astrocytoma (II)	14	None	42 ¶	Methylated	R132H	Intact	ATRX
			Recurrence	Glioblastoma (IV)				Methylated	R132H	Intact	ATRX
26	Female	36	Initial tumor	Astrocytoma (II)	15	None	47 ¶	Methylated	R132H	Intact	TP53, ATRX
			Recurrence	Anaplastic astrocytoma (III)				Methylated	R132H	Intact	TP53, ATRX
27	Male	56	Initial tumor	Astrocytoma (II)	12	None	54 ¶	Methylated	R132H	Intact	TP53, ATRX
			Recurrence	Anaplastic astrocytoma (III)				Methylated	R132H	Intact	TP53, ATRX
28	Male	30	Initial tumor	Astrocytoma (II)	85	None	137 ¶	Methylated	R132H	19q deletion	TP53, ATRX
			Recurrence	Anaplastic astrocytoma (III)				Methylated	R132H	19q deletion	TP53, ATRX
29	Male	52	Initial tumor	Astrocytoma (II)	57	None	109	Methylated	R132H	Intact	TP53, ATRX
			Recurrence	Anaplastic astrocytoma (III)				Methylated	R132H	Intact	TP53, ATRX

* Non-silent mutations
† Recurrent surgery for residual disease, no evidence of radiographic progression
‡ Including a month each of TMZ plus either Accutane or Thalidomide
§ Patient received TMZ prior to initial resection only
|| Patient lost to follow-up
¶ Patient alive

Table 2.1. Summary of the clinical and molecular features of each tumor in the cohort.

Table 2.2

Patient ID	Sample Type	Aligned Reads	Percent Duplication	Bait Set	PF Unique Reads Aligned	Mean Target Coverage	Percent Targets Zero Coverage	Percent Target Bases 30X Coverage
Patient01	Normal	164189752	0.5194	Agilent_SureSelect_Human_All_Exon_50mb	73268999	55.18	0.0440	0.6527
Patient01	Initial	123191094	0.3604	Agilent_SureSelect_Human_All_Exon_50mb	71631056	55.03	0.0406	0.6755
Patient01	Recurrence	152187820	0.3973	Agilent_SureSelect_Human_All_Exon_50mb	84006557	65.83	0.0380	0.7273
Patient02	Normal	155497580	0.2985	Agilent_SureSelect_Human_All_Exon_50mb	100712175	64.57	0.0380	0.7154
Patient02	Initial	130362234	0.4392	Agilent_SureSelect_Human_All_Exon_50mb	67316662	52.85	0.0473	0.6254
Patient02	Recurrence	175262920	0.3902	Agilent_SureSelect_Human_All_Exon_50mb	98664035	79.45	0.0382	0.7672
Patient04	Normal	369017302	0.6461	Agilent_SureSelect_Human_All_Exon_50mb	122144743	87.52	0.0294	0.8551
Patient04	Initial	423149028	0.6231	Agilent_SureSelect_Human_All_Exon_50mb	149033959	108.56	0.0292	0.8670
Patient04	Recurrence1	388829972	0.4979	Agilent_SureSelect_Human_All_Exon_50mb	179237717	127.44	0.0297	0.8621
Patient04	Recurrence2	422622590	0.7246	Agilent_SureSelect_Human_All_Exon_50mb	109327464	76.52	0.0329	0.8009
Patient04	Recurrence3	411818172	0.5571	Agilent_SureSelect_Human_All_Exon_50mb	169340111	115.41	0.0289	0.8573
Patient05	Normal	105960550	0.2563	Agilent_SureSelect_Human_All_Exon_50mb	72679670	62.00	0.0408	0.7256
Patient05	Initial	115647024	0.4357	Agilent_SureSelect_Human_All_Exon_50mb	59514886	48.41	0.0734	0.5210
Patient05	Recurrence	209762250	0.3547	Agilent_SureSelect_Human_All_Exon_50mb	123714928	105.95	0.0287	0.8671
Patient06	Normal	122123454	0.4186	Agilent_SureSelect_Human_All_Exon_50mb	65149273	52.82	0.0576	0.5902
Patient06	Initial	126528998	0.5423	Agilent_SureSelect_Human_All_Exon_50mb	53462276	45.65	0.0762	0.5166
Patient06	Recurrence	112361636	0.4096	Agilent_SureSelect_Human_All_Exon_50mb	60725056	50.79	0.0655	0.5474
Patient07	Normal	400920082	0.6515	Agilent_SureSelect_Human_All_Exon_50mb	130504743	92.45	0.0296	0.8525
Patient07	Initial	398488840	0.4811	Agilent_SureSelect_Human_All_Exon_50mb	186969969	147.14	0.0271	0.8916
Patient07	Recurrence	397446092	0.5190	Agilent_SureSelect_Human_All_Exon_50mb	172359540	134.21	0.0287	0.8782
Patient08	Normal	378004132	0.5802	Agilent_SureSelect_Human_All_Exon_50mb	148134183	112.71	0.0285	0.8668
Patient08	Initial	348591182	0.7179	Agilent_SureSelect_Human_All_Exon_50mb	91902268	64.63	0.0305	0.8092
Patient08	Recurrence	384059252	0.5680	Agilent_SureSelect_Human_All_Exon_50mb	155227207	104.78	0.0298	0.8414
Patient09	Normal	265686178	0.6010	Agilent_SureSelect_Human_All_Exon_50mb	99001945	76.37	0.0318	0.8170
Patient09	Initial	368731284	0.7178	Agilent_SureSelect_Human_All_Exon_50mb	97645482	67.45	0.0310	0.7994
Patient09	Recurrence	446906980	0.5909	Agilent_SureSelect_Human_All_Exon_50mb	171222128	112.72	0.0275	0.8765
Patient10	Normal	386974478	0.5749	Agilent_SureSelect_Human_All_Exon_50mb	153915240	117.04	0.0303	0.8779
Patient10	Initial	363569212	0.7389	Agilent_SureSelect_Human_All_Exon_50mb	89113474	56.05	0.0338	0.7617
Patient10	Recurrence	349262384	0.5322	Agilent_SureSelect_Human_All_Exon_50mb	152272606	119.87	0.0302	0.8764
Patient11	Normal	332902458	0.5519	Agilent_SureSelect_Human_All_Exon_50mb	139043307	105.76	0.0325	0.8423
Patient11	Initial	437288102	0.6113	Agilent_SureSelect_Human_All_Exon_50mb	160052465	104.97	0.0315	0.8595
Patient11	Recurrence	464429120	0.5988	Agilent_SureSelect_Human_All_Exon_50mb	174826967	111.94	0.0292	0.8671
Patient12	Normal	399729530	0.5896	Agilent_SureSelect_Human_All_Exon_50mb	149009736	109.53	0.0278	0.8735
Patient12	Initial	404170618	0.5052	Agilent_SureSelect_Human_All_Exon_50mb	181316672	141.54	0.0407	0.8475
Patient12	Recurrence	396288996	0.5074	Agilent_SureSelect_Human_All_Exon_50mb	176766260	143.55	0.0282	0.8854
Patient13	Normal	428823802	0.6245	Agilent_SureSelect_Human_All_Exon_50mb	149462438	112.50	0.0279	0.8851
Patient13	Initial	375900864	0.5671	Agilent_SureSelect_Human_All_Exon_50mb	151991452	109.46	0.0290	0.8614
Patient13	Recurrence	395698014	0.6377	Agilent_SureSelect_Human_All_Exon_50mb	133988496	102.04	0.0296	0.8706
Patient15	Normal	356565546	0.6944	Agilent_SureSelect_Human_All_Exon_50mb	101923977	70.57	0.0322	0.8276
Patient15	Initial	361842770	0.4976	Agilent_SureSelect_Human_All_Exon_50mb	170167006	122.04	0.0296	0.8732
Patient15	Recurrence	324249284	0.6581	Agilent_SureSelect_Human_All_Exon_50mb	103127635	83.20	0.0334	0.8169
Patient16	Normal	413818772	0.6015	Agilent_SureSelect_Human_All_Exon_50mb	153920547	103.69	0.0442	0.8409
Patient16	Initial	366868764	0.6760	Agilent_SureSelect_Human_All_Exon_50mb	111294559	72.80	0.0323	0.8339
Patient16	Recurrence	374388258	0.6902	Agilent_SureSelect_Human_All_Exon_50mb	108358244	75.66	0.0356	0.7973
Patient17	Normal	410213150	0.6211	Agilent_SureSelect_Human_All_Exon_50mb	143137416	120.30	0.0294	0.8637
Patient17	Initial A	432939976	0.5020	Agilent_SureSelect_Human_All_Exon_50mb	197954550	149.50	0.0283	0.8827
Patient17	Initial B	282682304	0.3351	Agilent_SureSelect_Human_All_Exon_50mb	172706817	155.02	0.0294	0.8779
Patient17	Initial C	262086408	0.3354	Agilent_SureSelect_Human_All_Exon_50mb	160023188	147.40	0.0299	0.8732
Patient17	Recurrence A	434106670	0.5463	Agilent_SureSelect_Human_All_Exon_50mb	181195751	151.37	0.0287	0.8783
Patient17	Recurrence B	221369134	0.2609	Agilent_SureSelect_Human_All_Exon_50mb	149300799	132.72	0.0286	0.8682
Patient17	Recurrence C	212469564	0.1887	Agilent_SureSelect_Human_All_Exon_50mb	157206847	124.38	0.0277	0.8589
Patient17	Recurrence D	192665284	0.1457	Agilent_SureSelect_Human_All_Exon_50mb	150388708	125.24	0.0283	0.8581
Patient18	Normal	377796074	0.5557	Agilent_SureSelect_Human_All_Exon_50mb	156092218	135.47	0.0294	0.8792
Patient18	Initial A	431458900	0.6291	Agilent_SureSelect_Human_All_Exon_50mb	148164517	120.08	0.0290	0.8652
Patient18	Initial B	183170435	0.1739	Agilent_SureSelect_Human_All_Exon_50mb	139028086	125.30	0.0303	0.8553
Patient18	Initial C	222901260	0.2239	Agilent_SureSelect_Human_All_Exon_50mb	159192535	143.28	0.0302	0.8658
Patient18	Initial D	224755464	0.1387	Agilent_SureSelect_Human_All_Exon_50mb	178164963	162.41	0.0296	0.8725
Patient18	Recurrence	413777728	0.5623	Agilent_SureSelect_Human_All_Exon_50mb	166488017	147.02	0.0288	0.8787
Patient21	Normal	236557208	0.1435	Agilent_SureSelect_Human_All_Exon_50mb	185759526	157.45	0.0272	0.8842
Patient21	Initial	232154134	0.3955	Agilent_SureSelect_Human_All_Exon_50mb	128220620	116.32	0.0299	0.8482
Patient21	Recurrence	219025678	0.1963	Agilent_SureSelect_Human_All_Exon_50mb	160524299	142.22	0.0294	0.8671

Patient ID	Sample Type	Aligned Reads	Percent Duplication	Bait Set*	PF Unique Reads Aligned	Mean Target Coverage	Percent Targets Zero Coverage	Percent Target Bases 30X Coverage
Patient24	Normal	291504104	0.3240	Agilent_SureSelect_Human_All_Exon_50mb	194079378	181.33	0.0511	0.8545
Patient24	Initial	222865850	0.2199	Agilent_SureSelect_Human_All_Exon_50mb	171015094	160.38	0.0509	0.8418
Patient24	Recurrence1	239585468	0.1378	Agilent_SureSelect_Human_All_Exon_50mb	203306203	188.47	0.0478	0.8606
Patient24	Recurrence2	206208472	0.2346	Agilent_SureSelect_Human_All_Exon_50mb	155461755	146.81	0.0537	0.8260
Patient25	Normal	263944886	0.1502	Agilent_SureSelect_Human_All_Exon_50mb	220616782	198.89	0.0440	0.8644
Patient25	Initial	272637310	0.1571	Agilent_SureSelect_Human_All_Exon_50mb	226300494	208.45	0.0459	0.8688
Patient25	Recurrence	278331906	0.1945	Agilent_SureSelect_Human_All_Exon_50mb	220670256	205.99	0.0478	0.8532
Patient26	Normal	235450112	0.2639	Agilent_SureSelect_Human_All_Exon_50mb	170904446	164.20	0.0564	0.8501
Patient26	Initial A	252423160	0.2216	Agilent_SureSelect_Human_All_Exon_50mb	192962251	176.03	0.0483	0.8540
Patient26	Initial B	202072954	0.0905	Agilent_SureSelect_Human_All_Exon_50mb	182046280	156.63	0.0529	0.8370
Patient26	Recurrence A	297876622	0.1722	Agilent_SureSelect_Human_All_Exon_50mb	243072156	230.40	0.0492	0.8753
Patient26	Recurrence B	146455804	0.0974	Agilent_SureSelect_Human_All_Exon_50mb	130989039	114.68	0.0592	0.7855
Patient27	Normal	276349302	0.2756	Agilent_SureSelect_Human_All_Exon_50mb	197384884	192.76	0.0527	0.8554
Patient27	Initial A	274051470	0.2790	Agilent_SureSelect_Human_All_Exon_50mb	194508972	183.82	0.0514	0.8583
Patient27	Initial B	129212828	0.0788	Agilent_SureSelect_Human_All_Exon_50mb	117837436	103.57	0.0582	0.7699
Patient27	Recurrence A	237800232	0.2553	Agilent_SureSelect_Human_All_Exon_50mb	174652244	169.53	0.0546	0.8445
Patient27	Recurrence B	175012226	0.0906	Agilent_SureSelect_Human_All_Exon_50mb	157511203	136.02	0.0529	0.8123
Patient28	Normal	242922762	0.2327	Agilent_SureSelect_Human_All_Exon_50mb	183537404	172.10	0.0516	0.8540
Patient28	Initial	279674136	0.2839	Agilent_SureSelect_Human_All_Exon_50mb	197100039	186.75	0.0503	0.8563
Patient28	Recurrence	261646218	0.1745	Agilent_SureSelect_Human_All_Exon_50mb	212857278	212.54	0.0514	0.8531
Patient29	Normal	302528350	0.3323	Agilent_SureSelect_Human_All_Exon_50mb	198828704	186.87	0.0507	0.8591
Patient29	Initial	261084924	0.2341	Agilent_SureSelect_Human_All_Exon_50mb	196936334	185.18	0.0519	0.8550
Patient29	Recurrence	318538994	0.1382	Agilent_SureSelect_Human_All_Exon_50mb	270007841	247.57	0.0443	0.8794
<hr/>								
Patient24	Normal	291504104	0.3240	Agilent_SureSelect_Human_All_Exon_V4	194079378	206.67	0.0004	0.9683
Patient24	Initial	222865850	0.2199	Agilent_SureSelect_Human_All_Exon_V4	171015094	182.73	0.0004	0.9621
Patient24	Recurrence1	239585468	0.1378	Agilent_SureSelect_Human_All_Exon_V4	203306203	215.22	0.0003	0.9726
Patient24	Recurrence2	206208472	0.2346	Agilent_SureSelect_Human_All_Exon_V4	155461755	166.69	0.0005	0.9472
Patient25	Normal	263944886	0.1502	Agilent_SureSelect_Human_All_Exon_V4	220616782	227.72	0.0004	0.9750
Patient25	Initial	272637310	0.1571	Agilent_SureSelect_Human_All_Exon_V4	226300494	238.18	0.0003	0.9760
Patient25	Recurrence	278331906	0.1945	Agilent_SureSelect_Human_All_Exon_V4	220670256	235.59	0.0004	0.9670
Patient26	Normal	235450112	0.2639	Agilent_SureSelect_Human_All_Exon_V4	170904446	185.27	0.0016	0.9642
Patient26	Initial A	252423160	0.2216	Agilent_SureSelect_Human_All_Exon_V4	192962251	200.94	0.0015	0.9685
Patient26	Initial B	202072954	0.0905	Agilent_SureSelect_Human_All_Exon_V4	182046280	177.99	0.0017	0.9557
Patient26	Recurrence A	297876622	0.1722	Agilent_SureSelect_Human_All_Exon_V4	243072156	262.40	0.0014	0.9795
Patient26	Recurrence B	146455804	0.0974	Agilent_SureSelect_Human_All_Exon_V4	130989039	130.12	0.0019	0.9173
Patient27	Normal	276349302	0.2756	Agilent_SureSelect_Human_All_Exon_V4	197384884	219.61	0.0005	0.9711
Patient27	Initial A	274051470	0.2790	Agilent_SureSelect_Human_All_Exon_V4	194508972	208.77	0.0005	0.9707
Patient27	Initial B	129212828	0.0788	Agilent_SureSelect_Human_All_Exon_V4	117837436	117.06	0.0009	0.9014
Patient27	Recurrence A	237800232	0.2553	Agilent_SureSelect_Human_All_Exon_V4	174652244	192.07	0.0005	0.9614
Patient27	Recurrence B	175012226	0.0906	Agilent_SureSelect_Human_All_Exon_V4	157511203	154.88	0.0006	0.9397
Patient28	Normal	242922762	0.2327	Agilent_SureSelect_Human_All_Exon_V4	183537404	195.17	0.0004	0.9665
Patient28	Initial	279674136	0.2839	Agilent_SureSelect_Human_All_Exon_V4	197100039	212.96	0.0003	0.9702
Patient28	Recurrence	261646218	0.1745	Agilent_SureSelect_Human_All_Exon_V4	212857278	244.10	0.0003	0.9722
Patient29	Normal	302528350	0.3323	Agilent_SureSelect_Human_All_Exon_V4	198828704	212.74	0.0004	0.9709
Patient29	Initial	261084924	0.2341	Agilent_SureSelect_Human_All_Exon_V4	196936334	210.95	0.0003	0.9694
Patient29	Recurrence	318538994	0.1382	Agilent_SureSelect_Human_All_Exon_V4	270007841	283.59	0.0003	0.9813

* For comparison, quality control metrics for patients 24-29 are presented for both SureSelect bait sets

Table 2.2. Quality control metrics for exome sequencing.

Table 2.3

Patient ID	Total number of mutations in initial tumor	Mutations shared with recurrence	Percent shared	Major pattern of genetic evolution
Patient17	18	1	6%	branched
Patient05	36	6	17%	branched
Patient06	4	1	25%	branched
Patient07	126	44	35%	intermediate
Patient18	66	25	38%	intermediate
Patient12	22	10	45%	intermediate
Patient24	13	6	46%	intermediate
Patient25	46	22	48%	intermediate
Patient13	30	15	50%	intermediate
Patient15	26	13	50%	intermediate
Patient04	17	9	53%	intermediate
Patient01	48	26	54%	intermediate
Patient16	36	20	56%	intermediate
Patient21	27	15	56%	intermediate
Patient29	43	27	63%	intermediate
Patient02	26	17	65%	intermediate
Patient11	27	19	70%	intermediate
Patient08	24	17	71%	intermediate
Patient10	36	26	72%	intermediate
Patient28	20	15	75%	linear
Patient09	9	7	78%	linear
Patient26	27	23	85%	linear
Patient27	36	31	86%	linear

Table 2.3. Patterns of genetic evolution between initial and recurrent gliomas

Table 2.4

Patient ID	Sample Type	CpA	CpT	CpG	CpC	ApA	ApT	ApG	ApC	Overall
Patient01	Initial	1.41	0.72	6.19	3.24	1.09	0.65	1.44	2.09	1.73
Patient01	Recurrence	42.74	135.31	51.39	261.99	1.89	2.57	1.88	6.49	67.39
Patient02	Initial	0.49	0.50	6.88	1.45	0.56	0.67	0.25	0.73	0.93
Patient02	Recurrence	0.44	0.45	9.00	0.50	0.27	0.63	0.00	0.33	0.80
Patient04	Initial	0.41	0.21	4.64	0.00	0.00	0.92	0.21	0.00	0.49
Patient04	Recurrence 1	0.63	0.85	7.14	0.45	0.26	0.92	0.43	1.25	1.01
Patient04	Recurrence 2	1.07	1.31	4.41	0.70	0.00	0.62	0.00	0.32	0.80
Patient04	Recurrence 3	1.46	2.56	4.76	1.14	0.52	1.84	0.43	1.87	1.57
Patient05	Initial	2.01	0.89	9.20	0.74	0.30	1.09	0.59	3.81	1.57
Patient05	Recurrence	42.99	148.61	89.35	233.83	2.62	6.50	2.82	8.56	70.12
Patient06	Initial	0.00	0.00	3.52	0.00	0.30	0.00	0.00	0.00	0.18
Patient06	Recurrence	0.82	0.85	5.26	0.69	0.00	1.06	0.56	0.81	0.86
Patient07	Initial	2.68	4.20	9.74	2.44	0.78	2.13	2.94	8.30	3.61
Patient07	Recurrence	3.53	3.80	11.11	4.26	2.33	1.83	3.17	11.45	4.54
Patient08	Initial	0.42	1.08	3.72	0.00	0.78	0.31	0.43	0.32	0.66
Patient08	Recurrence	2.52	2.14	9.65	1.60	0.78	1.54	0.86	0.94	1.97
Patient09	Initial	0.21	0.00	1.26	0.47	0.26	0.31	0.43	0.00	0.30
Patient09	Recurrence	1.25	2.33	4.09	1.12	1.04	0.31	1.27	0.93	1.40
Patient10	Initial	0.65	0.44	11.51	0.24	0.79	1.24	0.44	0.64	1.18
Patient10	Recurrence	46.29	185.65	82.49	339.40	1.55	2.75	2.31	3.08	90.87
Patient11	Initial	0.42	0.21	5.91	0.90	0.26	0.61	1.06	0.31	0.85
Patient11	Recurrence	0.63	0.00	6.45	1.80	0.52	0.92	1.27	0.31	1.11
Patient12	Initial	0.43	0.22	4.02	0.48	0.00	0.62	0.44	0.96	0.61
Patient12	Recurrence	0.41	0.00	7.48	0.89	0.26	1.22	0.21	0.31	0.84
Patient13	Initial	0.21	0.63	3.50	0.67	0.00	0.61	0.00	0.62	0.55
Patient13	Recurrence	0.62	0.84	5.20	0.89	0.00	0.92	0.21	0.00	0.78
Patient15	Initial	0.42	0.42	6.99	0.90	0.26	0.00	0.21	0.31	0.75
Patient15	Recurrence	0.42	1.08	8.01	1.16	0.26	1.24	0.65	0.95	1.20
Patient16	Initial	0.21	0.22	8.89	0.94	0.79	1.24	0.65	0.95	1.10
Patient16	Recurrence	0.22	1.10	10.48	0.96	0.79	0.93	1.32	0.64	1.35
Patient17	Initial A	0.83	0.21	2.31	0.22	0.52	1.22	0.21	0.00	0.55
Patient17	Initial B	0.83	0.00	1.72	0.22	0.52	0.92	0.21	0.00	0.45
Patient17	Initial C	0.83	0.21	2.30	0.00	0.26	0.92	0.00	0.00	0.42
Patient17	Recurrence A	0.62	0.63	6.35	1.34	1.03	1.53	1.05	1.54	1.36
Patient17	Recurrence B	0.62	0.84	7.46	1.56	0.52	2.75	1.48	1.85	1.65
Patient17	Recurrence C	0.62	0.64	5.79	1.12	0.52	1.23	0.63	1.55	1.14
Patient17	Recurrence D	0.62	0.84	7.47	1.12	0.78	2.75	1.27	1.54	1.55
Patient18	Initial A	2.71	1.70	4.11	1.58	1.56	3.06	1.69	0.93	2.02
Patient18	Initial B	2.50	1.91	5.84	1.81	1.56	2.76	0.42	1.55	1.99
Patient18	Initial C	2.28	1.69	5.79	1.35	1.56	3.98	1.90	0.93	2.14
Patient18	Initial D	2.28	1.69	4.02	1.12	1.55	3.05	1.47	0.92	1.84
Patient18	Recurrence	53.69	176.51	77.24	290.79	1.29	5.19	3.57	4.61	83.78
Patient21	Initial	0.83	1.06	4.07	1.13	0.26	0.00	0.00	0.62	0.78
Patient21	Recurrence	31.22	69.48	35.95	91.67	0.26	2.14	2.31	4.31	31.94
Patient24	Initial	0.21	0.21	1.55	0.64	0.00	0.31	0.21	0.00	0.32
Patient24	Recurrence 1	0.20	0.62	4.11	0.64	0.26	0.31	0.21	0.31	0.60
Patient24	Recurrence 2	47.58	171.75	55.34	278.01	0.79	3.11	1.68	4.32	79.54
Patient25	Initial	1.23	0.62	6.15	1.06	0.79	0.93	0.21	0.91	1.14
Patient25	Recurrence	1.64	2.08	16.00	2.56	1.05	1.55	1.46	0.61	2.52
Patient26	Initial A	0.41	0.42	3.61	0.21	0.53	1.55	0.83	0.31	0.76
Patient26	Initial B	0.41	0.21	3.65	0.22	0.53	1.24	1.26	1.54	0.90
Patient26	Recurrence A	0.82	0.62	6.14	0.64	1.05	1.23	1.04	0.61	1.17
Patient26	Recurrence B	0.63	0.21	5.90	0.44	0.53	0.94	1.28	1.88	1.11
Patient27	Initial A	0.82	0.62	5.66	0.85	0.52	0.93	0.42	1.22	1.05
Patient27	Initial B	1.48	0.86	7.57	2.00	0.80	0.95	0.43	1.26	1.51
Patient27	Recurrence A	1.03	1.04	6.21	1.71	0.79	0.93	1.04	1.23	1.44
Patient27	Recurrence B	0.83	0.85	5.27	1.52	0.79	0.94	0.63	1.55	1.26
Patient28	Initial	0.00	0.00	2.57	1.07	0.26	0.93	0.42	0.31	0.54
Patient28	Recurrence	0.00	0.63	4.65	0.86	0.79	0.62	0.83	0.00	0.80
Patient29	Initial	1.03	0.83	8.76	0.85	0.26	1.24	0.00	1.22	1.24
Patient29	Recurrence	1.22	1.03	6.65	1.06	0.52	0.62	0.62	0.91	1.23

Mutation rates are given in number of mutations/Mb of sequence

Table 2.4. Mutation rates in each tumor at each dinucleotide context and overall.

Table 2.5

Patient ID	Mutation	Additional initial tumor pieces	Mutational Status
Patient01	MTOR S2215F	6	Absent
Patient01	ERBB2 D989N	6	Absent
Patient01	RB1 c.2520+1G>A	6	Absent
Patient01	NF1 T685I	6	Absent
Patient05	CDK4 K84N	6	Absent
Patient05	CDK6 D311N	6	Absent
Patient05	CDKN2A P114L	6	Absent
Patient10	CDKN2A P114L	3	Absent
Patient10	TSC1 E646K	3	Absent
Patient18	NF1 L1475F	3	Absent
Patient18	NF1 T1951I	3	Absent
Patient18	PIK3CA E542K	3	Absent

Table 2.5. Sanger sequencing of TMZ-associated mutations.

CHAPTER 3:

**DNA METHYLATION AND SOMATIC MUTATIONS CONVERGE ON CELL CYCLE
AND DEFINE SIMILAR EVOLUTIONARY HISTORIES IN BRAIN TUMORS**

3.1 Abstract

The evolutionary history of tumor cell populations can be reconstructed from patterns of genetic alterations. In contrast to stable genetic events, epigenetic states are reversible and sensitive to the microenvironment, prompting the question whether epigenetic information can similarly be used to discover tumor phylogeny. We examined the spatial and temporal dynamics of DNA methylation in a cohort of low-grade gliomas and their patient-matched recurrences. Genes transcriptionally upregulated through promoter hypomethylation during malignant progression to high-grade glioblastoma were enriched in cell cycle function, evolving in parallel with genetic alterations that deregulate the G₁/S cell cycle checkpoint. Moreover, phyloepigenetic relationships robustly recapitulated phylogenetic patterns inferred from somatic mutations. These findings highlight widespread co-dependency of genetic and epigenetic events throughout brain tumor evolution.

3.2 Significance

Deciphering the evolutionary history of a tumor illuminates the sequence of events that occurred in tumorigenesis prior to diagnosis. The earliest events may provide ideal targets for precision therapeutic approaches, as these alterations are present in nearly all cells of a tumor. Here, we show that spatial and temporal patterns of either reversible DNA methylation or irreversible somatic mutations produce remarkably similar evolutionary histories. Phenotypically, mutations and promoter region DNA hypomethylation converge to deregulate the cell cycle as indolent low-grade tumors progress to high-grade malignancies. This study suggests strong interdependency of genetic and epigenetic alterations in these human brain tumors.

3.3 Introduction

Cancers develop through a process of clonal evolution in which ongoing genetic and epigenetic diversification allows for repeated cycles of subclonal selection and expansion (90, 91). As a result, human tumors can display substantial intratumoral heterogeneity, including discordant genetic alterations between initial tumors and their associated local recurrences or distant metastases (93, 99, 100, 123). While genomic profiling of spatially or temporally separated tumor samples can be used to reconstruct the evolutionary history and underlying clonal architectures of individual tumors (124), this view is incomplete without a parallel analysis of the heterogeneity and evolution of the epigenome, an approach only rarely attempted (95, 125).

In low-grade glioma, the course of tumor evolution is particularly clinically significant. World Health Organization (WHO) grade II gliomas (low-grade gliomas) are diffuse, infiltrative tumors that frequently recur and may unpredictably undergo malignant progression to a higher grade with a worse prognosis (3). Recurrences that progress to highly malignant WHO grade IV glioblastoma (GBM) acquire genetic alterations in the RB and AKT-mTOR pathways (17, 113, 126). In fact, adjuvant treatment with alkylating chemotherapeutics such as temozolomide (TMZ) can induce hypermutation that emerges in recurrent tumors (15), and we recently linked treatment-associated driver mutations in these two pathways to malignant progression of grade II glioma to GBM (126). It remains unknown, however, how epigenetic alterations contribute to the different courses of evolution of low-grade gliomas and how or if they relate to concurrent mutational evolution.

The critical role that epigenetic alterations play in the development and therapeutic response of gliomas is increasingly being appreciated (127). Epigenetic mechanisms can alter gene expression, and have been shown to affect tumor suppressors and oncogenes in gliomas (56, 128-132). Somatic mutations in *IDH1* or *IDH2* may be the first genetic driver in the

development of many low-grade gliomas (104, 126, 133). Genetic mutations in IDH genes induce a pattern of early epigenetic alterations known as the glioma CpG island methylator phenotype (G-CIMP) characterized by extensive remodeling of the DNA methylome (68, 78, 82, 134). The inactivation of other genes mutated in low-grade gliomas, such as *ATRX* (103) and *SMARCA4* (126), is known to induce specific DNA methylation changes as well (135, 136). Of clinical importance is DNA hypermethylation of the *MGMT* promoter, which is associated with loss of SP1 binding, closed chromatin and transcriptional silencing in GBM cells (58, 137) and increased survival in GBM patients treated with TMZ (13). Whether the DNA methylation status at this locus predicts the same survival benefit in low-grade glioma patients is unclear (18, 138-141). Although there has been extensive characterization of tumor methylomes using a single sampling per tumor, little is known about intratumoral heterogeneity at the epigenetic level or of temporal evolution of the low-grade glioma methylome and its relationship to the genome. An integrated model of the genomic and epigenomic evolutionary trajectory of initially low-grade gliomas may suggest strategies for delaying or treating recurrent disease, identify biomarkers for predicting the clinical course of a low-grade glioma, and also shed light on dynamic relationships between the genome and epigenome in other cancer types.

3.4 Results

We profiled the DNA methylomes of 19 clinically-annotated initial grade II gliomas and their patient-matched recurrences (Table 3.1) using the Illumina HumanMethylation450 bead array (Illumina 450K) (Figure 3.2A). We also performed transcriptome sequencing on the initial and recurrent tumors of 13 patients (Table 3.2). All gliomas profiled here are *IDH1*-mutant (18, 126) and are therefore expected to possess the characteristic methylation patterns associated with G-

CIMP (78, 82, 83). From these methylation array data we confirmed that G-CIMP was present in all initial tumors and always maintained at recurrence (Figure 3.2B), highlighting that these epigenetic changes arise very early and are potentially tumor-initiating.

3.4.i Patterns of DNA methylation are patient-specific and evolve in a manner specific to the grade of recurrence

To determine the extent to which these tumors had altered methylomes beyond the ubiquitous G-CIMP methylation patterns, we identified the most variable CpG sites across all initial and recurrent gliomas and performed unsupervised hierarchical clustering. Initial and recurrent tumors from the same individual clustered together (Figure 3.2C). This result reflects patient-specific methylation patterns, consistent with a previous report on glioma (142), and may be indicative of normal inter-individual epigenetic variation, patient-specific aberrant methylation from early stages of gliomagenesis, or both. Within the clustering, six of the seven patients who recurred with GBM formed a distinct subgroup, suggesting there may be a shared methylation pattern associated with malignant progression to GBM relative to a lower grade of recurrence. To further evaluate this pattern, we performed unsupervised clustering with progressively more lenient selections of variable CpG sites to discover additional global DNA methylation patterns. At intermediate cutoffs, a gradual switch in clustering patterns was evident (Figure 3.2D). At the most lenient cutoff, the methylation patterns separated GBM recurrences, as well as two initial tumors that recurred as GBM, from the grade II and III gliomas (Figure 3.1A). This further supports a GBM recurrence-specific methylation pattern and suggests extensive evolution of the methylome during malignant progression to GBM (Figure 3.2E). This unique pattern of epigenome evolution was prominent across GBM recurrences that arose in the absence of

adjuvant therapy as well as in GBMs that arose in a treatment-associated manner, adding to our prior genetic findings that spontaneous and treatment-associated progression to GBM have convergent genetic alterations (126). Interestingly, clustering of the transcriptome segregated some of the grade III recurrences with GBM samples (Figures 3.2F and 3.2G), indicating transcriptional changes are complementary to, but not exclusively overlapping with, changes in the DNA methylome during malignant progression. Thus, integrating the methylome and transcriptome may provide important insight into the functional epigenetic events that underlie malignant progression to GBM.

3.4.ii Identification of CpGs that lose methylation specifically during malignant progression to GBM

We next examined changes in the methylome and transcriptome to determine whether there is a signature of methylation or expression changes associated with recurrence. We calculated the change in methylation (β value, methylated fraction at a CpG site) from initial to recurrent tumor at each CpG site in each patient, and then identified CpG sites with consistent methylation changes upon recurrence across all patients. This powerful intra-patient approach controls for differences in DNA methylation that are age-related or reflect germline genetic effects, which confound inter-patient comparisons. DNA methylation differences between normal brain and glioma may be aberrant events in the tumor, or may reflect differences between the normal brain tissue sample and the methylation patterns of the tumor's cell of origin (143, 144). In contrast, the differences we report between initial and recurrent tumors are more likely to be aberrant changes attributable to tumor progression rather than cell of origin. We also applied an equivalent model to the transcriptome sequencing data and identified genes that commonly

increase or decrease in expression from initial to recurrent glioma (Figure 3.2H). The separation by grade in the methylation clustering suggested that a specific pattern of DNA methylation changes may underlie malignant progression to GBM. To discover this pattern in detail, we stratified patients by grade of recurrence. There were few common methylation changes evident in patients that recurred at grade II or III, whereas a strong pattern of hypomethylation was associated with malignant progression to GBM (Figures 3.1B, 3.1C and 3.2I). Patients that recurred at grade II or III were combined into a single group for further analysis.

To determine which methylation changes were specific to recurrence as GBM, we compared the change in methylation from initial to recurrence in patients who recurred as GBM versus those that recurred at grades II or III. We identified 1,953 CpG sites that were specifically hypomethylated upon recurrence as GBM (Figure 3.3A) (145). Given the G-CIMP-associated hypermethylation in these tumors, we first set out to determine if the hypomethylation in GBM recurrences affected G-CIMP genes. Noushmehr et al. identified 50 genes that were hypermethylated and downregulated in a G-CIMP specific manner. Only two of those genes (*ACSS3* and *RAB36*) showed GBM-specific hypomethylation, but in neither case did the genes show concurrent decreased expression. Further examination of these sites of decreasing methylation revealed a surprising enrichment for CpG sites that undergo age-related increased methylation in a comparison of normal fetal and adult brain (odds ratio 4.64, p value < 10^{-4} , permutation test). This is contrary to the typical pattern in cancer in which CpG sites that are hypermethylated during aging are also hypermethylated in cancer (68, 146). To further investigate whether the methylation changes alter gene regulation, we integrated active regulatory regions defined from histone H3K4me3, H3K4me1 and H3K27ac ChIP-seq in adult normal brain and primary GBM tissue (Figure 3.4A and Table 3.2) and found that sites of GBM-

specific DNA hypomethylation were enriched for candidate active enhancers (odds ratio 1.68, p value $< 10^{-4}$, permutation test). These hypomethylated loci thus may have gene regulatory effects. To enrich for functional methylation changes and exclude passenger events, we next integrated our transcriptome sequencing data with the DNA methylation analysis.

3.4.iii Cell cycle genes are specifically hypomethylated upon malignant progression

We applied an analysis similar to that of the methylation data and identified 528 genes with GBM-specific over-expression (Figure 3.3A). Of these, 39 genes showed GBM-specific hypomethylation of at least one CpG site within their promoter regions (Figures 3.4B, 3.4C) (145). Among genes with GBM-specific promoter hypermethylation, only *NTSR2* showed consistent transcriptional downregulation. We additionally identified four genes with consistent downregulation and gene body hypomethylation (145). Strikingly, the set of 39 hypomethylated and over-expressed genes was significantly enriched for cell cycle genes (Figure 3.3B) (145). Ki-67 is a marker of actively proliferating cells, and staining in initial and recurrent tumors confirmed that a statistically significantly higher fraction of cells (p value = 0.026, two-sided Wilcoxon rank sum test) were actively proliferating among the GBM recurrences (Figures 3.3C and 3.3D). Increased proliferation is a hallmark of GBM. These results thus highlight an epigenetic mechanism that may contribute to increased proliferation, concurrent with genetic alterations in key members of the RB pathway (126) that abrogate the G₁/S cell cycle checkpoint.

Among the epigenetically modified cell cycle genes, we noted that the hypomethylation in *TP73* was at an internal gene body promoter. Indeed, *TP73* possesses a gene body CpG island (CGI) that we identified recently as recurrently hypomethylated in primary GBM (130). The gene body CGI spans the transcription start site of a truncated, oncogenic form of TP73

($\Delta Np73$), which is correspondingly expressed in primary GBM. Similarly, increased expression (Figure 3.4D) and hypomethylation of $\Delta Np73$ was observed only in GBM recurrences. Due to the limited resolution of the Illumina 450K array, we identified only one significantly hypomethylated CpG site in this gene body CGI specifically upon recurrence as GBM, although several nearby CpG sites showed a similar trend. To gain greater resolution across the full CGI and other regions genome-wide, we examined our whole-genome shotgun bisulfite sequencing (WGBS) data on the initial and recurrent tumors of Patient01 (Figure 3.3E and Table 3.2). These data show that the pattern of hypomethylation indeed extends across the local genomic region. Among other hypomethylated and over-expressed genes, significant probes from array-based data were similarly indicative of a local effect including multiple CpG sites (Figures 3.4E and 3.4F), consistent with previous literature showing that the methylation levels of CpG sites within 1kb are highly correlated (147).

The functional effect of DNA hypomethylation of cell cycle genes specifically upon recurrence as GBM parallels the known GBM-specific genetic events that inactivate the G₁/S cell cycle checkpoint (17, 113, 126). These convergent genetic and epigenetic signals, in addition to the well-characterized functional relationships between genetic and epigenetic aberrations (82, 83, 135, 136, 148), prompted us to explore evolutionary relationships among different tumor cell populations within a tumor, as has been previously done with genetic data, and then compare the relationships inferred from DNA methylation to those inferred from somatic mutation in the same samples.

3.4.iv Reconstruction of tumor evolution from intratumoral and longitudinal DNA methylation patterns

We first examined the evolutionary relationships of tumor samples that were previously genetically characterized (126). We performed methylation profiling of seven spatially distinct pieces of tumor tissue from Patient17, three from the initial tumor and four from the recurrent tumor, and built a phyloepigenetic tree (Figure 3.5A, left) (145). The phyloepigenetic tree presented an intriguing model with early divergence between the initial and recurrent tumors, and more subtle divergences among the samples within each time point (initial A vs. initial B/C; recurrence A/C vs. recurrence B/D). We then used exome sequencing data of these same spatially distinct tumor samples to independently construct a phylogenetic tree (Figure 3.5A, right) (126, 145). The genetically defined relationships among tumor cell clones were consistent with those determined from DNA methylation data. We quantified this similarity as the correlation between the distance matrices that were used to build the phyloepigenetic and phylogenetic trees (Spearman's $\rho = 0.90$).

To identify the CpG sites underlying each branch point in the phyloepigenetic tree, we applied singular value decomposition to the methylation data from each patient to weigh the influence of individual CpG sites on separating particular subsets of samples (Figure 3.6A). For Patient17, the first singular vector (SV1), which accounts for the most methylation variability, mimicked the first major branch point of the phyloepigenetic tree (Figure 3.5B). We then selected the most influential CpGs for each singular vector and inferred that these underlie a particular branch point. The most highly weighted CpG sites within SV1 from Patient17 clearly showed differential methylation between the initial and recurrent tumor samples (Figure 3.5C). We examined the potential implications of these methylation changes by focusing on those affecting active promoters and enhancers in normal brain and primary GBM tissue and performed a gene ontology enrichment analysis. For Patient17, the CpG sites that underlie the

first major branch point were enriched for a variety of developmental, biosynthetic and metabolic processes, indicating that methylation changes during tumor progression may influence cellular metabolic states, in parallel with the genetic events disrupting cell cycle that separate these two main branches on the phylogenetic tree (145).

We then looked specifically at the evolutionary relationships of tumor samples from patients that underwent chemotherapy-associated malignant progression (18, 126). We performed methylation profiling of four spatially distinct pieces of the initial tumor and three pieces of recurrent tumor from Patient01 and inferred a phyloepigenetic tree (Figure 3.5D, left; Figure 3.6B) (145). While the four pieces of the initial tumor clustered together, the recurrent tumor consisted of two distinct populations. Recurrence B was relatively closely related to the initial tumor, while a long branch separated it from recurrences A and C, indicating significant evolutionary distance. A phylogenetic tree from these same tumor pieces (Figure 3.5D, right; Table 3.2) (145) similarly demonstrates the large evolutionary distance between recurrence B and recurrences A and C (Spearman's $\rho = 0.83$). In the phylogenetic tree, this longest branch corresponds to the development of a hypermutated population in the recurrent tumor. Intriguingly, this same branch is the longest in the phyloepigenetic tree, indicating that the hypermutated cells also have the greatest methylation change. Similarly, in Patient18, the phyloepigenetic tree identified three epigenetically similar pieces of the initial tumor, a piece of the initial tumor that branched off at an earlier evolutionary time point, and a recurrence that diverged even earlier – relationships that are accurately recapitulated in the phylogenetic tree (Spearman's $\rho = 0.90$) (Figures 3.5E and 3.6C) (145). Thus, even in extreme evolutionary events such as chemotherapy-associated hypermutation, both DNA methylation changes and mutational landscapes encode similar tumor evolutionary relationships. In these two cases with

TMZ-associated hypermutation (Figures 3.5D and 3.5E), the longest branch length in both the phyloepigenetic and phylogenetic trees is the hypermutated recurrence. These results suggest a potentially quantitative relationship between the number of mutations and epimutations in each tumor cell clone.

To determine if the strong correlations between phylogenetic and phyloepigenetic trees depend on the large-scale hypomethylation during malignant progression to GBM, we next compared the evolutionary relationships only in lower grade initial and recurrent tumors. Six pieces of tissue from the initial tumor and two pieces of tissue from the grade II recurrence from Patient90 were subjected to DNA methylation profiling. Construction of a phyloepigenetic tree revealed three distinct clusters of samples, with the initial tumor separating into two populations, and the recurrence forming a third (Figure 3.5F, top; Figure 3.6D) (145). We then performed exome sequencing of these same pieces of tissue to identify somatic mutations and constructed a phylogenetic tree (Figure 3.5F, bottom; Table 3.2) (145). This phylogenetic tree mirrored the evolutionary relationships defined from DNA methylation (Spearman's $\rho = 0.56$). We further pursued this question with Patient49 who underwent a single resection for an initial tumor from which we profiled six spatially distinct pieces. Construction of a phyloepigenetic tree revealed that the six pieces separate into two groups, in agreement with the phylogenetic tree derived from exome sequencing of the same pieces of tissue (Spearman's $\rho = 0.64$) (Figures 3.5G and 3.6E; Table 3.2) (145). Thus, even in the absence of malignant progression to GBM, DNA methylation changes among tumor cell clones yielded a very similar evolutionary trajectory as was inferred from somatic mutations.

3.4.v Enhanced model of tumor evolution derived from variation between phyloepigenetic and phylogenetic trees

To further address phyloepigenetic relationships over time, we examined tumor samples from Patient04, who had four sequential surgical resections over five years. We profiled six spatially distinct pieces of tumor from the initial surgery, and one from each of the three subsequent surgeries for tumor recurrence. The phyloepigenetic tree reveals two distinct populations within the initial tumor and an evolutionary trajectory shared among the three recurrences, with a relatively closer relationship between recurrences 2 and 3 (Figure 3.7, left; Figure 3.8) (145). The phylogenetic tree again reveals many similar clonal relationships, but also reveals differences that may be informative (Figure 3.7, right; Table 3.2) (145) (Spearman's $\rho = 0.78$). Based on somatic mutations, the first recurrence shares evolutionary history with the initial tumor, while the second recurrence diverged earlier in the evolution of the tumor and therefore independently progressed to grade III (126). Despite divergent genetic paths, methylation patterns are shared among the first recurrence and the second and third recurrences. This raises the possibility that the last common ancestor of the first and second recurrences was primed for progression with a set of DNA methylation changes required for progression to a higher grade. This case illustrates how differences in genetic and epigenetic phylogenies may bring to light an enhanced understanding of the evolution of a tumor.

3.4.vi Gene-level genetic and epigenetic convergence

The common evolutionary histories defined from mutations and DNA methylation led us to examine if there was also convergence at the level of individual genes. We identified a small number of intra-patient single gene convergence events in which some samples from a patient

had a mutation, while other samples which lack the mutation show differential methylation at the same gene (Table 3.3). We also identified a small number of genes with inter-patient convergence. These are genes that are mutated in one patient but show methylation alteration in another patient (Table 3.3). However, the vast majority of mutations and methylation changes occur in different sets of genes, consistent with our prior low-resolution analysis of gliomas (60). In contrast to single genes, single pathways such as the cell cycle pathway are commonly altered by multiple genetic (126) and epigenetic (Figure 3.3B) alterations within and across tumor samples.

3.5 Discussion

DNA methylation patterns record a remarkable breadth of information about cells, including their chronological age, developmental history and differentiation potential. Here, we show that despite epigenome plasticity, chemotherapy, and the ubiquitous *IDH1* mutation-driven G-CIMP pattern, patient-specific tumor phyloepigenetic analyses replicated and extended tumor phylogenetic analyses. From this striking result, we conclude that the precise chronological order of epigenetic changes, from initiating to late events, can be determined from intratumoral methylation patterns, thus surpassing prior binary categorization of epigenetic events as early or late. While our study is focused on methylation and somatic mutations in *IDH1*-mutant gliomas, a study of prostate cancer and prostate cancer metastasis showed a complementary unified model of evolution for DNA methylation and copy number alterations (95). Thus, genomic-epigenomic co-dependency may be a feature of multiple types of cancer, and may span somatic mutations, copy number, and DNA methylation.

The importance of epigenetic variation within individual human tumors is just beginning to be uncovered. Recent work in chronic lymphocytic leukemia suggests that stochastic changes in the methylome lead to increased heterogeneity, allowing for selection of more malignant epigenotypes coupled with an adverse clinical outcome (149). Somatic genetic events, like *IDH1* mutations, have been directly linked to alterations in the methylome (78, 82), while germline variants have been indirectly associated with specific DNA methylation patterns (148, 150, 151). Consistent with these theories, the widespread correlation between somatic mutations and DNA methylation patterns suggests that in addition to *IDH1* mutation and G-CIMP, other epigenetic patterns might be directly or indirectly induced by mutations, or vice versa. It will be of interest to determine the extent to which these findings hold for *IDH1*-wild-type low-grade gliomas and their recurrences.

We also discovered a convergence of genetic and epigenetic changes driving aberrant cell cycle function (Figure 3.9). We previously found that recurrent tumors that underwent malignant progression to GBM acquired somatic mutations in the RB pathway that inactivate the G₁/S cell cycle checkpoint (126). Here we identified a pattern of functional DNA hypomethylation specific to recurrence as GBM that alters cell cycle genes. This phenotypic convergence of genetic and epigenetic mechanisms on the same pathway underscores the importance of cell cycle deregulation on the process of malignant progression, while also raising questions about how these two processes might be connected. Of note, we identify hypomethylation at *TP73* as a recurrent event. Transcription of *TP73* is upregulated by E2F1 (152), a transcription factor that itself activates cell cycle progression-related genes following inactivation of the RB pathway (153), which is deregulated by genetic mechanisms in these tumors. Further work will be required to deconvolute these relationships. By combining the information from somatic

mutations, copy number alteration and DNA methylation patterns, we derived a comprehensive model of glioma evolution (Figure 3.9). Chronological ordering of *IDH1*, *TP53*, and *ATRX* mutations and copy number alterations was derived from our previous tumor phylogenetic analyses (126), other studies (104, 133), and additional data presented here. This model is derived from a total of 32 patients with paired initial and recurrent samples and includes 70 DNA methylation profiles, 26 mRNA expression profiles and 130 exome sequencing profiles. The model extends from the initiating genetic and epigenetic lesions and captures clinically divergent paths at recurrence, including an evolutionary path driven by treatment.

These findings underscore the power of integrated genetic and epigenetic analyses of tumors. Deregulated cell cycle control is among the essential phenotypes of cancer cells, and we demonstrate that this deregulation is encoded in both the genome and epigenome, raising the question of the extent to which this reflects a functional interaction between genetics and epigenetics. This finding also raises the possibility that other critical molecular phenotypes, such as genomic instability, angiogenesis or invasion may leave their imprint on DNA methylation patterns during tumor evolution.

Figure 3.1

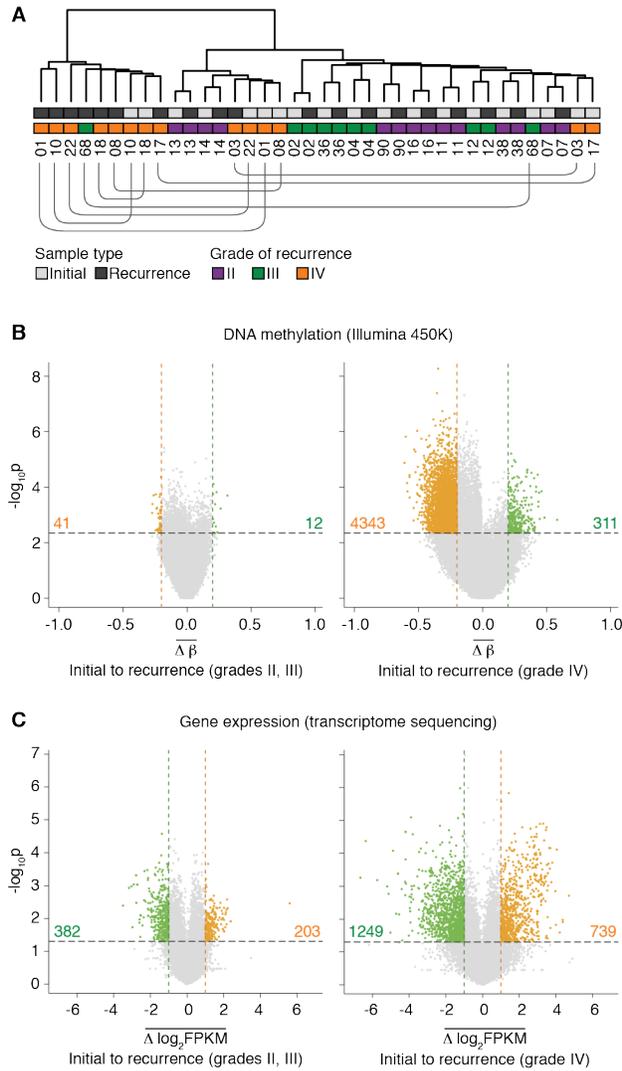


Figure 3.1. Evolutionary dynamics of the methylome and transcriptome in initial and recurrent glioma pairs. (A) Unsupervised hierarchical clustering of the top 50% most variable CpG sites. Annotations of sample type, grade of recurrence, and patient identification numbers are provided. The lines beneath the patient identification numbers connect initial and recurrent tumors from the same patient that are not adjacent to each other. **(B)** The average methylation change from initial low-grade tumor to recurrence at each CpG site measured in patients that do not (left) or do (right) undergo malignant progression to GBM (grade IV). Colored dots represent

CpG sites that show significant hypomethylation (orange dots, total count provided) or hypermethylation (green dots, total count provided) at recurrence ($p_{\text{value}_{\text{adjust}}} < 0.05$ and $|\Delta\beta| > 0.2$). (C) Average gene-level expression changes from initial to recurrence in patients that do not (left) or do (right) undergo malignant progression to GBM. Significantly differentially expressed genes are highlighted in green (down-regulated at recurrence, total count provided) and orange (up-regulated at recurrence, total count provided) ($p_{\text{value}} < 0.05$ and $|\Delta\log_2\text{FPKM}| > 1$).

Figure 3.2

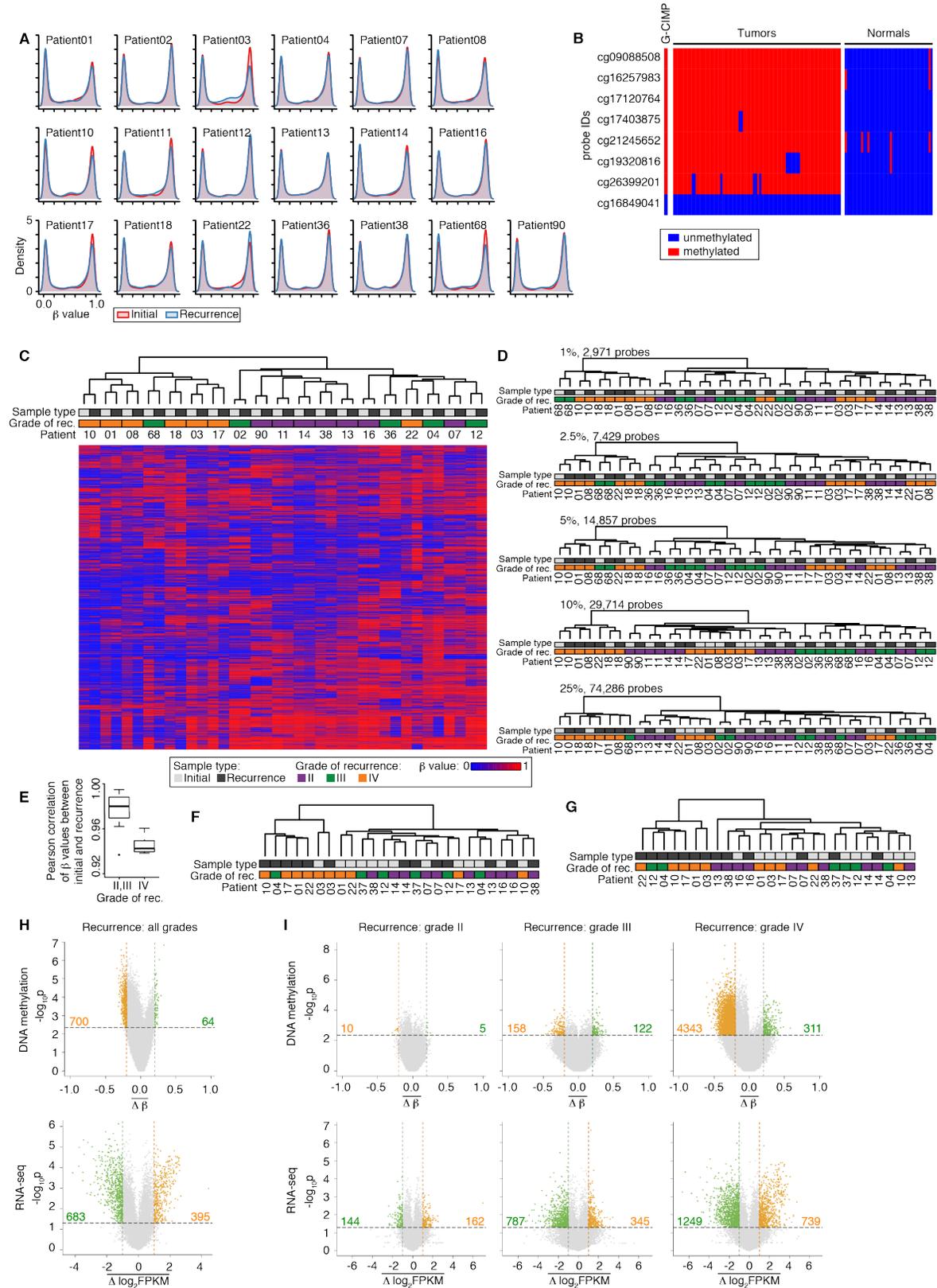


Figure 3.2. Plots supporting Figure 3.1 (A) Density plots of background corrected and normalized beta values in each initial and recurrent tumor. (B) Confirmation of the presence of the glioma CpG island methylator phenotype (G-CIMP) in all initial tumors and maintenance of G-CIMP at recurrence (tumor n=70). G-CIMP is absent from all normal brain tissues examined (normal brain n=38). (C) Unsupervised hierarchical clustering of the top 0.5% most variable CpG sites and heatmap of beta values. (D) Unsupervised hierarchical clustering of the most variable CpG sites at intermediate (top 1%, 2.5%, 5%, 10%, 25%) cutoffs. (E) Boxplot summarizing Pearson correlations of beta values between initial and recurrent tumors for each patient, grouped by the grade of the recurrent tumor, show decreased correlations in patients that recur as GBM. The box encompasses data points between the first and third quartiles, with a horizontal line indicating the median value. Whiskers extend to 1.5 x interquartile range, and any data points beyond that range are shown as individual dots. (F, G), Unsupervised hierarchical clustering of the top 1% (F) or top 50% (G) most variably expressed genes across the cohort. (H) The methylation change (top) and expression change (bottom) from initial low-grade tumor to recurrence at each CpG site (top) and gene (bottom), averaged across all patients in the cohort. Colored dots represent CpG sites (top) and genes (bottom) that show significant changes at recurrence. The number of significant CpG sites (top) and genes (bottom) are provided in each quadrant. (I) Methylation (top) and expression (bottom) changes from initial to recurrent tumor, subdivided by the grade of the recurrent tumor.

Figure 3.3

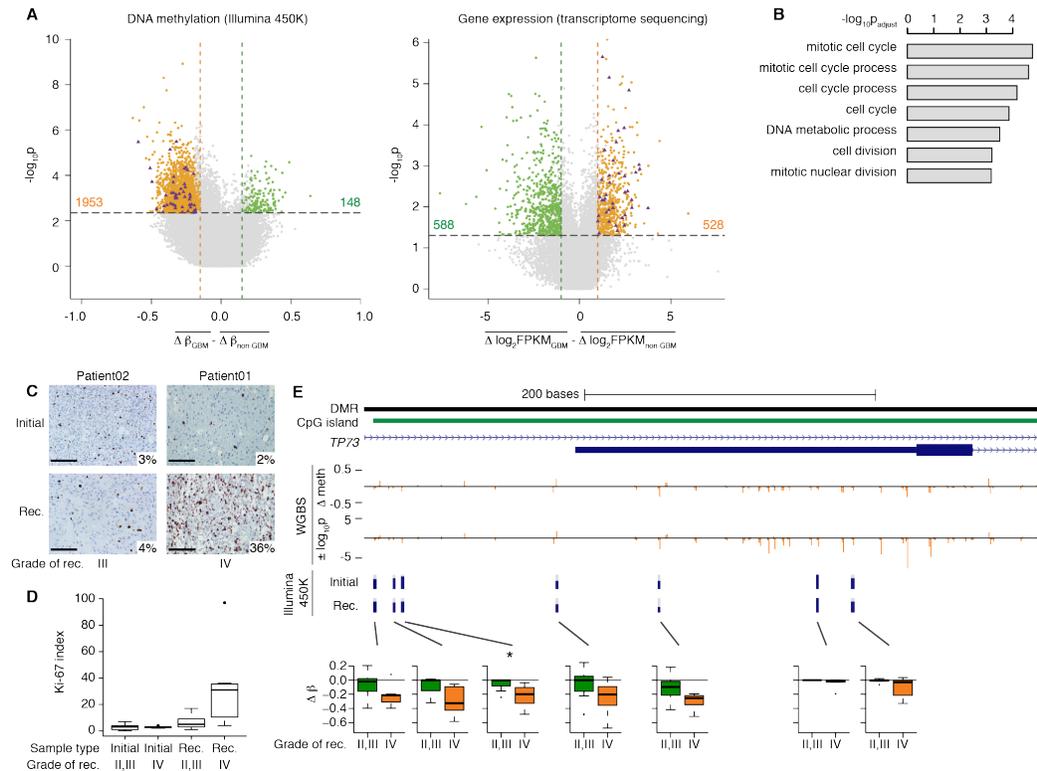


Figure 3.3. Cell cycle genes are hypomethylated and over-expressed specifically upon recurrence as GBM, coordinately with an increase in actively cycling cells. (A) Left panel shows a scatter plot of differences between GBM and non-GBM recurrent tumors in methylation changes from initial grade II to recurrent gliomas. Right panel shows an equivalent representation of differences in expression changes between GBM and non-GBM recurrent tumors. Colored points indicate significant differences. Purple triangles highlight genes that become hypomethylated at promoter CpGs (left) and over-expressed (right) during malignant progression to GBM. **(B)** Barplot of the top results of a gene ontology analysis of genes that are both significantly hypomethylated and over-expressed specifically upon recurrence as GBM. **(C)** Representative staining for Ki-67 in a patient that recurred at grade III (left) and a patient that recurred at grade IV (right). Bars represent 100 μ m. **(D)** Boxplot representing the Ki-67 labeling

index of tumors in the cohort (n=16 patients), subdivided by grade of recurrence (p value = 0.026, two-sided Wilcoxon rank sum test between GBM recurrences and recurrences at grades II or III). The box encompasses data points between the first and third quartiles, with a horizontal line indicating the median value. Whiskers extend to 1.5 x interquartile range, and any data points beyond that range are shown as individual dots. (E) Whole-genome shotgun bisulfite sequencing data (WGBS) of Patient01 across an intragenic CpG island in the *TP73* locus. From top to bottom, tracks represent: a differentially methylated region (DMR) reported in primary GBM (Nagarajan et al., 2014); CpG island; *TP73* full-length and truncated transcripts; change in methylation level from initial to recurrent tumor by WGBS; statistical significance of the WGBS methylation changes, where positive values indicate hypermethylation at recurrence and negative values indicate hypomethylation; methylation levels from Illumina 450K array in Patient01 at the seven CpG sites assayed on the array. Box plots present the methylation change in all patients in the cohort across the same seven CpG sites. Boxplots are drawn as in panel D.

Figure 3.4

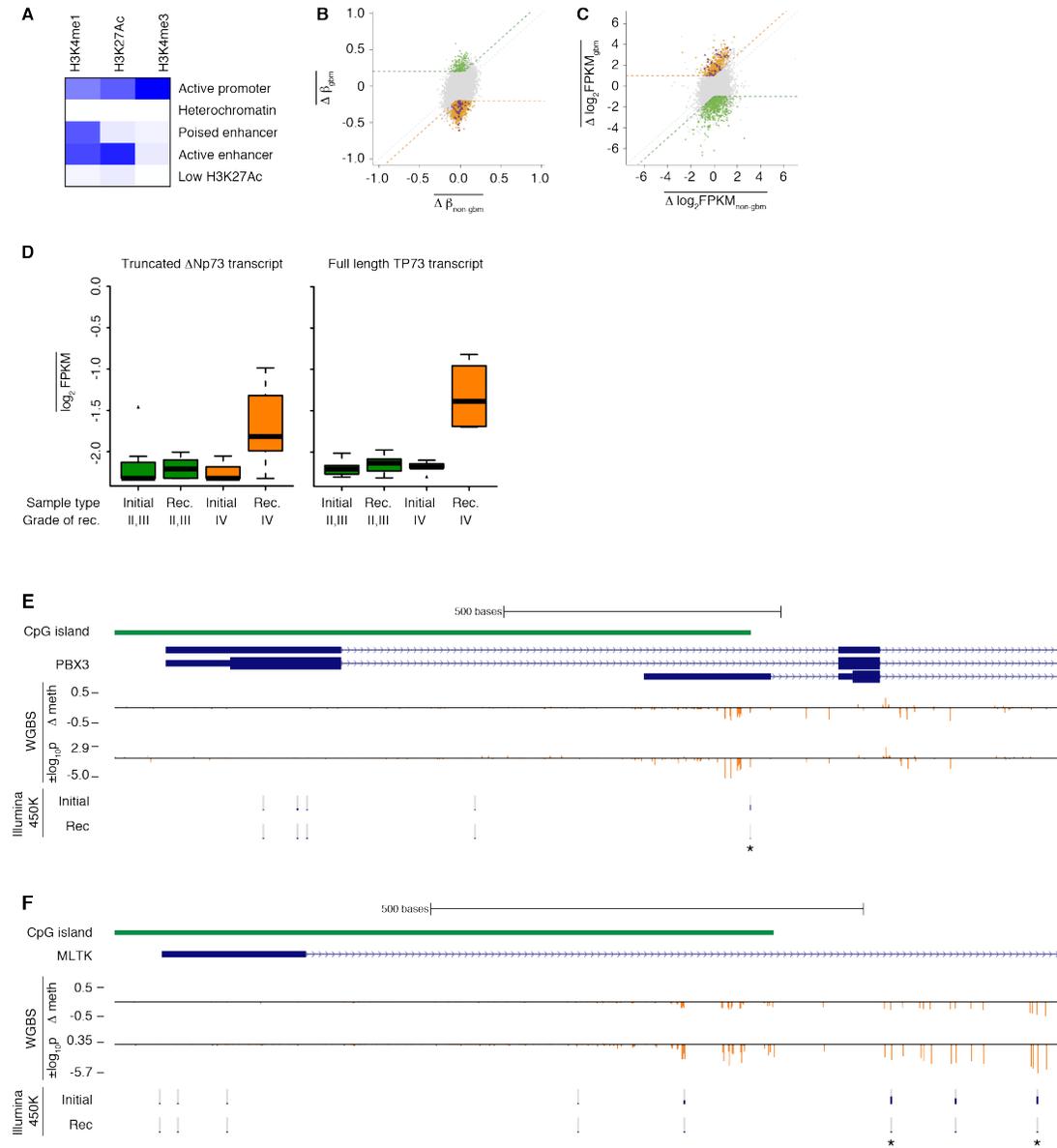


Figure 3.4. Plots supporting Figure 3.3 (A) Chromatin state outputs from ChromHMM applied to ChIP-seq for H3K4me1, H3K4me3, and H3K27ac. The intensity of the color in each box indicates the probability that a particular mark is present in each state. (B, C) Scatterplots show how the average change from initial to recurrent tumor in methylation (B) and expression (C) for each CpG site or gene differs between patients that recur as GBM (y-axis) and those that recur at

grades II or III (x-axis). Purple triangles highlight genes that become hypomethylated at promoter CpGs (B) and over-expressed (C) during malignant progression to GBM (see Supplemental Experimental Procedures). (D) Boxplot of \log_2 FPKM of all full-length TP73 (ENST00000346387.4, ENST00000354437.4, ENST00000357733.3, ENST00000378295.4, ENST00000603362.1, ENST00000604074.1, ENST00000604479.1) and truncated Δ Np73 (ENST00000378280.1, ENST00000378285.1, ENST00000378288.4) transcripts, averaged per patient. Both transcripts are uniquely expressed in grade IV recurrences. The box encompasses data points between the first and third quartiles, with a horizontal line indicating the median value. Whiskers extend to 1.5 x interquartile range, and any data points beyond that range are shown as individual dots. (E, F) WGBS data surrounding CpG sites within the promoter regions of PBX3 (E) and MLTK (F) that were hypomethylated specifically upon recurrence as GBM based on Illumina 450K data. An asterisk marks the CpG sites identified as hypomethylated on the array. Both genes show local regions of hypomethylation in the WGBS data and were upregulated upon recurrence based on transcriptome sequencing. From top to bottom, tracks represent: CpG island; gene transcripts; change in methylation level from initial to recurrent tumor by WGBS; statistical significance of the WGBS methylation changes, where positive values indicate hypermethylation at recurrence and negative values indicate hypomethylation; methylation levels from Illumina 450K array in Patient01 at the CpG sites assayed on the array.

Figure 3.5

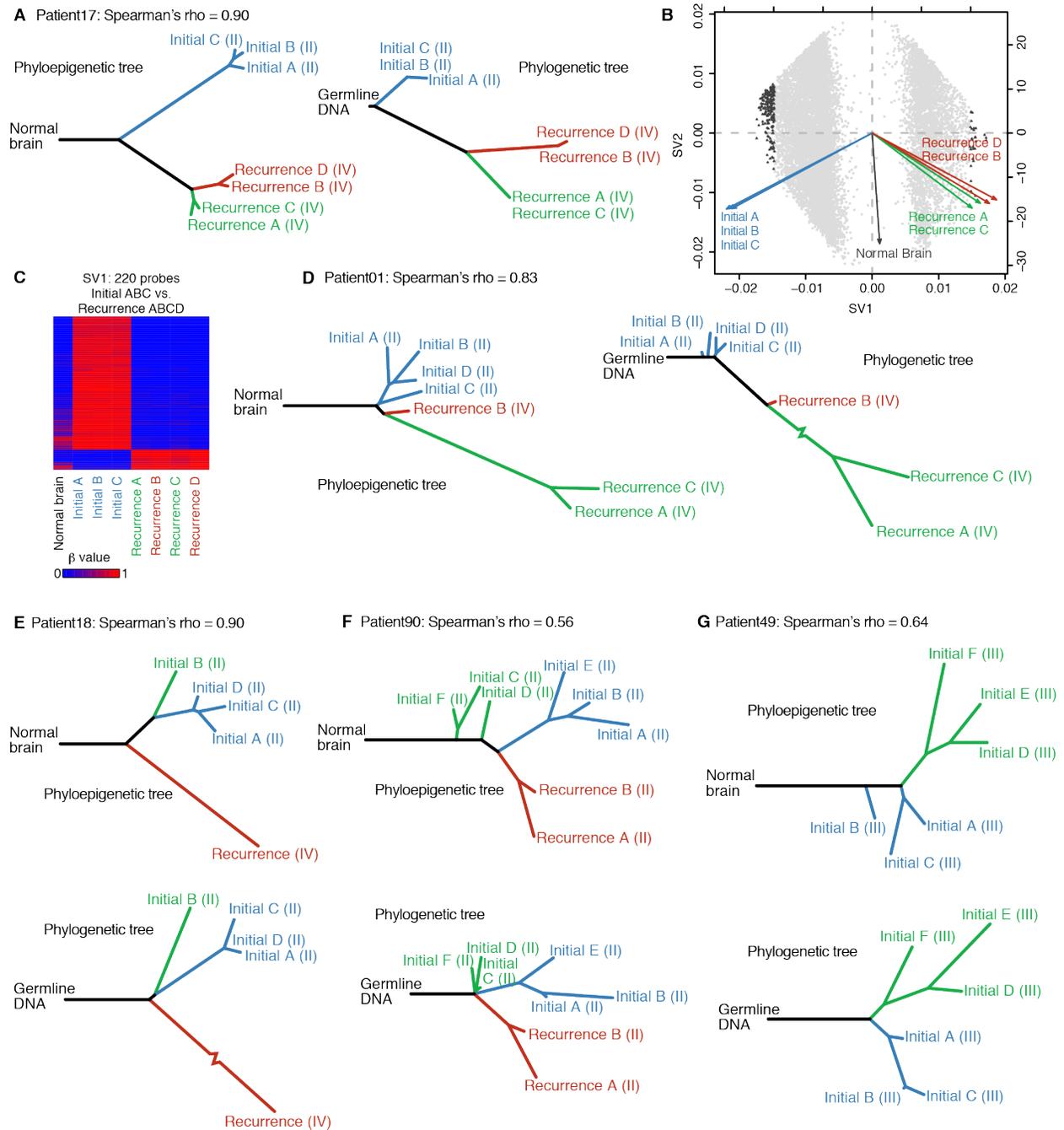
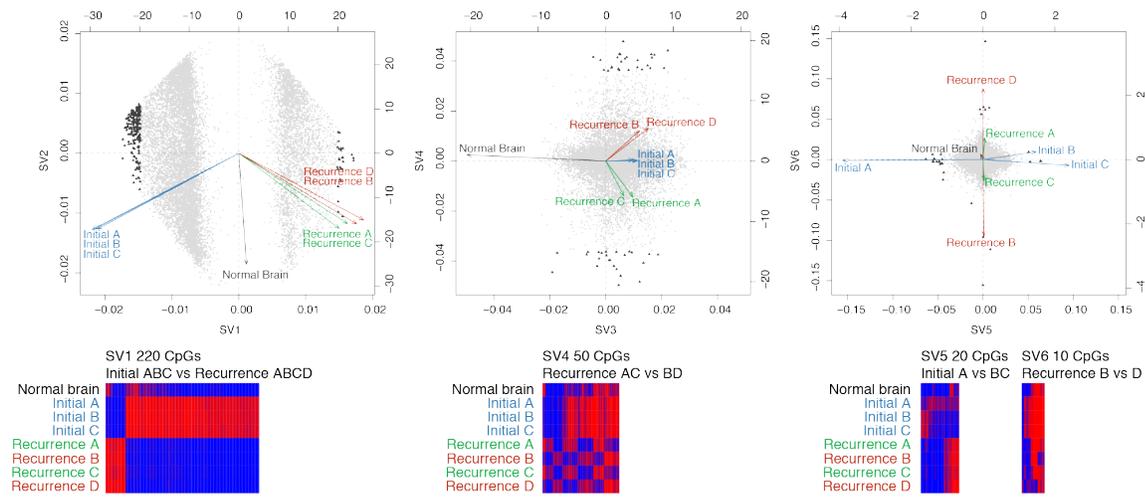


Figure 3.5. The spatial and temporal patterns of tumor evolution observed from DNA methylation dynamics and somatic mutations yield similar evolutionary histories. (A) A phyloepigenetic tree constructed from seven samples from Patient17 (left) and a phylogenetic

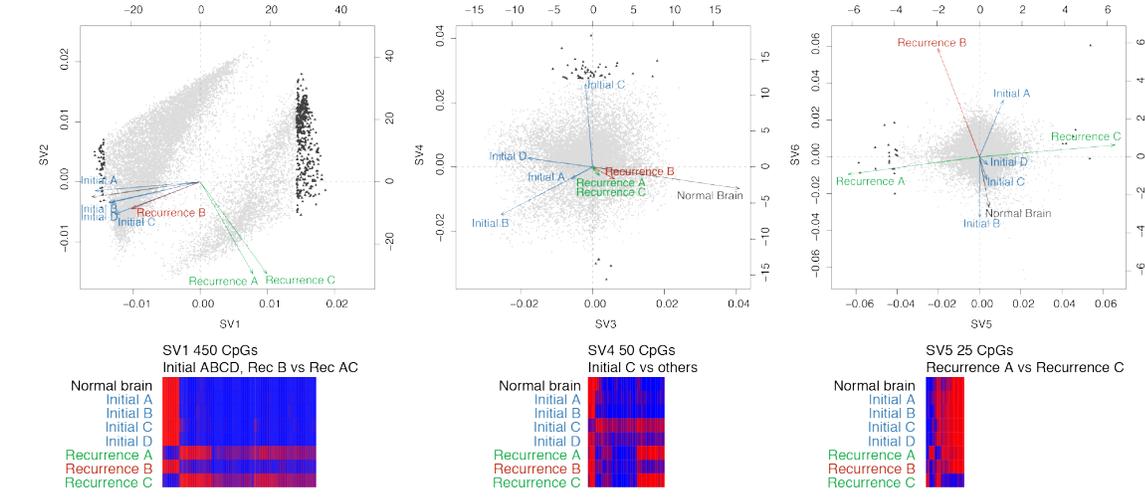
tree derived from somatic mutations from exome sequencing of the same DNA samples (right) (Spearman's $\rho = 0.90$). Tumor grade is provided in parentheses after each sample name. **(B)** Singular value decomposition biplot shows the probes involved in separating tumor samples. Each probe used to build the phyloepigenetic tree in (A) is plotted (grey dots). The most highly weighted probes are highlighted (triangles). **(C)** A heatmap of the beta values at the 220 probes most highly weighted by SV1. **(D)** A phyloepigenetic tree (left) and a phylogenetic tree (right) were constructed to infer the evolutionary relationships within and between the initial and recurrent tumors of Patient01 (Spearman's $\rho = 0.83$). Tumor grade is provided in parentheses after each sample name. **(E-G)** Phyloepigenetic (top) and phylogenetic trees (bottom) for Patient 18 (E, Spearman's $\rho = 0.90$), Patient90 (F, Spearman's $\rho = 0.56$) and Patient49 (G, Spearman's $\rho = 0.64$). Tumor grade is provided in parentheses after each sample name.

Figure 3.6

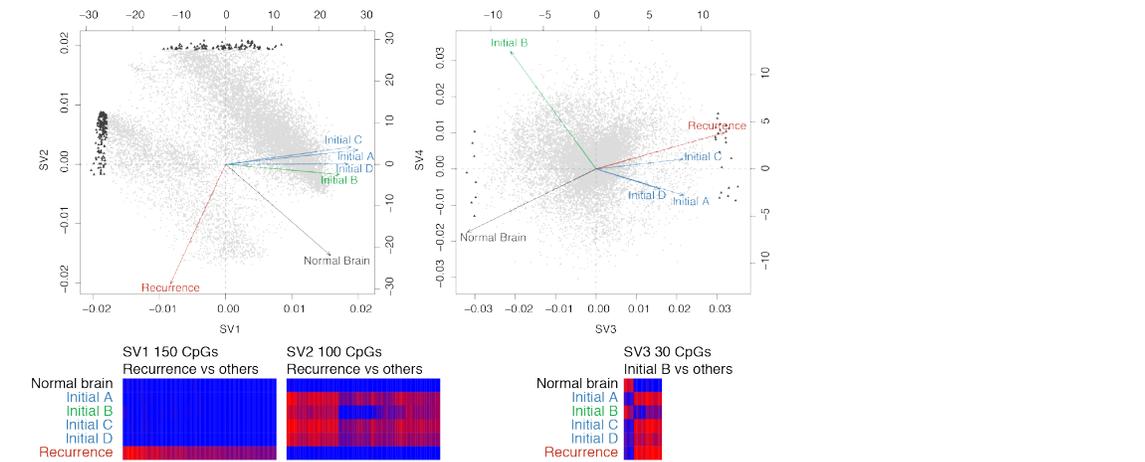
A Patient17



B Patient01



C Patient18



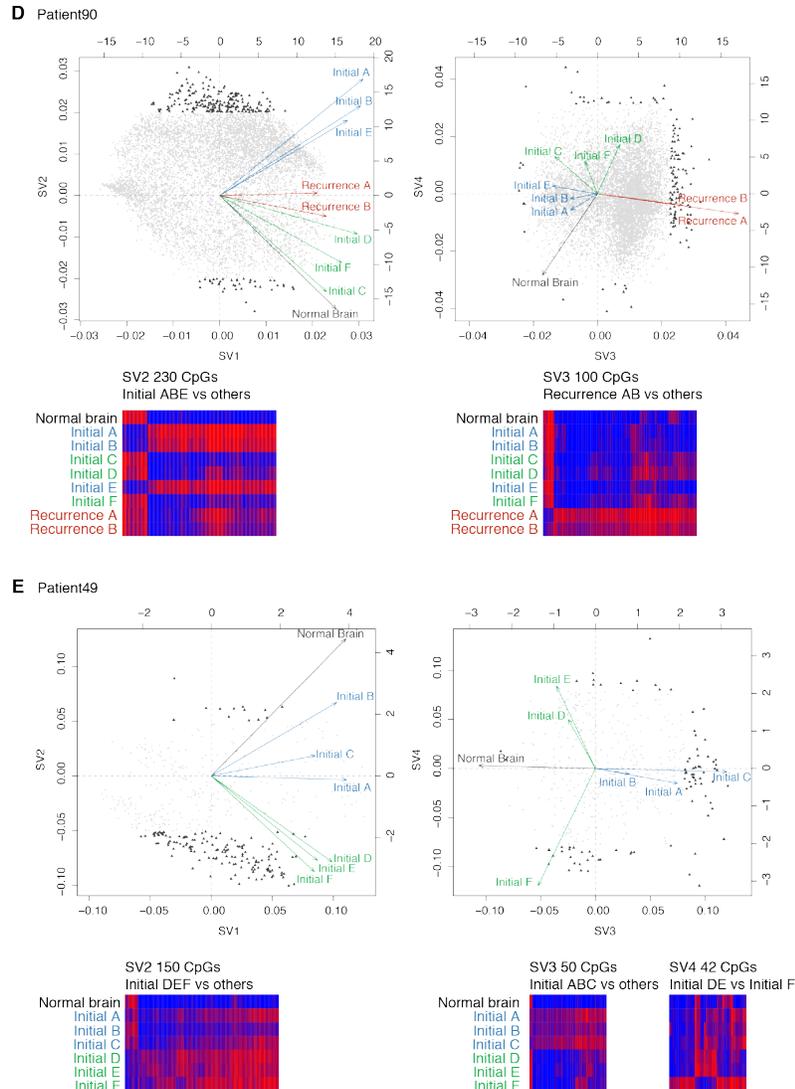


Figure 3.6. Plots supporting Figure 3.5 (A-E) Singular value decomposition biplots show the probes involved in separating tumor samples for Patient17 (A), Patient01 (B), Patient18 (C), Patient90 (D) and Patient49 (E). Each probe used to build the phyloepigenetic tree is plotted (grey dots). The most highly weighted probes for each selected SV are highlighted as black triangles. Below each biplot, a heatmap shows the beta values at the most highly weighted probes.

Figure 3.7

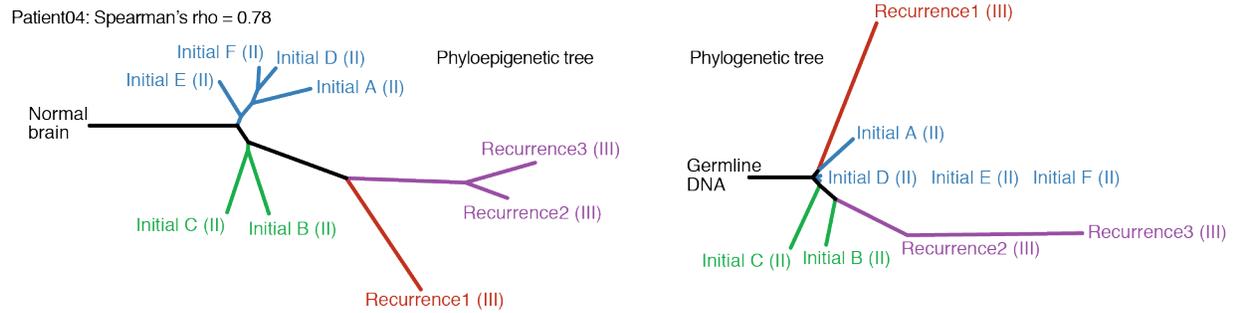


Figure 3.7. Phyloepigenetic trees coupled with phylogenetic trees from a low-grade glioma patient with three recurrences reveal an enhanced understanding of evolutionary relationships. Phyloepigenetic (left) and phylogenetic (right) trees of Patient04 present evolutionary relations across four surgical time points (Spearman's rho = 0.78). Tumor grade is provided in parentheses after each sample name.

Figure 3.8

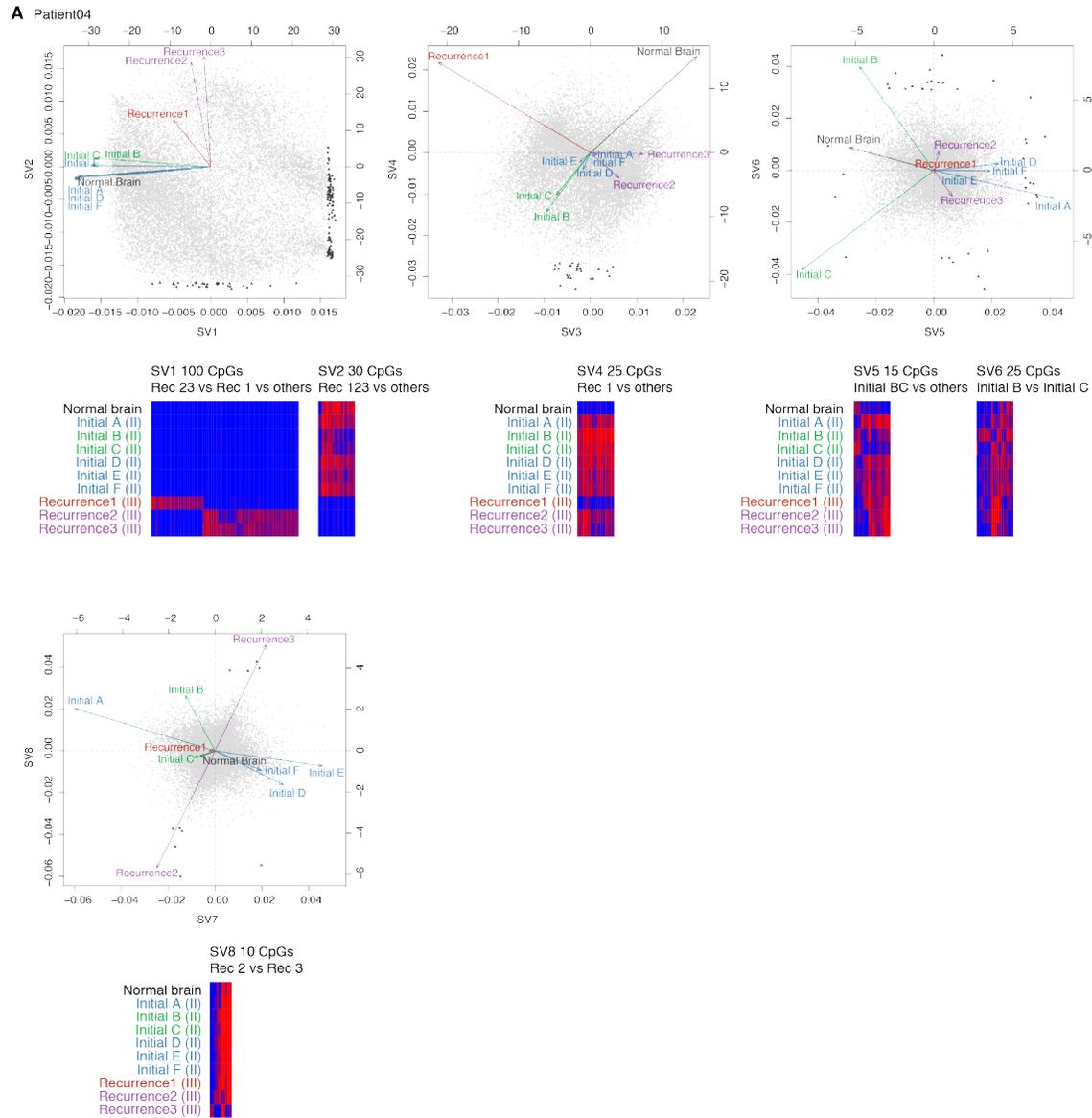


Figure 3.8. Plots supporting Figure 3.7 Singular value decomposition biplots show the probes involved in separating tumor samples for Patient04. Each probe used to build the phyloepigenetic tree is plotted (grey dots). The most highly weighted probes for each selected SV are highlighted as black triangles. Below each biplot, a heatmap shows the beta values at the most highly weighted probes.

Figure 3.9

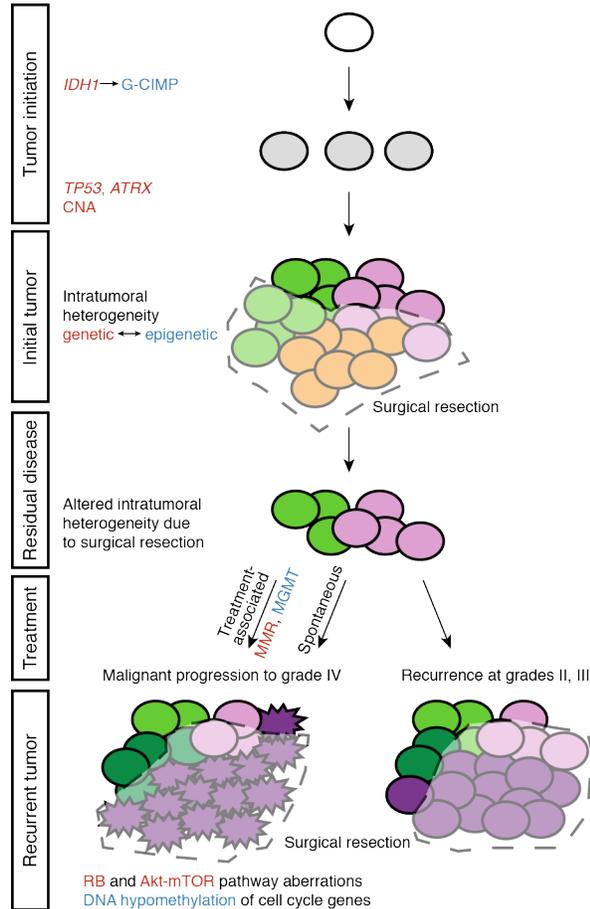


Figure 3.9. A genomic and epigenomic co-dependency model of clonal evolution. Low-grade gliomas exhibit intratumoral heterogeneity at initial presentation, with subclones that share the initiating genetic (*IDH1* followed by *TP53* and *ATRX* and copy number alterations, CNA) and epigenetic (*IDH1*-associated glioma CpG island methylator phenotype, G-CIMP) alterations, but further develop distinct genetic and epigenetic characteristics. Following surgical resection, the outgrowth from residual disease may be grade II or III, while still continuing to evolve subclones with genetic and co-dependent epigenetic features that are distinct from the initial tumor. In other patients, residual disease may undergo malignant progression to GBM, either spontaneously or as a consequence of treatment-associated mutations, in either case acquiring genetic defects in the RB and Akt-mTOR pathways and promoter hypomethylation and activation of cell cycle

genes. Treatment associated progression to GBM is uniquely associated with an increased epigenetic silencing of MGMT (*18*) and acquisition of genetic defects in mismatch repair genes.

Table 3.1

Patient	Gender	Age at diagnosis	Tumor sample	Diagnosis (WHO grade)	Surgical interval (months)	Non-surgical treatment (months)	Overall survival (months)	IDH1 status	1p19q status	Illumina 405K array	RNA-seq
01	Male	28	Initial tumor	Astrocytoma (II)	31	TMZ (14) ^b	58	R132H	intact	yes	yes
			Recurrence	Glioblastoma (IV)				R132H	intact	yes	yes
02	Female	26	Initial tumor	Oligoastrocytoma (II)	74	None	79 ^c	R132H	intact	yes	no
			Recurrence	Anaplastic astrocytoma (III)				R132H	intact	yes	no
03	Female	28	Initial tumor	Astrocytoma (II)	76	TMZ (7), TMZ (11)	85	R132H	intact	yes	yes
			Recurrence	Glioblastoma (IV)				R132H	intact	yes	yes
04	Male	22	Initial tumor	Astrocytoma (II)	15	None	61	R132C	intact	yes	yes
			Recurrence 1	Anaplastic astrocytoma (III)				R132C	intact	yes	yes
			Recurrence 2	Anaplastic astrocytoma (III)				R132C	intact	yes	no
			Recurrence 3	Anaplastic astrocytoma (III)				R132C	intact	yes	no
07	Male	30	Initial tumor	Astrocytoma (II)	105	XRT (1)	148	R132H	intact	yes	yes
			Recurrence	Astrocytoma (II)				R132H	intact	yes	yes
08	Male	44	Initial tumor	Oligoastrocytoma (II)	40	None	103	R132H	intact	yes	no
			Recurrence	Glioblastoma (IV)				R132H	intact	yes	no
10	Female	41	Initial tumor	Astrocytoma (II)	25	TMZ (9)	44	R132H	intact	yes	yes
			Recurrence	Glioblastoma (IV)				R132H	intact	yes	yes
11	Female	30	Initial tumor	Oligoastrocytoma (II)	132	XRT (1), TMZ (26)	186	R132H	intact	yes	no
			Recurrence	Oligoastrocytoma (II)				R132H	intact	yes	no
12	Male	35	Initial tumor	Astrocytoma (II)	17	None	82 ^d	R132H	intact	yes	yes
			Recurrence	Anaplastic astrocytoma (III)				R132H	intact	yes	yes
13	Male	24	Initial tumor	Oligoastrocytoma (II)	21	None	106	R132G	intact	yes	yes
			Recurrence	Oligoastrocytoma (III)				R132G	intact	yes	yes
14	Male	25	Initial tumor	Astrocytoma (II)	30	None	149 ^d	R132H	intact	yes	yes
			Recurrence	Astrocytoma (II)				R132H	intact	yes	yes
16	Female	35	Initial tumor	Astrocytoma (II)	5 ^a	None	38	R132H	intact	yes	yes
			Recurrence	Astrocytoma (II)				R132H	intact	yes	yes
17	Male	27	Initial tumor	Oligodendroglioma (II)	30	TMZ (12)	59 ^d	R132H	intact	yes	yes
			Recurrence	Glioblastoma (IV)				R132H	intact	yes	yes
18	Male	49	Initial tumor	Oligoastrocytoma (II)	94	TMZ (11)	106 ^d	R132H	intact	yes	no
			Recurrence	Glioblastoma (IV)				R132H	intact	yes	no
22	Male	22	Initial tumor	Astrocytoma (II)	56	XRT (1)	70	R132H	intact	yes	yes
			Recurrence	Glioblastoma (IV)				R132H	intact	yes	yes
36	Female	31	Initial tumor	Astrocytoma (II)	71	None	73 ^d	R132H	intact	yes	no
			Recurrence	Anaplastic astrocytoma (III)				R132H	intact	yes	no
37	Male	31	Initial tumor	Oligoastrocytoma (II)	57	None	105 ^d	R132H	intact	no	yes
			Recurrence	Anaplastic oligoastrocytoma (III)				R132H	intact	no	yes
38	Female	21	Initial tumor	Astrocytoma (II)	20	None	25 ^d	R132H	intact	yes	yes
			Recurrence	Astrocytoma (II)				R132H	intact	yes	yes
49	Male	23	Initial tumor	Anaplastic oligodendroglioma (III)			14 ^d	R132H	codel	yes	no
68	Female	31	Initial tumor	Oligoastrocytoma (II)	17	TMZ (12)	23 ^d	R132H	intact	yes	no
			Recurrence	Anaplastic oligoastrocytoma (III)				R132H	intact	yes	no
90	Female	39	Initial tumor	Oligodendroglioma (II)	22	None	64 ^d	R132H	intact	yes	no
			Recurrence 1	Oligoastrocytoma (II)				34	None	R132H	intact

^a Recurrent surgery for residual disease, no evidence of radiographic progression

^b Including a month each of TMZ plus either Accutane or Thalidomide

^c Patient lost to follow-up

^d Patient alive

Table 3.1. Summary of the data types, clinical and molecular features of each tumor in the cohort

Table 3.2

data type	patient	sample type	total_reads	pass filter reads	aligned reads	duplication rate	number of lanes	Coverage (genome)	Coverage (CpG)	avg CpC conversion*	avg CpC conversion	lambda phage*	bait set	mean target coverage	% target bases 10X coverage	% target bases 30X coverage
RNA-seq	Patient01	Initial	123,924,953	123,171,115	116,989,098											
RNA-seq	Patient01	Recurrence	166,883,463	164,772,191	154,800,636											
RNA-seq	Patient03	Initial	147,201,293	145,906,915	137,957,083											
RNA-seq	Patient03	Recurrence	236,677,850	232,112,330	217,068,889											
RNA-seq	Patient04	Initial	167,168,530	166,084,546	141,513,388											
RNA-seq	Patient04	Recurrence	227,722,835	226,145,054	209,067,541											
RNA-seq	Patient07	Initial	199,736,768	199,011,577	184,971,747											
RNA-seq	Patient07	Recurrence	195,524,400	194,446,327	178,276,572											
RNA-seq	Patient10	Initial	177,970,142	177,041,115	158,630,888											
RNA-seq	Patient10	Recurrence	200,867,603	199,893,088	182,338,697											
RNA-seq	Patient12	Initial	178,012,434	177,071,050	155,542,029											
RNA-seq	Patient12	Recurrence	201,635,312	200,761,033	186,646,082											
RNA-seq	Patient13	Initial	197,483,106	196,595,752	183,918,468											
RNA-seq	Patient13	Recurrence	198,338,953	197,684,064	184,153,218											
RNA-seq	Patient14	Initial	178,664,250	177,616,588	158,626,487											
RNA-seq	Patient14	Recurrence	176,773,591	175,959,863	160,986,786											
RNA-seq	Patient16	Initial	219,645,663	218,389,856	201,549,578											
RNA-seq	Patient16	Recurrence	212,100,720	210,030,959	187,154,078											
RNA-seq	Patient17	Initial	218,416,773	217,159,274	203,966,280											
RNA-seq	Patient17	Recurrence	218,269,894	217,362,179	201,548,795											
RNA-seq	Patient22	Initial	293,208,305	214,887,275	159,439,445											
RNA-seq	Patient22	Recurrence	294,511,318	260,841,313	223,553,223											
ChIP-seq	GBM01	H3K4me3	54,948,899	47,335,427	28,357,473	0.401										
ChIP-seq	GBM01	H3K4me1	69,763,200	61,766,081	54,764,288	0.113										
ChIP-seq	GBM01	H3K27Ac	54,932,406	46,007,279	39,185,972	0.148										
ChIP-seq	GBM01	Input	68,292,233	55,286,557	48,595,209	0.121										
ChIP-seq	GBM02	H3K4me3	52,140,050	34,684,499	24,625,383	0.290										
ChIP-seq	GBM02	H3K4me1	95,979,196	80,753,283	74,806,291	0.074										
ChIP-seq	GBM02	H3K27Ac	87,286,516	59,692,832	53,612,453	0.102										
ChIP-seq	GBM02	Input	58,980,778	49,218,006	33,163,322	0.326										
ChIP-seq	GBM03	H3K4me3	92,323,976	71,232,629	36,767,363	0.484										
ChIP-seq	GBM03	H3K4me1	92,987,428	81,926,356	76,983,287	0.060										
ChIP-seq	GBM03	H3K27Ac	79,579,916	63,513,796	57,855,398	0.089										
ChIP-seq	GBM03	Input	59,823,120	48,949,192	47,381,335	0.032										
ChIP-seq	GBM04	H3K4me3	52,609,302	43,095,282	29,792,539	0.309										
ChIP-seq	GBM04	H3K4me1	92,265,478	75,900,287	70,665,228	0.069										
ChIP-seq	GBM04	H3K27Ac	55,631,920	39,930,644	36,287,616	0.091										
ChIP-seq	GBM04	Input	62,087,390	51,614,061	48,702,088	0.056										
ChIP-seq	Anterior_Caudate	H3K4me3			35,616,706											
ChIP-seq	Anterior_Caudate	H3K4me1			39,992,528											
ChIP-seq	Anterior_Caudate	H3K27Ac			26,285,811											
ChIP-seq	Anterior_Caudate	Input			32,718,096											
ChIP-seq	Cingulate_Gyrus	H3K4me3			34,477,121											
ChIP-seq	Cingulate_Gyrus	H3K4me1			34,535,603											
ChIP-seq	Cingulate_Gyrus	H3K27Ac			34,537,642											
ChIP-seq	Cingulate_Gyrus	Input			35,971,424											
ChIP-seq	Mid_Frontal_Lobe	H3K4me3			37,296,152											
ChIP-seq	Mid_Frontal_Lobe	H3K4me1			37,339,589											
ChIP-seq	Mid_Frontal_Lobe	H3K27Ac			34,362,604											
ChIP-seq	Mid_Frontal_Lobe	Input			36,503,506											
ChIP-seq	Hippocampus_Middle	H3K4me3			35,625,648											
ChIP-seq	Hippocampus_Middle	H3K4me1			34,257,403											
ChIP-seq	Hippocampus_Middle	H3K27Ac			22,634,126											
ChIP-seq	Hippocampus_Middle	Input			34,233,676											
ChIP-seq	Inferior_Temporal_Lobe	H3K4me3			31,124,677											
ChIP-seq	Inferior_Temporal_Lobe	H3K4me1			31,418,935											
ChIP-seq	Inferior_Temporal_Lobe	H3K27Ac			27,823,926											
ChIP-seq	Inferior_Temporal_Lobe	Input			27,992,918											
WGBS	Patient01	Initial	1,157,558,570		746,445,946		6	23x	23	0.990	0.994					
WGBS	Patient01	Recurrence	813,068,944		689,518,038		6	20x	12	0.992	0.998					
exome-seq	Patient01	Initial B	85,755,020	83,084,541	71,149,603	0.032							SeqCap	69	0.018	0.877
exome-seq	Patient01	Initial C	93,700,578	90,351,374	77,208,921	0.036							SeqCap	75	0.018	0.898
exome-seq	Patient01	Initial D	133,041,790	127,455,019	109,321,390	0.043							SeqCap	105	0.017	0.939
exome-seq	Patient01	Recurrence1 B	442,952,598	161,688,372	148,024,153	0.641							SeqCap	86	0.071	0.730
exome-seq	Patient01	Recurrence1 C	94,062,994	91,312,908	78,405,727	0.030							SeqCap	76	0.017	0.891
exome-seq	Patient04	Initial B	208,661,984	162,375,158	148,608,674	0.224							Agilent	137	0.031	0.863
exome-seq	Patient04	Initial C	236,030,112	174,833,241	160,344,090	0.262							Agilent	147	0.030	0.873
exome-seq	Patient04	Initial D	229,159,672	195,519,016	179,177,460	0.148							Agilent	167	0.031	0.875
exome-seq	Patient04	Initial E	220,332,424	191,769,242	175,565,569	0.131							Agilent	160	0.030	0.872
exome-seq	Patient04	Initial F	187,373,578	153,497,653	140,027,285	0.182							Agilent	128	0.031	0.856
exome-seq	Patient49	Normal	97,394,618	93,881,911	80,570,083	0.037							SeqCap	77	0.018	0.913
exome-seq	Patient49	Initial A	84,894,514	81,794,598	70,422,613	0.037							SeqCap	68	0.018	0.880
exome-seq	Patient49	Initial B	310,905,878	297,567,270	258,061,290	0.044							SeqCap	244	0.015	0.969
exome-seq	Patient49	Initial C	100,398,442	96,468,453	83,511,385	0.040							SeqCap	77	0.017	0.910
exome-seq	Patient49	Initial D	91,382,174	88,172,249	76,021,465	0.036							SeqCap	72	0.018	0.888
exome-seq	Patient49	Initial E	85,821,408	82,664,659	71,339,302	0.038							SeqCap	69	0.018	0.884
exome-seq	Patient49	Initial F	113,366,324	108,940,069	94,353,472	0.040							SeqCap	91	0.018	0.927
exome-seq	Patient90	Normal	93,718,592	90,355,271	78,191,682	0.037							SeqCap	75	0.019	0.915
exome-seq	Patient90	Initial A	92,244,704	88,801,130	76,273,833	0.038							SeqCap	72	0.019	0.881
exome-seq	Patient90	Initial B	92,771,024	89,240,891	76,795,203	0.039							SeqCap	73	0.019	0.889
exome-seq	Patient90	Initial C	97,949,538	94,392,106	81,426,584	0.037							SeqCap	77	0.019	0.917
exome-seq	Patient90	Initial D	101,963,298	98,706,431	85,434,106	0.033							SeqCap	81	0.018	0.925
exome-seq	Patient90	Initial E	108,390,990	104,441,391	90,177,238	0.037							SeqCap	88	0.019	0.924
exome-seq	Patient90	Initial F	98,222,164	94,683,292	81,856,931	0.037							SeqCap	78	0.019	0.920
exome-seq	Patient90	Recurrence1 A	89,614,982	86,834,177	75,239,724	0.032							SeqCap	74	0.020	0.898
exome-seq	Patient90	Recurrence1 B	94,134,364	90,906,985	78,852,256	0.035							SeqCap	74	0.019	0.910

* calculated per lane and averaged

Table 3.2. QC of all sequencing datasets

Table 3.3

Intra-patient

	IlmnID	gene_name	t_stat	pval	Initial A	Initial B	Initial C	Rec1 A	Rec1 B	Rec1 C	Rec1 D	
Patient17	cg02995664	TMEM131	16.32875	1.21E-07	0.52	0.51	0.53	0.14	0.11	0.11	0.09	
Patient17	cg02152120	SPAG17	-14.23887	3.68E-07	Initial A	Initial B	Initial C	Rec1 A	Rec1 B	Rec1 C	Rec1 D	
Patient17	cg03850057	SPAG17	9.499175	9.09E-06	0.20	0.13	0.12	0.52	0.50	0.43	0.56	
Patient90	cg04912843	GIPC2	5.086217	2.97E-04	Initial A	Initial B	Initial C	Initial D	Initial E	Initial F	Rec1 A	Rec1 B
					0.57	0.52	0.07	0.04	0.61	0.09	0.25	0.18

Inter-patient

		Epigenetics of:			
		Patient04	Patient17	Patient49	Patient90
Genetics of:	Patient04	-	MGAT4C, PDZD2, ACSF2, ARHGEF17, SPATA6, GULP1, HPS4, LAMA3, APOBEC4, WDFY4	SPATA6	TEAD3, HSP90AB1, MGAT4C, OR10V1, ZP3, GULP1, APOBEC4, CCR4, LAMA3
	Patient17	ABCC9, TSHB, SPAG17, TMPRSS11B, NOTCH4, SP6, UGT3A1, GPR142, SLC4A3, WFDC12, SLC22A25, C7orf10, TLN2	-	SPAG17, TMEM63B	SP6, LRRC16B, GPR142, SPAG17, SDAD1, CDHR3, NOTCH4, ABCC9, SLC22A25
	Patient49	NAV2, SHE, OXTR, LCE1D, RAPGEF6, VRTN	ST3GAL2, TLR4, ALOX5, RASD2, SHE, NAV2	-	TLR4, NAV2
	Patient90	AJAP1, ERICH1, F13A1, EXPH5, KLF14, CRYBB3, MEGF6, HOXA6, APBB1IP, CPEB4, COL12A1, HNF4A, RYR1, C1orf158, PLS1, RGMA, REST	IL1RAP, KLF14, RAC3, COL12A1, F13A1, APBB1IP, DNAH12, CASKIN2, PPP1R1B, HOXA6, EXPH5, AJAP1, MEGF6, PCNXL2, GIPC2, REST	MEGF6	-

Table 3.3. Intra- and inter-patient gene-level convergence

CHAPTER 4:
EXTREME DNA HYPOMETHYLATION AND TUMOR PROGRESSION
FOLLOWING DELETION OR AMPLIFICATION OF MUTANT *IDH1* IN GLIOMA

4.1 ABSTRACT

Heterozygous mutations in *IDH1* are the most common genetic alteration in low-grade gliomas (LGGs) (154). The neomorphic mutant IDH1 enzyme produces the putative oncometabolite 2-hydroxyglutarate (2HG) (76, 77), which leads to alterations in DNA methylation (glioma CpG island methylator phenotype, G-CIMP) (78, 79, 82) and histone methylation (80, 81). While mutation in *IDH1* is believed to initiate tumorigenesis, its role in sustained tumor growth is less clear. We therefore undertook a longitudinal analysis of copy number at the *IDH1* locus in a cohort of initially *IDH1*-mutant LGGs and their patient-matched recurrences and identified six cases in which the ratio of the mutant to wild-type allele was altered. With either deletion or amplification of the mutant allele, 2HG levels were consistently decreased. While G-CIMP status was uniformly maintained, several cases exhibited a signature of extreme DNA hypomethylation outside of CpG islands (CGIs). Deletion of the mutant *IDH1* allele followed by clonal expansion suggests that sustained growth of initially *IDH1*-mutant tumors can proceed in the absence of both the mutant IDH1 protein and high levels of 2HG. These results have implications for targeted inhibition of mutant IDH1 and provide *in vivo* evidence for differentiating between drivers of tumor initiation and sustained tumor growth.

4.2 MAIN TEXT

Mutations in *IDH1* or *IDH2* (collectively: *IDH*) are the earliest known alterations in LGG (104, 126) and mark the subset of adult glioma cases with the best outcomes (7, 8). These mutations are heterozygous allowing for hetero-dimerization of wild-type and mutant protein for maximal production of 2HG (155-158). 2HG is a competitive inhibitor of α -ketoglutarate (α KG) dependent enzymes including TET2 and histone demethylases (79, 80). We and others have

previously shown that *IDH* mutations are retained upon tumor recurrence (104, 126, 159), supporting the potential of inhibiting mutant IDH for therapeutic benefit (84, 160). However, several case reports suggest that both the mutant and wild-type allele can be lost during tumor progression as well as in culture (161-164). We therefore set out to discover the role of *IDH* mutations at tumor recurrence and downstream consequences of alterations at *IDH*.

We performed exome-sequencing on an expanded cohort of 50 paired initial LGGs and their patient-matched recurrences (18, 126, 165). All of the initial tumors in this cohort had mutations in either *IDH1* or *IDH2*. However, there were two recurrences from *IDH1*-mutant initial tumors (Patient14 Recurrence2 A and Patient169 Recurrence1 A) in which our exome pipeline did not call an *IDH1* mutation. In each case, several other mutations were shared with the earlier tumors (Figure 4.1A), suggesting that these were outgrowths of the earlier tumors that deleted the mutant allele of *IDH1* rather than being *de novo* *IDH*-wild-type gliomas. More sensitive analysis of the exome-sequencing data showed that these tumors did have evidence of an *IDH1* mutation, but the variant frequency was below the 10% minimum required by our mutation calling pipeline (Table 4.1).

To address the suspected copy number change underlying the loss of the *IDH1* mutation in these cases, we performed a loss of heterozygosity (LOH) analysis and a read-density based analysis of total copy number (TCN). For both recurrent tumors, there was a decrease in heterozygosity, suggesting an allelic imbalance, which was confirmed as single copy loss in the TCN analysis (Figures 4.1B and 4.2A). We next undertook a more comprehensive analysis of this cohort by looking for LOH spanning the *IDH1* or *IDH2* locus in all tumors. Through this analysis, we identified four additional cases with allelic imbalance at the *IDH1* locus in *IDH1*-mutant tumors (Table 4.1 and Figures 4.2B-E). However assessment of CNV in these four cases

was inconclusive; while two cases (Patients 17 and 27) clearly showed gain at the *IDH1* locus, the other two cases (Patients 21 and 68), which both showed LOH across the entirety of chromosome 2, showed no copy number change. This could reflect copy-neutral-LOH or a polyploid genome. While Patient27 showed evidence of the alteration at initial presentation, the remainder acquired these alterations later and presented with aberrant copy number at recurrence only. We did not identify LOH at the *IDH2* locus in the two *IDH2*-mutant cases in this cohort.

These exome-sequencing-based analyses were limited to single samples of each tumor and thus represent only a fraction of the cells that make up each tumor. We therefore set out to investigate how well these single samples represented the tumor bulk. We had previously published exome-seq data for additional, spatially distinct samples of Patients 17 (Initial BC, Recurrence1 BCD) and we were able to acquire additional, spatially distinct pieces of tissue from Patients 14 (Recurrence2 B), 68 (Recurrence1 B) and 169 (Initial B) that we profiled by exome-sequencing. These additional pieces replicated our initial findings (Figure 4.2), suggesting that these alterations are present throughout the recurrent tumor.

To address the clonality of these changes in a larger number of samples per tumor, and to determine the impact of these copy number changes on protein levels, we performed immunohistochemistry (IHC) with an antibody specific to R132H-mutant IDH1 on all available formalin-fixed paraffin-embedded (FFPE) blocks (Table 4.2). Unfortunately, the data for Patient27 was acquired from an outside institution so that case is excluded from all further analyses. We found that all of these genetic alterations manifest at the protein level, ranging from the complete absence of mutant IDH1 protein following deletion of the mutant allele in Patient14 Recurrence2 to dramatically increased staining intensity in Patient21 Recurrence1 (Figures 4.3, 4.4 and Table 4.2).

The differential intensity of staining by IHC supported the LOH results that both Patients 21 and 68 did have copy number aberrations affecting *IDH1*, so we set out to conclusively determine the copy number state of *IDH1* with fluorescence in situ hybridization (FISH). We performed FISH at the *IDH1* locus in all cases to confirm the exome-based copy number (Figure 4.3 and Table 4.2). Patients 21 and 68 both showed either subclonal (Patient68) or clonal (Patient21) gain of chromosome 2. Combined with the earlier exome-based copy number analysis, this suggested a polyploidy genome. To confirm this, we performed FISH with probes for the centromeres of chromosomes 2, 3 and 15 and confirmed that cells in these tumors had additional copies of multiple chromosomes. As a result of these analyses, we identified two cases of deletion of the mutant allele (Patients 14 and 169), one case with subclonal gain of the wild type chromosome (Patient68), one case with subclonal gain of the mutant allele (Patient17) and one case with gain of the mutant chromosome with additional focal gain of the *IDH1* locus (Patient21) (Table 4.1).

We next set out to understand the downstream consequences of these copy number changes. Mutant IDH1 generates 2HG, leading to cellular accumulation of 2HG and genome-wide changes to DNA methylation and histone modifications (76-78, 80). Maximal 2HG production requires both mutant and wild-type IDH1 protein (157, 158, 162), suggesting that the tumors with copy number change at the *IDH1* locus, including both loss or gain of the mutant allele, should lead to decreased production of 2HG. We used NMR analysis of flash-frozen tumor tissue to calculate the concentration of 2HG and subsequently performed exome-seq on the same piece of tissue. However, due to tissue availability we were only able to profile a subset of cases in this analysis. In two samples from Patient14 Recurrence2 we did not detect 2HG, as expected following loss of the mutant allele. Patient21 Recurrence1 had decreased concentration

of 2HG, supporting *in vitro* studies that an abundance of mutant IDH1 protein relative to wild type can actually decrease 2HG production (157, 158, 162). Patient17 Recurrence1 had a normal 2HG concentration, likely due to the subclonal nature of the copy number alteration. However, the initial tumor from Patient169 surprisingly showed little to no detectable 2HG. This is inconsistent with the subsequent exome-sequencing of those pieces of tissue (Initial B, C), which indicates an intact chromosome (Figure 4.2F). IHC for mutant IDH1 was negative in some regions of that tumor, consistent with the 2HG results, suggesting that the deletion of the mutant allele already occurred by the time of the initial surgery for this patient. For all other tumors, the results of exome-sequencing on the post-NMR tissue was consistent with all other samples of that tumor (Figure 4.2 and Table 4.2)

Mutant IDH1 and 2HG have both been shown to slow cellular proliferation *in vitro* (157, 166) and it has been theorized that this may be related to the slow growth of *IDH*-mutant gliomas (157). We asked whether the observed copy number alterations at *IDH1*, and subsequent decrease in 2HG, could produce a more proliferative tumor. To address this, we took three tumors with heterogeneous IDH1 R132H IHC staining and quantified the Ki67 index in each region. In all three cases, the region of tumor with abnormal IDH1 R132H staining (higher intensity following amplification in Patients 17 and 21; absent following deletion in Patient169) had a higher Ki67 index than the region with “normal” staining (consistent with retention of the heterozygous mutation) (Table 4.3). Taken together with previous studies, this suggests that reduction of 2HG may enable a more proliferative state.

Given the widespread DNA methylation changes associated with *IDH1* mutation, we next used the Infinium 450K array to investigate the DNA methylation profile of these tumors. *IDH1* mutations are associated with a CpG island (CGI) hypermethylation phenotype (G-CIMP). We

therefore set out to determine the G-CIMP status of the tumors with copy number changes at the *IDH1* locus. Surprisingly, all tumors retained G-CIMP based on both an eight-site minimal definition (78) (Figure 4.5A) and by clustering with *IDH* mutant and wild type tumors from TCGA (8, 167) (Figure 4.5B). However, both of these methods use a small fraction of CpG sites. For a more global view, we applied principal component analysis (PCA) to the TCGA samples and found that PC1 and PC2 together separated samples according to *IDH* mutation status (Figure 4.5C). Interestingly, when we plotted the *IDH1* copy number change cases, several of these tumors clustered with the *IDH*-wild type TCGA tumors (Figure 4.5C). This suggested that despite the retention of G-CIMP, these tumors have globally altered DNA methylation. Indeed, density plots and the average methylation per tumor both indicate a massive, global decrease in methylation in these tumors (Figure 4.6), well beyond the decrease that is characteristic in tumors that recur as GBM (165). Although small molecule inhibition of mutant IDH1 in glioma cell lines produced no change in DNA methylation (84), our data supports a large-scale change in methylation, possibly reflecting the substantial differences of *in vitro* models relative to patient tumor samples. This may also be a characteristic of specific inhibitors, as a recent study of a different IDH1 inhibitor in primary AML cells showed a similar pattern of non-CGI hypomethylation (168).

To further understand these methylation changes, we leveraged our cohort of paired tumors to investigate the average change in methylation from the initial tumor to the *IDH*-copy number changed recurrence (for Patient14, we compared Recurrence2 to Recurrence1). There was a clear loss of methylation following copy number change at *IDH1* (Figure 4.7A), even compared to the significant hypomethylation we discovered in GBM recurrences (165). Given the retention of G-CIMP, which indicates retained hypermethylation within CGIs, we next asked

if this hypomethylation was primarily found outside of CGIs. Indeed, it was clear that non-CGI CpG sites were the source of this hypomethylation (Figure 4.7B). While the pattern of hypomethylation was very strong in the two cases of deletion of the mutant allele (Patients 14 and 169) and the clonal gain of the mutant allele (Patient21), it was moderate in the case of a subclonal gain of the wild-type allele (Patient68) and not visible with subclonal gains of the mutant allele (Patient17). It is likely that the subclonal nature of the copy number changes in Patients 17 and 68 muted the signal of any aberrant hypomethylation.

The tumors from Patients 14, 169 and 21 show a strong hypomethylation signal, specifically from CpG sites outside of CGIs. If these changes are due to mechanistic effects from the decrease in 2HG as a consequence of copy number change at *IDH1*, then there should be a common set of CpG sites altered in these three recurrences. We calculated the change in methylation between all initial and recurrent tumors (for Patient14, we compared Recurrence1 to Recurrence2). Patients 14, 169 and 21 cluster together based on unsupervised hierarchical clustering (Figure 4.7C) or PCA (Figure 4.7D). We also calculated all pairwise correlations and found that the three highest pairwise correlations were among Patients 14, 169 and 21 (Figure 4.7E). Together, these analyses demonstrate that similar CpG sites are altered following copy number change at the *IDH1* locus suggesting a common effector of these changes.

In summary, this study identified six cases of copy number change encompassing the *IDH1* locus through a combination of exome sequencing, IHC and FISH. These tumors retain G-CIMP positivity and yet also show dramatic hypomethylation relative to their patient-matched initial tumors along with decreased 2HG. This hypomethylation is driven by CpG sites outside of CGIs, while CpG sites within CGIs show a more consistent level of DNA methylation compared to patient-matched initial tumors, consistent with retention of G-CIMP. Interestingly, global

hypomethylation is a consequence of both deletion of the mutant IDH1 allele as well as amplification of the mutant allele. Furthermore, the existence of tumor cells that lose the mutant *IDH1* allele and then clonally expand suggests that sustained growth of initially *IDH1* mutant tumors can in some cases proceed in the absence of the mutant IDH1 protein. These results may have implications for targeted inhibition of mutant IDH1. Our results further emphasize the importance of differentiating between mutations that drive tumor initiation and those that sustain tumor growth.

Figure 4.1

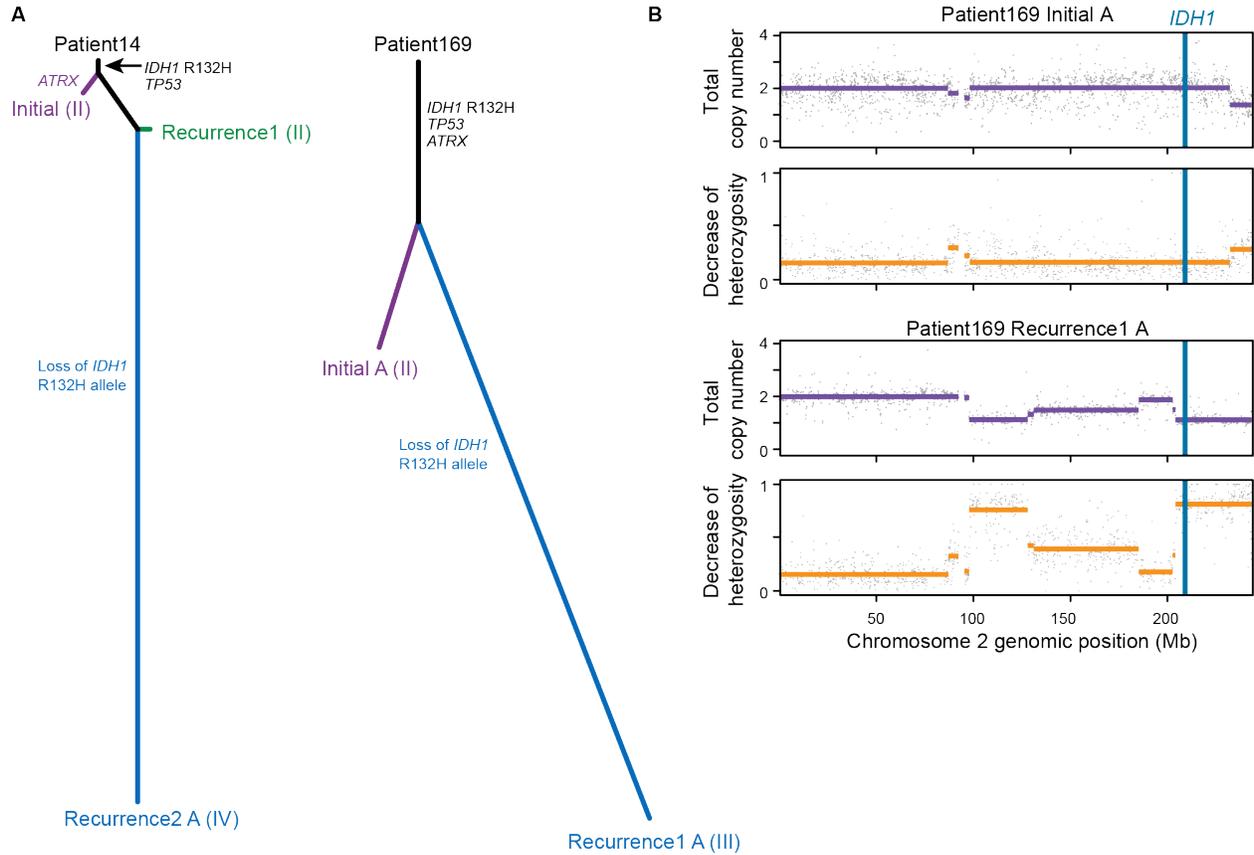
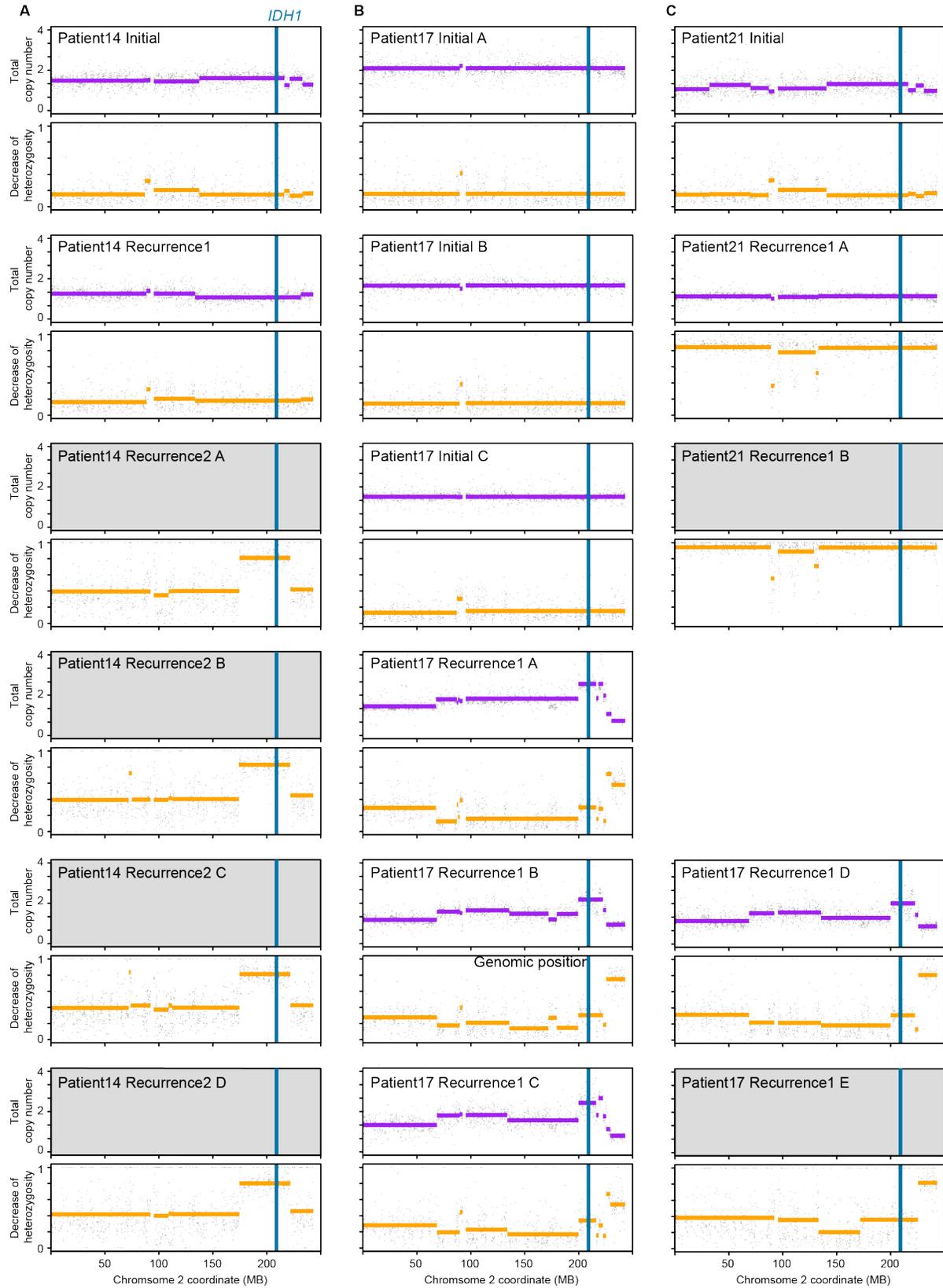


Figure 4.1. Identification of two cases with deletion of mutant *IDH1*. (A) Phylogenetic tree of Patients 14 (left) and 169 (right) showing shared mutations despite the absence of *IDH1* mutation at Patient14 Recurrence2 and Patient169 Recurrence1. (B) Decrease of heterozygosity (DoH) and total copy number (TCN) plots of chromosome 2 for Patient169 Initial (top) and Recurrence1 (bottom) show single copy loss spanning the *IDH1* locus at recurrence.

Figure 4.2



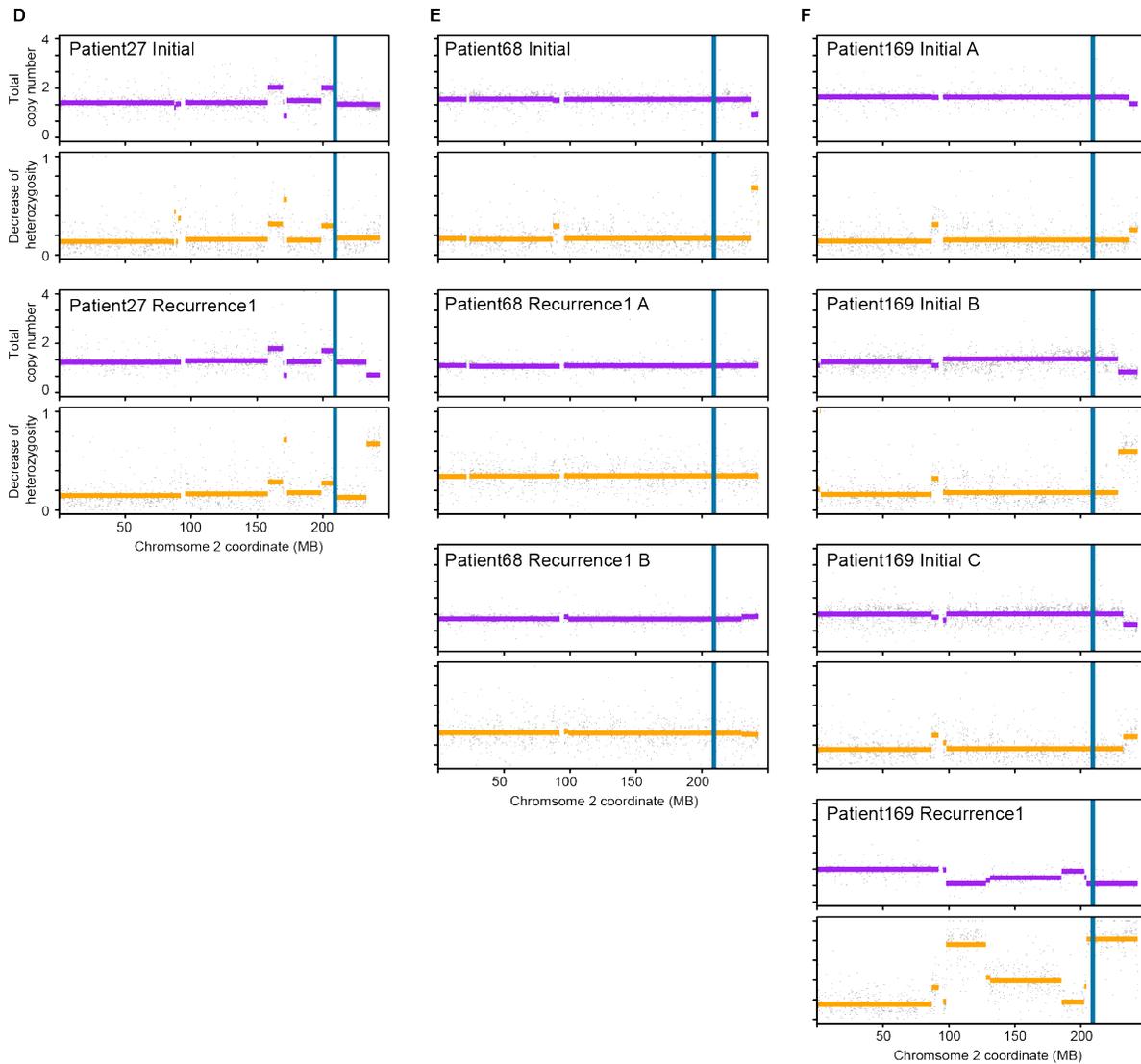


Figure 4.2. DoH and TCN plots for all cases with allelic imbalance at *IDH1*. Plots for (A) Patient14, (B) Patient17, (C) Patient21, (D) Patient27, (E) Patient68 and (F) Patient169. In grey are several exomes that were generated with a different exome kit than the patient-matched normal; in these cases, TCN analysis was inconclusive. 450K array based copy number analysis demonstrated single copy loss in Patient14 Recurrence2 ABCD, gain in Patient17 Recurrence1 E and no copy number change in Patient21 Recurrence1 B.

Figure 4.3

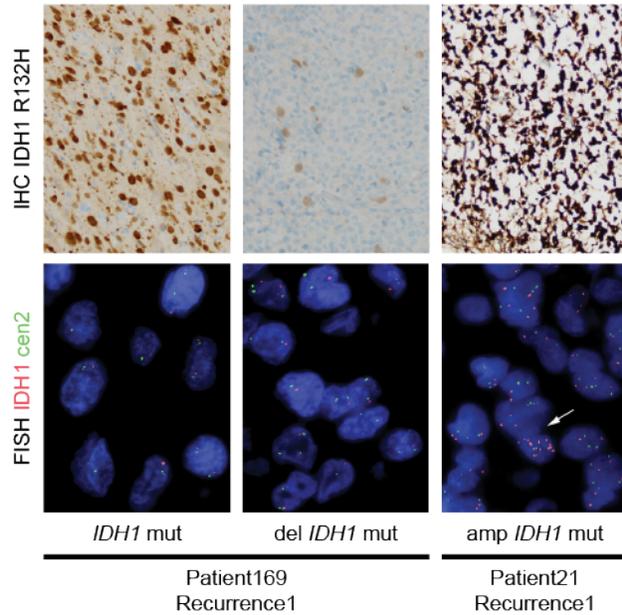


Figure 4.3. IHC and FISH from Patients 169 and 21. Representative images from IHC against R132H mutant IDH1 (top) and FISH for the centromere of chromosome 2 (green) and the *IDH1* locus (red) are shown for two distinct regions of Patient169 Recurrence1 (left, heterozygous *IDH1* mutation; center, deletion of mutant allele) and Patient21 Recurrence 1 (right, gain of chromosome 2 and the *IDH1* locus).

Figure 4.4

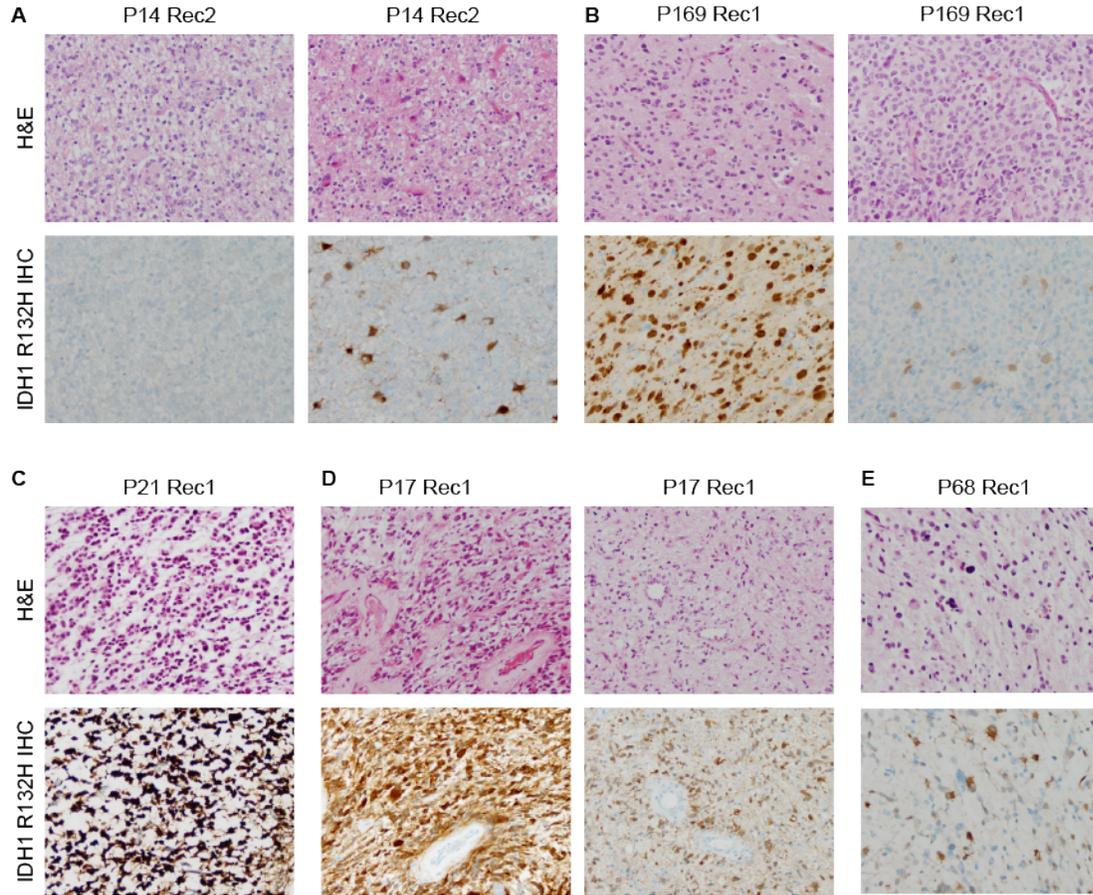


Figure 4.4. Representative images from H&E and IDH1 R132H IHC. (A) Patient14 Recurrence2 shows no positivity for IDH1 R132H across most of the tumor (left) but a handful of positive cells remain (right). (B) Patient169 Recurrence1 has two distinct tumor populations: some cells retain IDH1 R132H positivity (left) while other regions show almost complete absence (right). (C) Patient21 Recurrence1 shows hyper-intense staining for IDH1 R132H. (D) Patient17 Recurrence1 has two distinct tumor populations: some cells stain intensely for IDH1 R132H (left) while others have a more traditional staining intensity (right). (E) Patient68 Recurrence1 has overall weak staining for IDH1 R132H.

Figure 4.5

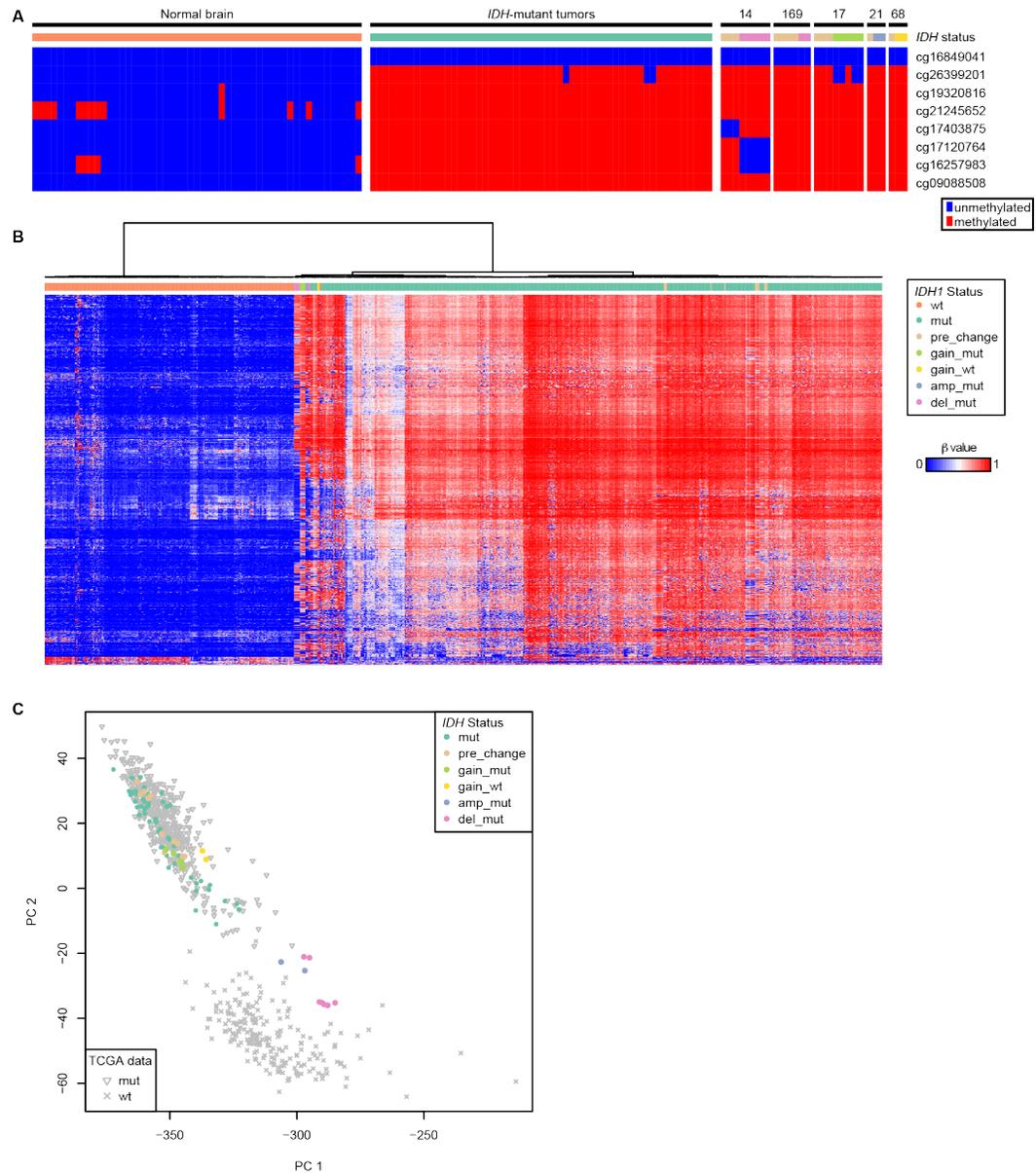


Figure 4.5. Classification of *IDH1* copy number change tumors relative to *IDH* mutant and wild type gliomas. (A) Determination of G-CIMP status based on 8 CpG sites. (B) Unsupervised hierarchical clustering with TCGA GBM and LGG samples (top 0.5% most variable by standard deviation). (C) PCA built on TCGA GBM and LGG samples and applied to this cohort of samples.

Figure 4.6

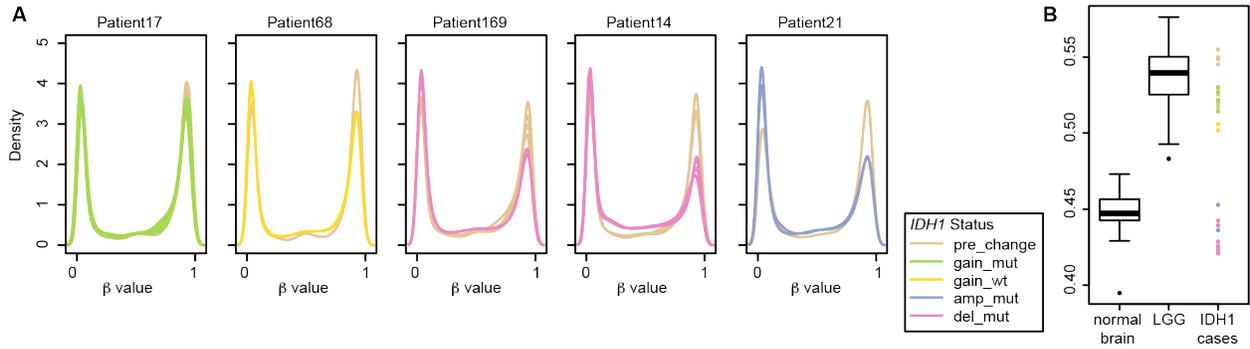


Figure 4.6. Average methylation levels following change in *IDH1* copy number. (A) Density plots of β values in all samples from the five patients with *IDH1* copy number changes. (B) Boxplot of average β value separated into normal brain, LGG and the *IDH1* copy number change tumors.

Figure 4.7

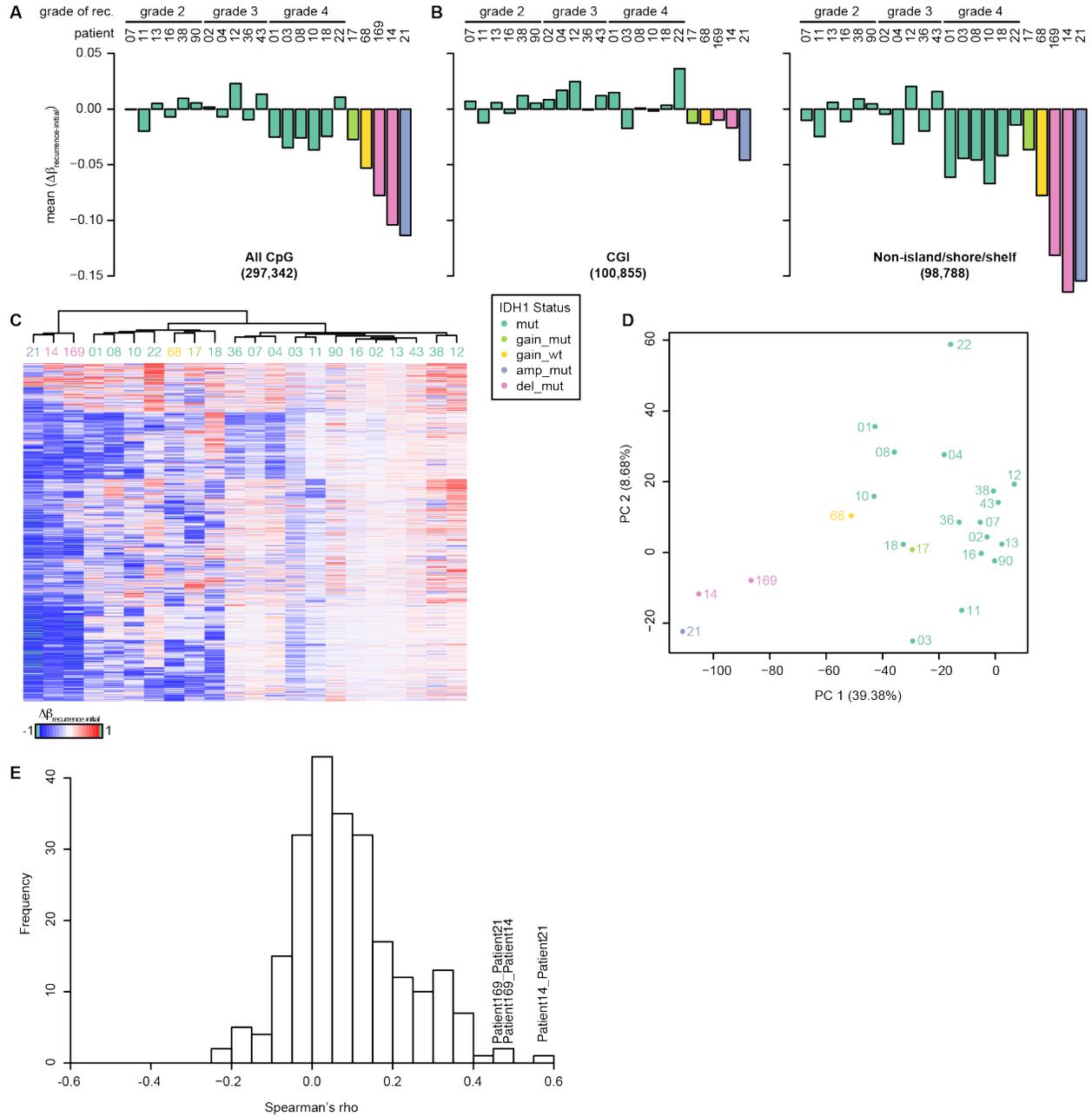


Figure 4.7. Similarity of methylation change following deletion or amplification of mutant *IDH1*. (A and B) The average change in methylation from initial to recurrence for each patient in the cohort, using all CpG sites (a) or only those within or outside CGIs (b). (C) Unsupervised hierarchical clustering of the change in methylation from initial to recurrence in each patient (top

1% most variable probes). Patient numbers are provided under the dendrogram. **(D)** PCA of the change in methylation from initial to recurrence in each patient. PC1 and PC2 are plotted with the percent of variation accounted for. **(E)** Histogram of all pair-wise correlations of methylation change from primary to recurrence for each patient. The top three most correlated pairs are labeled.

Table 4.1

	Patient	Surgery	Tumor grade	Sample	IDH1 variant frequency	Normalized 2HG area	Interpretation
less mutant	Patient14	Initial	2	A	0.33	-	heterozygous IDH1 R132H mutation
		Recurrence1	2	A	0.32	-	heterozygous IDH1 R132H mutation
	Recurrence2	4		A	0.03	-	deletion of mutant allele
				B	0	-	
				C	0.01	0	
				D	0	0	
	Patient169	Initial	2	A	0.44	-	heterozygous IDH1 R132H mutation
				B	0.45	-	
				C	0.36	0.0148	
	Recurrence1	3	A	0.07	-	deletion of mutant allele	
Patient68	Initial	2	A	0.56	-	heterozygous IDH1 R132H mutation	
	Recurrence1	3	A	0.38	-	gain of wt allele (subclonal)	
more mutant	Patient17	Initial	2	B	0.43		-
				C	0.42	-	
				A	0.54	-	
				B	0.59	-	
				C	0.56	-	
	Recurrence1	4	C	0.56	-	gain of mutant allele (subclonal)	
	D	0.61	-				
	E	0.68	2.88				
	Patient21	Initial	2	A	0.43	-	heterozygous IDH1 R132H mutation
		Recurrence1	4	A	0.71	-	gain of mutant allele + amp of mutant gene
Patient27	Initial	2	B	0.94	0.0516		
			A	0.78	-	gain of mutant allele (clonality unknown)	
			B	0.62	-		
			Recurrence1	3	A		0.48
B	0.69	-					

Table 4.1. Summary of patients identified and samples profiled with copy number change at *IDH1*

Table 4.2

	Patient	Surgery	number of slides examined by IHC	IDH1 R132H IHC	FISH (chr2, IDH1)	clonality assessment
less mutant	Patient14	Initial	5	positive	intact	
		Recurrence1	4	positive	intact	
		Recurrence2	7	negative; rare positive cells	hemizygous deletion	clonal hemizygous deletion
	Patient169	Initial	7	positive with interspersed negative	intact	
		Recurrence1	6	regions of positive and regions of negative	intact region and hemizygous deletion region	two distinct populations, one hemizygous deletion and one intact
	Patient68	Initial	11	positive	intact	
Recurrence1		6	intermixed positive and negative	intact; occasional cell with high level polysomy/gain	subclonal (intermixed) polysomy	
more mutant	Patient17	Initial	1	positive	intact	
		Recurrence1	4	positive; some areas with strong staining	intact; occasional cell with high level polysomy/gain	subclonal polysomy
	Patient21	Initial	0	N/A	N/A	N/A
		Recurrence1	1	positive; some areas with very strong staining	polysomy plus IDH1 amp	polysomy with subclonal focal amplification

Table 4.2. Summary of IHC and FISH results

Table 4.3

	IDH1 R132H present region		IDH1 R132H intense region	
	"het" = higher 2HG		amp mut = lower 2HG	
	ki-67	cells/200x field	ki-67	cells/200x field
Patient17_Recurrence1	5.98%	242.5	18.39%	368
Patient21_Recurrence1	27.64%	354.5	43.56%	375

	IDH1 R132H present region		IDH1 R132H absent region	
	"het" = higher 2HG		del mut = lower 2HG	
	ki-67	cells/200x field	ki-67	cells/200x field
Patient169_Recurrence1	31.90%	153.6	48.94%	307
			50.44%	343

Table 4.3. Results of Ki67 quantification in patients with regionally heterogeneous IDH1 R132H staining patterns

CHAPTER 5:
MATERIALS AND METHODS

5.1 Sample acquisition

All tumor samples were collected during surgical resection and were either snap frozen in liquid nitrogen and stored at -80°C , or were formalin fixed and paraffin embedded (FFPE). In cases where more than one sample from a tumor was investigated, those samples were independent, geographically distinct pieces derived from multiple time points during surgery. Patient-matched normal samples were peripheral blood or muscle tissue. Samples were obtained from the Neurosurgery Tissue Bank at the University of California San Francisco (UCSF). Sample use was approved by the Committee on Human Research at UCSF and research was approved by the institutional review board at UCSF. Additional samples were obtained from: the University of Tokyo hospital with the approval of the Ethics Committee of the University of Tokyo; Erasmus Medical Center with the approval of the Medical Ethics Committee at Erasmus Medical Center Rotterdam; and the OncoNeuroTheque tissue bank at Groupe Hospitalier Pitié-Salpêtrière with the approval of the Ethics Committee. All patients provided informed written consent.

Snap frozen normal human post-mortem brain tissue from two males (55 and 56 years of age respectively) was obtained from the National Disease Research Interchange (NDRI) and frontal cerebral cortex gray matter was macrodissected. Normal brain tissues were also acquired from post-mortem human fetal neural tissues (obtained from two cases of twin nonsyndrome fetuses whose deaths were attributed to environmental/placental etiology) and adult insula normal brain (obtained from an autopsy of a case whose death was not related to brain malignancy). Tissues were obtained with written consent according to Partner's Healthcare/Brigham and Women's Hospital IRB and UCSF IRB guidelines. Data from additional fetal and adult normal brain tissue were obtained from previously published datasets (*169*) and publicly available data (CopyNumber450k R package).

5.2 DNA and RNA isolation

Genomic DNA was extracted with either a QIAGEN DNA extraction kit (Qiagen, Valencia, CA) following the manufacturer's instructions or isolated by PCI extraction. For PCI extraction, tissues were digested with 1mg/ml proteinase K in lysis buffer (50mM Tris, pH 8.0, 1mM EDTA pH 8.0, 0.5% SDS) overnight at 55C. After RNase treatment, DNA was phenol/chloroform extracted, precipitated with ethanol and resuspended in TE. For DNA extraction from FFPE, PCI extractions were preceded by xylene washes to remove paraffin. RNA was isolated with Trizol (Invitrogen, Carlsbad, CA, USA) according to manufacturer's instruction.

5.3 Exome sequencing

5.3.i Hybrid capture, sequencing and alignment

Exome capture was performed using either Agilent (SureSelect Human All Exon 50MB, SureSelect Human All Exon v4, SureSelect Human All Exon v5) or NimbleGen (SeqCap EZ Exome v3) exome capture kits according to manufacturer's protocol. All sequencing reported here acquired paired-end reads of 76bp or 100bp in length from Illumina HiSeq 2000 and 2500 instrumentation. Paired-end sequencing data from exome capture libraries were aligned to the reference human genome (build hg19) with the Burrows-Wheeler Aligner (*170*). All sequenced and aligned libraries (uniquely aligned reads only) were further processed with both the Picard suite (<http://picard.sourceforge.net/>) and the Genome Analysis Toolkit (*171, 172*), including de-duplication, base quality recalibration, and multiple-sequence realignment.

5.3.ii Mutation identification

Single-nucleotide variants (SNVs) were detected with MuTect, a Bayesian framework for the detection of somatic mutations (173). SNVs with less than a 10% variant frequency in the tumor, with more than 5 variant reads in the patient-matched normal, or greater than a 10% variant frequency in the patient-matched normal were excluded from further analysis. Indels under 50bp were detected with Pindel (174), and those with fewer than 6 supporting reads in the tumor, any supporting reads or less than 14 total reads in the patient-matched normal, and replacements for which the deletion and non-template inserted sequence were of the same length were excluded. If multiple indels were present at the same genomic coordinates, only the indel with the most supporting reads was retained. All indels and SNVs were annotated for their mutational context and effect using ANNOVAR (175). Only protein-coding or splice-site mutations were retained for further analysis. Mutations were additionally annotated for their presence in dbSNP (Build ID: 132) (<http://www.ncbi.nlm.nih.gov/SNP/>) or 1000Genomes (Phase 1, November 2010 release) (176) data sets. To generate a list of only the highest quality variants for phylogenetic tree construction (Chapter 3 only), further filtering was applied by excluding all SNVs that were not classified as “covered” by MuTect in all samples for that patient, SNVs with any variant reads detected in the patient-matched normal and all indels, unless validated by Sanger sequencing.

In the original exomes from the initial and recurrent gliomas of patients 26 and 27, a total of 44 mutations were categorized as private. After the generation of additional exomes from distinct regions of each tumor, 3 mutations that were formerly private were now shared: 1 mutation in the recurrence of patient 26 and 2 mutations in the recurrence of patient 27 were called in the additional exomes of their patient-matched initial gliomas, for an overall misclassification rate of $3/44=7\%$.

TMZ-associated mutations were defined as C>T/G>A transition mutations at any CpN dinucleotide context which were unique to a TMZ-treated hypermutated recurrent tumor. To ensure this was a conservative determination, any TMZ-associated mutation identified in a recurrence that showed evidence of the alternate allele (1 or more reads of base quality greater than or equal to 20) in the patient-matched initial tumor (or first recurrent tumor for patient 24) or failed Sanger validation were not labeled candidate TMZ-associated mutations for the purposes of further analysis. Nevertheless, it is possible that a subset of candidate TMZ-associated mutations were misattributed as such, having existed in the initial tumor but were not identified from exome sequence data due to intratumoral heterogeneity. Therefore, we compared the TMZ-associated mutations identified in the recurrent tumor of patient 18 with exome sequencing data from three additional samples from geographically distinct regions of the initial tumor and from eight other non-patient matched initial tumor exomes with similar coverage (patients 04, 07, 12, 13, 15, 17, 24, and 26). The number of mutations that were misattributed as TMZ-associated was small and did not vary significantly between the patient-matched ($1.7\pm 0.08\%$) and non-patient-matched ($1.5\pm 0.6\%$) comparisons (p-value = 0.5; Wilcoxon rank sum test).

5.3.iii Mutation validation: Sanger sequencing

Candidate mutations were validated with PCR amplification of genomic DNA followed by Sanger biochemistry according to conventional protocols. While most primers were designed with Primer3 (177), primers for the IDH1 mutation were taken from Christensen et al (178). All primer sequences are available upon request. PCR was performed using either the KOD-plus (TOYOBO) or Phusion (New England Biolabs) high-fidelity DNA polymerases under optimized

thermal conditions. PCR products were evaluated on agarose gels and sequenced in both directions by Quintara Biosciences. Additional sequencing was performed using Big Dye Terminator reactions and subsequent loading on an ABI 3130xl capillary sequencer (Applied Biosystems).

To confirm mutations of interest as well as the large number of somatic mutations in the hypermutated recurrent tumors, we used Sanger sequencing and validated 50 of 51 randomly selected mutations and 213 of 291 targeted mutations, many of which were assayed for their likely false-positive status (102). To expand the number of loci assayed, we used available transcriptome sequencing data for the initial and recurrent tumors of eight patients as a second orthogonal means of confirming somatic mutations. Of those putative mutations with at least 10 transcriptome sequencing reads, we confirmed 3,385 of 3,812 coding somatic point mutations (102). Our overall validation rate of 88% compares favorably with many recent cancer sequencing studies (179, 180).

5.3.iv Mutation validation: PCR analysis of ATRX deletion

Genomic DNA from the initial and recurrent tumors for patient 17 was amplified using the following primers: forward ACGCATCTTCATTTACAGTTTCA and reverse AACAAAGTATGTAGAATCAGATGATGA. These primers flank the 8bp deletion found in the recurrence, allowing for amplification of both the wild-type and deletion alleles. The PCR products were resolved on a 15% TBE polyacrylamide gel, stained with SybrSafe (Invitrogen) for 30min and then imaged. To determine the limit of detectability for the deletion allele, a dilution series (30ng down to 30pg) was performed with genomic DNA from the recurrent tumor.

5.3.v Mutation validation: droplet digital PCR

TaqMan assays against mutations in MLH1 and TP53 were designed and synthesized by Applied Biosystems: (1) fwd primer: CCTGATTGGATTACCCCTTCTGATT, rev primer: CAGTGGCTAGTCGAAGAATGAAGAT, VIC (wt) probe: ACTATGTGCCCCCTTTG, FAM (mut) probe: AACTATGTGCTCCCTTTG for mutation MLH1 chr3:37090054 C>T identified in patient 01 recurrence (2) fwd primer: TCCCTTGTCTTTTTCTGCAA, rev primer: GATAGGCAGTCCCTCCAAAGG, VIC (wt) probe: CTGATTGGATTACCCCTTC, FAM (mut) probe: CTGATTGGATTATCCCTTC for mutation MLH1 chr3:37090029 C>T identified in patient 10 recurrence (3) fwd primer: GCCAGTTGGCAAACATCTTGT, rev primer: TGACTTTCAACTCTGTCTCCTTCCT, VIC (wt) probe: CCTACAGTACTCCCCTGCC, FAM (mut) probe: CCTACAGTACTTCCCTGCC for mutation TP53 chr17:7578550 G>A identified in patient 17 recurrence. A TaqMan assay against BRAF V600E was synthesized by Applied Biosystems from primer and probe sequences described in BioRad Bulletin 6260: fwd primer: CTA CTGTTTTCTTTACTTACTACACCTCAGA, rev primer: ATCCAGACA ACTGTTCAA ACTGATG, VIC (mut) probe: TTGGTCTAGCTACAGAGAAAT, FAM (wt) probe: TTGGTCTAGCTACAGTGAAAT

All reactions were set up with the droplet PCR supermix and the QX100 Droplet Digital PCR system from Bio-Rad (Hercules, CA) according to manufacturer guidelines and analyzed using the QuantaSoft software in rare event detection (RED) mode. Assays were first optimized through a temperature gradient to maximize separation between the mutant and wild-type signals. Assays were then subject to titration experiments to determine if mutant allele detection was robust down to 0.1%.

The assays were then tested in genomic DNA from the tumors of interest. In all cases, genomic DNA was first subject to restriction enzyme digest (MluCI (patients 01, 10, 18) and AluI (patient 17), New England BioLabs) at 37°C for 15 minutes. Each assay was run with digested DNA from the tumor in which the mutation was identified (positive control), normal DNA from a different patient (negative control, in duplicate), the patient-matched tumor in which the mutation was not detected (in quadruplicate) and a no-DNA water control.

5.3.vi Mutation rates

Mutation rates were determined for each tumor from the total number of SNVs detected and the total number of base pairs sequenced to a sufficient depth and quality for mutations to be called by MuTect. Only mutations and sequencing data in protein coding regions and splice-sites were used.

Though it is not possible to determine with certainty whether any single mutation was directly induced by TMZ exposure, one can estimate the proportion of mutations attributable to such an event. Given the mechanisms of action of TMZ and its near-exclusive induction of C>T/G>A transitions across all CpN dinucleotide contexts (11, 15, 17), we can estimate R and P , which are the C>T/G>A mutation rates for the TMZ-treated recurrent tumor (R) and the TMZ-naïve initial tumor (P). Here, R is a mixture of both the spontaneous mutation rate of the tumor (measured as P) and the rate of mutation due to TMZ exposure ($R-P$). Then, $(R-P)/R$ estimates the proportion of C>T/G>A mutations in the TMZ-treated recurrent tumor attributable solely to TMZ exposure. Therefore, we determined that 98.7% to 99.8% of the mutations we identified in the TMZ-treated hypermutated tumors were attributable to TMZ-associated mutagenesis. Though other factors not modeled here could influence the recurrent tumor mutation rate, the

relative stability of mutation rates between initial and recurrent tumors in the untreated patients strongly suggests this is not the case.

5.3.vii Copy number analysis (Chapter 2)

Copy number segmentation was performed on all samples from exome sequence data with an adaptation of CBS segmentation (181). For each tumor (initial or recurrence) and matched normal pair, total aligned coverage was determined for each baited exon from the hybrid selection assay. After excluding exons from the analysis with coverage levels in the matched normal sample of fewer than 10 reads, a log₂ ratio of read coverage levels from the tumor-to-normal was determined and scaled by the total aligned sequence in each library. These were segmented with CBS and change-points were reverted if the standard deviation between the means of adjacent segments was less than 1.5. Segmented profiles were then normalized and copy number alterations determined with the RAE framework (182).

5.3.viii Copy number and decrease of heterozygosity analysis (Chapter 4)

Total copy number (TCN) is calculated as the ratio of aligned tumor and (matched) normal reads in non-overlapping bins. For each bin we also calculated the average decrease of heterozygosity (DoH), which represents how much the tumor and normal differ in allelic ratios across the SNPs in the bin. Specifically, we used samtools mpileup (183) on the tumor-normal BAM files followed by a joint sequenza::pileup2seqz() (184), to obtain tumor and normal TCN counts and DoH estimates at each genomic position. These estimates were then averaged in non-overlapping 100-kb bins resulting in (TCN,DoH) sample-mean estimates for each bin. Based on these binned averages, we used PSCBS::segmentedByPairedPSCBS() (185) to partition the (TCN,DoH)

estimates into genomic regions of (statistically) constant PSCN levels. For each PSCN segment we obtained a genomic start and stop position, TCN and DoH mean levels, and noise estimates. Minor and major CN segment levels (C1,C2) follow from the (bijective) relationship $C1 = 1/2 * (1 - DoH) * TCN$ and $C2 = TCN - C1$.

5.3.ix Tumor cell fraction analysis

Purity (fraction of cells in the sequenced sample that were tumor) and ploidy in each tumor were estimated with ASCAT (186) using B allele frequencies inferred for all heterozygous SNPs genotyped as non-reference in tumor and normal exome pairs. At each of these variant sites and in each sample, a log₂ copy number ratio was assigned from the average read coverages of the host exon. ASCAT was run with default parameters and the allelic frequencies of all somatic SNVs were converted into the fraction of tumor cells bearing that mutation as previously described (187). Calculation of tumor cell fraction does not adjust for hemizyosity in the normal, so SNVs on the X chromosome in male patients were excluded. One SNV (TP53 R213P, Patient 05) was miscalled as absent in the primary due to low coverage at the locus, but was validated as somatic in both tumors by Sanger sequencing, and is therefore presented as a shared variant. Indel variant frequencies were not available and therefore were excluded from this analysis. Purity and ploidy estimates for the first recurrence of patients 06 and 24 could not be estimated by ASCAT likely due to low tumor purity. Subclonal CNAs, including those arising in rare cell populations, were distinguished from intrinsic experimental noise in the copy number inference from tumor/normal exome coverage levels by comparing all lesions to the minimum allele frequencies (MAF) of heterozygous SNPs spanned by the event. Subclonal CNAs have a decrease in their log₂ copy number ratio that is accompanied by a similar change in MAF.

5.4 DNA methylation

5.4.i Illumina 450K array and preprocessing

Genomic DNA was bisulfite converted using the EZ DNA Methylation Kit (Zymo Research) and processed on Infinium HumanMethylation450 bead arrays (Illumina Inc.) according to the manufacturer's protocol. Probe-level signals for individual CpG sites were subject to both background and global dye-bias correction (188). Probes that map to regions with known germline polymorphisms (Illumina supplementary SNP list v1.2, downloaded Sept. 3, 2013), to multiple genomic loci (189), or to either sex chromosome were filtered out. 297,342 probes remained following filtering.

5.4.ii Differential methylation analysis (Chapter 3)

The glioma CpG island methylator phenotype (G-CIMP) was confirmed in all tumors profiled here by examining methylation levels at CpGs adjacent to eight previously defined markers (*ANKRD43*, *HFE*, *MAL*, *DOCK5*, *LGALS3*, *FAS-1*, *FAS-2*, *RHOF*) (78). To determine common and specific methylation profiles in the paired initial and recurrent tumors, we performed two-way unsupervised hierarchical clustering using Euclidean distance and Ward linkage on the most variable CpG sites across the cohort, with variability ranked by standard deviation (0.5% cutoff = 1,486 CpGs; 50% cutoff = 148,572 CpGs).

For all subsequent statistical analyses, beta values for individual CpG sites were made more Gaussian using the logit-transformation. We subtracted the transformed beta values between patient-matched recurrent and initial tumors and used Limma (190), an empirical Bayes approach utilizing a moderated t-statistic, to test for significant differences in individual CpG

sites between the group of patients that recurred as GBM and the group that did not. Differentially methylated CpGs were defined as those with both a nominal p value < 0.05 and an average methylation change upon recurrence ≤ -0.2 or ≥ 0.2 . The same empirical Bayes approach was also used to compare methylation differences between the GBM and non-GBM groups. Hypomethylated CpGs were defined as those with both a nominal adjusted p value < 0.05 and an average methylation change upon recurrence as GBM ≤ -0.2 and a difference of the average change between the GBM and non-GBM group of -0.15 . Genes associated with the promoter of these GBM-specific hypomethylated CpGs that were also over-expressed upon recurrence as GBM were subject to functional enrichment with clusterProfiler (191) against a background of all genes that have methylation probes in their promoter. Here, we defined promoters as 1.5kb upstream of the transcriptional start site (TSS) and 1kb downstream of the TSS.

5.4.iii Age-related methylation analysis

We used Limma to test for the differential methylation between 33 fetal and 8 adult brain tissues. We selected probes having both a nominal adjusted p value (derived from the previous analysis) < 0.05 and an average methylation change upon aging ≥ 0.2 .

5.4.iv Methylation analysis (Chapter 4)

The 8-site definition of G-CIMP was performed using the closest probes to eight previously defined markers: cg26399201 (*ANKRD43*), cg19320816 (*HFE*), cg21245652 (*MAL*), cg16849041 (*DOCK5*), cg17403875 (*LGALS3*), cg16257983 (*FAS-1*), cg17120764 (*FAS-2*), cg09088508 (*RHOF*) (78). For the unsupervised hierarchical clustering definition of G-CIMP, we downloaded 450K data from TCGA LGG and GBM (IDAT files, downloaded 10/29/2015

from TCGA data portal). We also downloaded copy number data (CBioPortal, 11/4/2015) and somatic mutations (TCGA data portal, 11/4/2015). We identified 647 cases with 450K array, copy number and somatic mutations which we processed through our pre-processing pipeline together with our data. We excluded 15 TCGA samples with >0.5% of probes with p-value >0.05. We selected the most variable probes by standard deviation (0.5% = 1,448 probes) (78) and performed two-way unsupervised hierarchical clustering using Euclidean distance and Ward linkage. PCA was performed on the 647 TCGA samples and the model was then applied to our data.

The mean methylation level was calculated across all probes, and across subsets of probes as defined by Illumina-provided annotations of CpG islands, shores and shelves. Methylation differences were calculated as the difference from initial to recurrence (or recurrence1 to recurrence2 for Patient14) at each probe. These values were averaged (mean), clustered on the most variable probes (top 0.5% by standard deviation, Euclidean distance, Ward linkage) and all pair-wise correlations were calculated (Spearman's rho).

5.4.v Whole genome shotgun bisulfite sequencing

One to 5µg of genomic DNA was sonicated to an approximate size range of 200–400 bp. DNA was quantified by fluorescent incorporation (Qubit, Invitrogen). Sonicated DNA was subjected to end-repair and phosphorylation with NEBNext™ or Illumina Sample Prep Kit reagents and addition of an 'A' base to the 3' end. Methylated adapters were ligated and size selection was performed to remove excess free adaptors. The ligated DNA was quantified by Qubit, and 100ng DNA was used for bisulfite conversion. Unmethylated lambda-phage DNA (NEB) ligated with methylated adaptor was used as an internal control for assessing the rate of bisulfite conversion.

The ratio of target library to lambda was 1600:1. The methylated adapter-ligated DNA fragments were subject to bisulfite conversion with Qiagen's Epitect Bisulfite Kit (FFPE Tissue Samples Protocol). Cleanup of the bisulfite-converted DNA was performed, followed by a second round of bisulfite conversion. Enrichment of adaptor-ligated DNA fragments was accomplished by dividing the template into five aliquots followed by eight cycles of PCR with adaptor primers. Post-PCR size-selection of the PCR products from the five reactions was achieved by PAGE gel. Libraries were subject to 100bp paired-end sequencing on Illumina instrumentation.

Individual sequencing lanes were chastity filtered, deduplicated, trimmed of low quality bases and adapter sequence (TrimGalore, http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and aligned to hg19 with Bismark v0.10.1 (192). All lanes sequenced from a single library were merged and deduplicated. Methylation information was extracted using Bismark.

The posterior distribution of methylation level at each CpG location was obtained using a binomial likelihood and an uninformative beta prior, where the likelihood gives the probability of finding the observed number of methylated cytosine among the total number of reads covering the base. To compute differential methylation between matched samples, the resulting posterior distribution of methylation at each CpG site from the initial tumor was used to compute the beta-binomial posterior predictive distribution and the p value for observing a given number of methylated cytosine at the corresponding site in the recurrence.

5.4.vi MGMT bisulfite sequencing analysis

In total, 1ug of DNA was bisulfite converted as previously described (193). Converted DNA was amplified with PCR using the following primers: forward GGATATGTTGGGATAGTT and

reverse TAAAAATCAAAACRACCCCACACC. Amplified DNA was gel extracted, cloned using the TOPO TA sequencing kit (Invitrogen) and 10-15 clones were sequenced. Bisulfite sequence data at the *MGMT* locus (194) was analyzed with BISMA (195).

5.4.vii MGMT methylation-specific PCR

Genomic DNA samples (250-1000ng each) were used for bisulfite reactions using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's protocol. DNA methylation status of the O6-methylguanine methyltransferase (*MGMT*) promoter was then determined by methylation-specific PCR as previously described (194).

5.5 Transcriptome sequencing

5.5.i Library construction and sequencing

PolyA⁺ RNA was purified using the MACS mRNA isolation kit (Miltenyi Biotec, Bergisch Gladbach, Germany), from 2-4ug of total RNA as per the manufacturer's instructions. The process included on-column DNaseI treatment (Invitrogen). cDNA was synthesized from the purified polyA⁺ RNA using the Superscript II Double-Stranded cDNA Synthesis kit (Invitrogen) and 200ng random hexamers (Invitrogen). After first strand synthesis, dNTPs were removed using 2 volumes of AMPure XP beads. GeneAmp 12.5mM dNTPs blend (2.5mM dCTP, 2.5mM dGTP, 2.5mM dATP, 5.0mM dUTP) was used in the second strand synthesis mixture. Double stranded cDNA was purified using 2 volumes of Ampure XP beads, fragmented using Covaris E series shearing (Covaris Inc. Woburn, MA, USA; 20% duty cycle, Intensity 5, 55 seconds), and used for paired-end sequencing library preparation following indexed plate-based library construction (196) with the following modifications: ligation reaction was extended to 2h, just

before library amplification uridine digestion was performed at 37°C for 30min following with 10min at 95°C in Qiagen Elution buffer (Qiagen Inc., Toronto, ON, Canada; 10mM Tris-Cl, pH 8.5) with 5 units of Uracil-N-Glycosylase (AmpErase UNG, Invitrogen).

5.5.ii Alignment and analysis (Chapter 2)

All transcriptome sequencing data from initial and recurrent tumor pairs were aligned with TopHat (v1.4.0) (197), with parameters `-r30 -library-type fr-firststrand`. To estimate transcript abundance, aligned data was processed with Cufflinks (v1.3.0) (198), with parameters `-b hg19.fa -G refGene_HG19.gtf -library-type fr-firststrand`. Alternative splicing was analyzed with ALEXA-seq (199). Gene set enrichment analysis (200) was performed using the MSigDB genetic and chemical perturbation gene sets (v3.0) of size 15-500, 1000 gene set permutations, the weighted enrichment statistic, and the log₂ ratio of classes as the gene ranking metric.

5.5.iii Alignment and analysis (Chapter 3)

All transcriptome sequencing data from initial and recurrent tumor pairs were aligned with TopHat (v2.0.12) (197) to the hg19 reference genome using a GENCODE transcriptome-guided alignment; the following parameters were used: `--transcriptome-index=hg19_GencodeCompV19 -library-type fr-firststrand`. The aligned data were then processed through custom quality-control scripts to remove unmapped, improperly-matched, multi-mapping, and chimeric reads, as well as accumulation in non-assembled chromosomes. To estimate transcript abundance, aligned data were processed with the `cuffnorm` and `cuffquant` commands from the Cufflinks package (v2.2.2) (198) against a Gencode reference transcriptome (downloaded from UCSC genome browser on 02/03/2014) that has its IDs already linked with official gene symbols. The `cuffquant` program

was run with parameters `--max-bundle-frags 50000000 -b hg19.fa --library-type fr-firststrand`; the cuffnorm program had the following parameters `--compatible-hits-norm --library-type fr-firststrand`.

For all subsequent statistical analyses, FPKM estimates for individual genes were made more Gaussian using a log₂-transformation. We subtracted the transformed FPKM estimates between patient-matched recurrent and initial tumors and used Limma to test for significant differences among individual genes within the group of patients that recurred as GBM and the group that did not. Differentially expressed genes were defined as those with both a nominal p value < 0.05 and an average log 2-fold change upon recurrence ≤ -1 or ≥ 1 . Limma was again used to compare methylation differences between the GBM and non-GBM groups. Upregulated genes were defined as those with both a nominal p value < 0.05 and an average log 2-fold change upon recurrence as GBM ≥ 1 and a difference of the average change between the GBM and non-GBM group of at least 1.

5.6 P value adjustment (Chapter 3)

Statistical tests for assessing significant differences in gene expression and methylation status were performed independently. The varying number of tests performed (~300k for methylation and ~25k for expression), makes it difficult to directly compare the resulting p values. While Storey's false discovery rate controlling for multiple-testing corrections are standard (201), our data show bimodal distribution of the RNA-seq analysis p values and do not satisfy the assumptions required to apply the method, resulting in incorrect estimation of the number of genes in the null distribution. Thus, we chose our p value cutoffs by identifying the value at which we would identify an equal number of false positives if all the test cases satisfied a null

hypothesis and the p values had a uniform distribution. Specifically, by using a .05 cut-off in the expression data, under our simplistic assumptions, to identify the same number of false positives in the methylation data, we would need to use a cutoff $p_{adjusted-methylation} = p_{methylation} * (N_{450k\ probes} / N_{genes})$. Our use of p values here is primarily to rank all probes and genes in our study and follow-up by selecting only those with the most consistent difference.

5.7 Phylogenetic and phyloepigenetic trees

5.7.i Phylogenetic trees (Chapter 2)

We constructed phylogenetic trees using mutations from exome sequencing data (202) of all samples of tumor tissue from patients 04 and 17, inferring ancestral relationships by clonal ordering (92, 93). A subset of the coding SNVs and indels from each branch was subject to validation by Sanger sequencing and their locations in the tree were adjusted (false negatives) or removed (false positives). One mutation (patient 04, *MUC4* p.3894_3910del) could not be assessed for technical reasons and was excluded. At least one mutation from each branch was confirmed by Sanger sequencing to validate the structure of the tree. The length of each branch in the phylogenetic tree is proportional to the number of mutations. Labeled genes are those for which there are more than 20 entries in the COSMIC database (version 58, March 2012 release) (118).

To confirm that the initial and recurrent tumors of patient 17 derived from a common cell of origin rather than being independently arising tumors, we identified non-coding mutations shared by all geographically distinct samples of the initial and recurrent tumor and validated 3 of them by Sanger sequencing.

5.7.ii Phylogenetic and phyloepigenetic trees (Chapters 3 and 4)

For the phylogeny analysis of both the genetic and epigenetic data, we employed an independent, but parallel, analysis of the methylation data and exome-seq mutations. For the exome-seq data, we used binary mutation calls to build a distance matrix for all samples from a patient using the Manhattan distance metric, including a normal tissue sample for which all mutations were absent. Similarly, for the methylation data, we used only the probes that had a beta value difference of at least 0.4 between any of the samples from a patient to build a Euclidean distance matrix. Using several other probe selection cut-offs produced similar results. A normal brain sample (adult insula tissue from a different individual) was not included in the probe selection, but was added to the distance matrix calculation to serve as the tree root. To compare the distance matrices from the mutation data and the methylation data, we calculated the Spearman's rho correlation. We then built the phylogeny trees using an ordinary least squares (OLS) minimum evolution (203) approach from the ape R package (204) using the distance matrices from the genetic and epigenetic data independently.

5.7.iii Identifying discriminative methylation probes by Singular Value Decomposition

To identify the probes most responsible for a particular bifurcation on a phyloepigenetic tree (similar to identifying mutations that differ between two branches of a phylogenetic tree), we used Singular Value Decomposition (SVD) of the methylation data matrix to calculate the left and right singular vectors that form orthonormal bases of the subspaces spanned by the columns and rows of the data matrix, respectively. Projecting the columns, corresponding to samples, onto the two-dimensional subspace spanned by the first two left singular vectors (SV) reduces the data to the first two principal components that maximally separate the samples in the probe

space. These projections are shown as arrows in Figure 3.5B, where the rows of the first two left SVs are plotted as scatter points representing probes. In this biplot, the probes that best separate samples have large absolute values in the SV1 direction.

The singular value decomposition (SVD) starts with a mean-centered $p \times n$ data matrix X , where the rows are probes and the columns are samples from a patient. A rank- k approximation of X is obtained from the SVD of X as $X_k = UDV^T$, where U contains the first k left singular vectors as columns, V contains the first k right singular vectors as columns, and D is a diagonal matrix of the first k singular values. We can rewrite X_k as $X_k = (UD^a)(D^{1-a}V^T) = GH$, where a determines the scaling of the probes and samples. A biplot uses $k=2$ and plots the rows of G as points and the columns of H as arrows. For the purpose of performing PCA on samples in the probe space, we used the parameter $a = 0$. The axes at the bottom and left of the biplot are the coordinate axes for the probes while the axes at the top and right of the biplot are the coordinate axes for the samples, allowing us to simultaneously represent both the separation of the samples and the magnitude of each probe contributing towards that separation.

5.8 Analysis of gene-level convergence

5.8.i Analysis of gene-level convergence within each patient

We compiled lists of genes with mutations and methylation changes that were present in at least one but not all pieces of tumor from a given patient (excluding hypermutated cases). We enriched those lists for functional events by counting only non-silent mutations and methylation changes in promoter regions. We then used Limma to determine if the methylation levels among the samples with a particular mutation were similar to each other and different from the methylation levels of those samples without the mutation – thus determining if the samples

without the mutation had different methylation levels than the samples with the mutation. Table 3.3 presents the 3 genes (4 CpG sites) for which the absolute value of the t-statistic was greater than 3.5.

5.8.ii Analysis of gene-level convergence among the patient cohort

On a patient level, we identified all genes with non-silent mutations and all genes with promoter methylation changes, excluding genes affected by both mutations and DNA methylation within the same patient, and also excluding hypermutated cases. We then compared those gene lists across patients to count the number of genes that are mutated in one patient but affected by DNA methylation changes in another patient (Table 3.3).

5.9 Gene ontology and enrichment analyses

For all gene ontology analysis, we used the clusterProfiler R package (191). In the analysis of methylation changes in annotated promoters, the genes that have methylation probes in their promoters were used as a background. For the analysis of methylation changes in the ChIP-seq-defined enhancer and promoter regions, the genes that have been identified within any regulatory element (everything outside heterochromatin) regions were used as a background. For the enhancer enrichment analysis, we counted probes with genomic coordinates within a region defined as an active enhancer in any of the normal brain or primary GBM samples, and then permuted the probe IDs 10,000 times.

5.10 Immunohistochemistry

Immunohistochemistry was performed at the Brain Tumor Research Center, University of

California San Francisco with a Ventana automated immunohistochemical staining processor using the following antibodies: IDH1 (Dianova #DIAH09), Phospho-4E-BP1 (Cell Signaling #2855), Phospho-S6 (Cell Signaling #2215) and MIB1 (CONFIRM anti-Ki-67 (30-9) Rabbit Monoclonal Primary Antibody, Ventana). Areas of maximal nuclear staining for MIB1 were selected for quantification of labeling index (LI), defined as the number of MIB1 positive cells divided by the total number of cells. At least 1,000 cells were evaluated (unless otherwise specified), and quantification was performed either by manual counting under a light microscope containing an eyepiece micrometer grid or through a semi-automated image analysis approach. Image acquisition was performed using an Olympus BX-41 microscope, 20X objective, and Olympus DP21 digital camera, and analysis was done using Image J (<http://imagej.nih.gov/ij/>) and the ImmunoRatio plug-in (<http://153.1.200.58/sites/default/files/software/immunoratio-plugin/index.html>).

5.11 Fluorescence in situ hybridization

Fluorescence in situ hybridization (FISH) was performed with commercially available probes according to manufacturer's instructions. FISH for IDH1 was performed with IDH1 (2q34, 5-TAMRA) and CEP2 (5-Fluorescein) probes from Empire Genomics (Buffalo, NY). FISH for polysomy was performed with centromeric FISH probes D2Z1 (SpectrumGreen), D3Z1 (SpectrumOrange) and D15Z1 (SpectrumAqua) (Vysis, Abbott Molecular, Abbott Park, IL).

5.12 Ex vivo HR-MAS spectroscopy

Image-guided tissue samples were loaded into a 35-ml zirconia rotor (custom-designed by Varian) with 3 ml of 99.9% atom-D deuterium oxide containing 0.75 wt % 3-

(trimethylsilyl)propionic acid (Sigma-Aldrich) for chemical shift referencing. Data were acquired at 11.7 T, 1°C, 2250 Hz spin rate in a 4-mm gHX nanoprobe with a Varian INOVA 500 MHz multinuclear spectrometer. The nanoprobe gHX is an inverse probe, optimized for the direct detection of protons and the indirect detection of X-nuclei (^{13}C , ^{31}P , ^{15}N) and was equipped with a magic angle gradient coil. A rotor-synchronized T2-weighted Carr-Purcell-Meiboom-Gill (CPMG) (205) pulse sequence was chosen for its ability to eliminate broad macromolecular signals, and run with a TR/TE = (4 s)/(144 ms), 768 scans, 40,000 acquired points, 90° pulse, and 20 kHz spectral width for a total time of 54 min.

Preprocessing of HR-MAS spectra was done in the frequency domain using the Chenomx processing software, which fits a basis set of known metabolites to a given spectrum (206). Preprocessing techniques included frequency referencing, .5 Hz apodization, baseline correction, and removal of the water signal for visualization of the full metabolic profile. Manual metabolite fitting was performed and standard area under the curve was calculated for quantitative analysis. Two experienced spectroscopists evaluated each spectrum to qualitatively assess goodness of metabolite fits, the presence of 2HG levels, and whether low resolution or SNR compromised its analysis.

5.13 Chromatin immunoprecipitation sequencing

Histone ChIP-seq and quality control were performed on four primary GBM frozen tissue samples as previously described (130). Briefly, histones marked with H3K4me3 (Cell signaling #9751), H3K4me1 (Diagenode #pAb-037-050), and H3K27Ac (Active Motif #39133), were immunoprecipitated using Sepharose beads coated in protein A/G, and then DNA purified. Pre-aligned ChIP-Seq data of the same histone modifications was downloaded for adult Inferior

Temporal Lobe, Hippocampus Middle, Mid Frontal Lobe, Cingulate Gyrus, and Anterior Caudate from the Human Epigenome Atlas (<http://www.genboree.org/epigenomeatlas/multiGridViewerPublic.rhtml>).

Genome-wide active promoter and enhancer states were generated from the aligned primary GBM and adult normal brain ChIP-seq data using ChromHMM v1.03 (207). The default parameters were used to binarize the bed files (chromHMM.jar binarizeBed), and the following parameters were used to learn the HMM Model: -xmx3g chromHMM.jar LearnModel 5 hg19. The hidden state showing co-occurrence of high H3K4me3 and H3K27ac marks was assigned as an 'Active Promoter.' Similarly, the state showing co-occurrence of high H3K4me1 and H3K27ac but no H3K4me3 was assigned as an 'Active Enhancer.'

5.14 Expression of exogenous mTOR and western blots

The pcDNA-FLAG-mTOR plasmid was obtained (a generous gift of Davide Ruggero Lab, UCSF) and mutated by site directed mutagenesis per the Quickchange II XL kit using primer sets (S2215F - CTGGCCAATGACCCAACATTTCTTCGGAAAAACCTC and GAGGTTTTTCCGAAGAAATGTTGGGTCATTGGCCAG; S2215Y - CTGGCCAATGACCCAACATATCTTCGGAAAAACCTC and GAGGTTTTTCCGAAGATATGTTGGGTCATTGGCCAG). HEK293 cells were transfected with pcDNA-FLAG-mTOR, S2215F, S2215Y, or pcDNA empty vector using Lipofectamine 2000 reagent overnight. Cells were then serum and nutrient starved by incubation in DMEM/0.1% FBS for 47hrs and then in sterile PBS for 1hr prior to harvest. Cells were lysed in Cell Lysis Buffer (Cell Signaling). Western blots were probed with the following antibodies

FLAG-M2 (Cell Signaling #8146), mTOR (#2983), RPS6 (#2217s), p-4E-BP1 (#2855s), 4E-BP1 (#9644s), p-RPS6 (#2211s), and GAPDH (Millipore mab374).

5.15 Radiologic analysis

Magnetic resonance imaging was acquired from the electronic radiology archives at UCSF. T1- and T2-weighted images representative of the tumor region were obtained from pre-surgical and post-surgical scans performed at either UCSF or outside institutions. T1-weighted images were acquired from a spoiled gradient echo (SPGR) pulse sequence after an injection of a paramagnetic gadolinium contrast agent. T2-weighted imaging was acquired from a fluid attenuated inversion recovery (FLAIR) sequence when available, and a standard fast spin echo (FSE) sequence when unavailable.

CHAPTER 6:
DISCUSSION

6.1 Hypermutation

We have shown that LGG can hypermutate as a consequence of TMZ and that hypermutated recurrences from initially LGG universally undergo malignant progression to highly malignant GBM. Moreover, we identified TMZ-associated mutations in genes associated with malignant progression, suggesting that TMZ may be directly driving these tumors down a malignant trajectory. This finding is highly concerning given the widespread use of TMZ in LGG patients, but there is still much work to be done to fully understand the ramifications of hypermutation.

There are three major questions that arise from the work done to date. First, what is the frequency of hypermutation in tumors that are treated with TMZ (or other alkylating chemotherapies)? Answering this question will require a much larger cohort of tumors than the 10 TMZ treated cases we published in 2014. To this end, we have collected additional cases in the intervening years and now have initial LGG and recurrence sample pairs from 50 patients treated with TMZ. However, beyond collecting cases to test for hypermutation, there are some inherent biases that will complicate the discovery of the true frequency of hypermutation. In order to classify a tumor as hypermutated or not, tissue must be acquired. This requirement for tissue will bias the patient population in two major ways: (1) tissue acquisition requires that surgery be the suggested clinical course, eliminating, for example, cases in which tumors are in less accessible regions of the brain, which may be a subset of tumors with unique features; (2) surgery must be performed at UCSF (or another major medical center) where tissue is more easily accessible for research, biasing towards patients with the means to travel to these centers.

In addition to biases in sample acquisition, the question of frequency is also complicated by the lack of uniformity in treatment regimens. TMZ may be given following a biopsy or subtotal resection or gross total resection, and may be given immediately following surgery or

after radiographic progression. Given that each of these factors affects the number of tumor cells in the brain at the time of treatment, and an increased number of cells could provide additional opportunities for one cell to acquire the MMR mutation necessary for hypermutation, it is possible that tumor burden may alter the frequency of hypermutation. The duration of treatment and intensity of doses could also impact the frequency. All of these factors will make it difficult to determine the frequency of hypermutation in an unbiased manner.

The second major question that arises is: what is the impact of hypermutation on progression-free survival (PFS) or overall survival (OS)? Similar to the previous questions, beginning to address this question requires a much larger cohort of post-treatment cases than is currently available. Moreover, this question is also complicated by the lack of uniformity in treatment regimens. Ideally, a large cohort of uniformly treated LGG cases from clinical trials of TMZ treatment should be used to minimize these variables.

Finally, the third major question is whether we can predict which tumors will hypermutate after TMZ treatment. Answering this question requires tissue samples from both before and after treatment with TMZ. While paired tissue samples like this are difficult to acquire, we have paired pre-treatment samples for almost all of our 50 post-treatment tumors. We have a wide range of data (germline SNPs, somatic mutations, copy number variation, RNA expression, DNA methylation) for each of these cases and work is ongoing to determine if we can predict the hypermutation status of a recurrence from the tissue acquired prior to treatment.

However, even before we have answers to the above questions, there is potential to translate this research finding into improved clinical care. Based on our finding that hypermutated recurrences universally activate the AKT-mTOR pathway, UCSF began a clinical trial (NCT02023905) in which patients are treated with TMZ for its anti-tumor activity along

with the mTOR inhibitor RAD001 to delay or prevent the outgrowth of malignantly transformed cells. As we accumulate a larger cohort we will be able to identify other commonly altered pathways, which may open up new therapeutic opportunities.

Additionally, there is growing evidence that immune checkpoint inhibitors are most efficacious against tumors with high mutational burden (208-211). This new family of therapies reactivates the immune system by turning off negative feedback signals from the tumor. However, even if the immune system is active in the microenvironment of the tumor, the tumor cells still need to present a tumor-specific antigen to trigger immune system cytotoxicity. Tumor-specific antigens, or neoantigens, are peptide fragments with abnormal sequence as a result of somatic mutations in the tumor cells. Since not all somatic mutations produce a neoantigen (212), it is theorized that the higher the mutational burden of a tumor, the more likely that tumor cells will present a neoantigen (213). There are currently several clinical trials that will test the efficacy of checkpoint inhibitors in GBM patients. Promisingly, two patients with germline MMR-deficient and somatically hypermutated pediatric GBM were successfully treated with a checkpoint inhibitor (214). The durable clinical response seen in both patients lends hope that these checkpoint inhibitors can be effective in the brain.

6.2 Co-evolution of genomics and epigenomics

We demonstrated that evolutionary histories built from intratumoral heterogeneity data look highly similar when derived from genomic (somatic mutations) or epigenomic (DNA methylation) data. However, we also demonstrated, as in the case of Patient04, that the differences between the evolutionary histories from these two data types can provide a novel understanding. Thus, to best understand tumor heterogeneity between tumor subclones and to

build a comprehensive evolutionary history of cancer progression, a novel analytical approach combining genetic and epigenetic data would be beneficial.

The similarity between genomic and epigenomic derived evolutionary histories suggests that genetic and epigenetic changes may co-evolve during tumorigenesis, raising the question of a mechanistic basis of this co-evolution. Aberrant epigenetic states may promote genetic instability or may arise from specific genetic alterations (215, 216). As an example, using single samples per patient from a large cohort of patients, comparison of samples with and without particular somatic mutations have identified associations between mutations and DNA methylation patterns. These associations reflect both mutations that drive altered DNA methylation, as with *IDHI* mutation in LGG (78, 82) and altered DNA methylation landscapes which allow for the acquisition of particular mutations, as with *BRAF* mutation in colorectal cancer (65, 66).

Despite the success of previous studies in identifying the relationship between a mutation and specific methylation changes, these statistical analyses required large cohorts of cases with and without each mutation to overcome the inherent inter-individual variability of DNA methylation arising from germline epigenetic differences, age, gender and other covariates. One approach to identify genetic-epigenetic associations in a smaller cohort is to use intratumoral heterogeneity of mutations and DNA methylation. Chromatin modifier genes, including *SMARCA4*, are often mutated as late events in tumorigenesis and therefore are present heterogeneously within a tumor (124, 126, 154). By profiling mITH and contrasting the samples with and without mutation in a particular gene, and then extending the analysis across a cohort of patients with similarly heterogeneous mutations in the same gene, substantial inter-patient heterogeneity can be excluded to then identify DNA methylation changes that result from the

specific mutation. Alternatively, cohorts of hundreds or thousands of cases might be sufficient to address the associations.

6.3 Clinical Implications

One of the difficulties of treating LGG is the lack of treatment options beyond radiation and chemotherapy with demonstrated clinical benefit. This problem is compounded by the time frame required for a clinical trial to determine the impact of a treatment on OS. Given that patients can live for more than a decade following diagnosis, clinical trials require long-term follow-up to detect an impact on OS. As an example, a major phase 3 trial testing the benefit of PCV in LGG patients (NCT00003375, RTOG-9802) enrolled patients from 1998 until 2002. In 2012, ten years after enrollment was completed, there was a significant difference in PFS but not OS (217). In a follow-up in 2016, nearly 20 years after the study began, an OS difference was apparent (4). This means that from the time a new therapeutic option is available in a clinical trial, it may take 20 years to determine if that treatment is beneficial in this patient population.

While we have demonstrated that TMZ can have a detrimental impact on tumor evolution in a histologically and genetically defined subset of LGG patients and there is no known OS benefit, anecdotal evidence suggests that TMZ improves quality of life in LGG patients (3). There are currently several phase 3 clinical trials testing TMZ in different subsets of LGG patients. While the analysis of RTOG-9802 was hampered by a lack of tissue for analysis, limiting the analysis of molecular subtypes (4), several of these newer TMZ studies include tissue requirements which will allow analysis to be performed on the basis of molecular subtypes in addition to histological subtypes. Unfortunately, while several of these studies stratify or limit patient enrollment on the basis of 1p19q status, none of them account for *IDH* status. This is

likely a consequence of timing, as several of these studies started in the same timeframe as the discovery of *IDH* mutations as a common event in glioma (159, 218). The tissue requirements will hopefully allow for post-analysis on the basis of *IDH* status.

Another major new therapeutic push is based around *IDH* mutations. Several companies have small molecule inhibitors of the mutant enzymes in clinical trials. There are also vaccine-based approaches, which inject a short peptide including the mutation-specific amino acid change to raise an immune response against cells presenting the mutant peptide (160). Currently two open phase 1 clinical trials are testing this vaccine-based approach (219). Our study of patients with copy number changes at the *IDH1* locus suggests that the mutant IDH1 protein is not required for tumor maintenance, and therefore that loss of the mutant protein is a potential mechanism of resistance to these targeted therapies. Loss of a neoantigen, a process called immunoediting, has been seen in GBM following treatment with an antibody against a mutant variant of EGFR (220). It will be interesting to see the results of these clinical trials, in light of our findings about the importance of *IDH1* mutations at later stages of tumor growth.

Finally, there is the very exciting field of immune checkpoint inhibitors. As discussed above, based on recent findings that checkpoint inhibitors are most active in tumors with a high mutational burden (208-211), there is hope that they will be similarly effective in recurrences that hypermutated following TMZ treatment.

While it will likely take several decades to have definitive evidence that any of these treatments improve OS for LGG patients, there is reason to be excited. The discoveries that have already come out of genomic analyses have opened up several new therapeutic opportunities. As the field continues to understand the events that drive tumor initiation and growth, and as the

evolutionary history of tumors is better understood, there will continue to be new opportunities for therapeutic intervention.

BIBLIOGRAPHY

1. Q. T. Ostrom *et al.*, CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007-2011. *Neuro Oncol* **16 Suppl 4**, iv1-63 (2014).
2. D. N. Louis *et al.*, The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* **114**, 97-109 (2007).
3. N. Sanai, S. Chang, M. S. Berger, Low-grade gliomas in adults. *J Neurosurg* **115**, 948-965 (2011).
4. J. C. Buckner *et al.*, Radiation plus Procarbazine, CCNU, and Vincristine in Low-Grade Glioma. *N Engl J Med* **374**, 1344-1355 (2016).
5. R. Stupp *et al.*, Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* **352**, 987-996 (2005).
6. Q. T. Ostrom *et al.*, The epidemiology of glioma in adults: a "state of the science" review. *Neuro Oncol* **16**, 896-913 (2014).
7. J. E. Eckel-Passow *et al.*, Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. *N Engl J Med*, 150610140038004 (2015).
8. Cancer Genome Atlas Research Network *et al.*, Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med* **372**, 2481-2498 (2015).
9. G. Cairncross *et al.*, Phase III trial of chemoradiotherapy for anaplastic oligodendroglioma: long-term results of RTOG 9402. *J Clin Oncol* **31**, 337-343 (2013).
10. M. J. van den Bent *et al.*, Adjuvant procarbazine, lomustine, and vincristine chemotherapy in newly diagnosed anaplastic oligodendroglioma: long-term follow-up of EORTC brain tumor group study 26951. *J Clin Oncol* **31**, 344-350 (2013).
11. W. J. Bodell, N. W. Gaikwad, D. Miller, M. S. Berger, Formation of DNA adducts and induction of lacI mutations in Big Blue Rat-2 cells treated with temozolomide: implications for the treatment of low-grade adult and pediatric brain tumors. *Cancer Epidemiol Biomarkers Prev* **12**, 545-551 (2003).
12. L. Liu, S. L. Gerson, Targeted modulation of MGMT: clinical implications. *Clin. Cancer Res.* **12**, 328-331 (2006).
13. M. E. Hegi *et al.*, MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med* **352**, 997-1003 (2005).
14. S. Yip *et al.*, MSH6 mutations arise in glioblastomas during temozolomide therapy and mediate temozolomide resistance. *Clin Cancer Res* **15**, 4622-4629 (2009).
15. C. Hunter *et al.*, A hypermutation phenotype and somatic MSH6 mutations in recurrent human malignant gliomas after alkylator chemotherapy. *Cancer Res* **66**, 3987-3991 (2006).
16. D. P. Cahill *et al.*, Loss of the mismatch repair protein MSH6 in human glioblastomas is associated with tumor progression during temozolomide treatment. *Clin Cancer Res* **13**, 2038-2045 (2007).
17. Cancer Genome Atlas Research Network, Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061-1068 (2008).
18. H. F. van Thuijl *et al.*, Evolution of DNA repair defects during malignant progression of low-grade gliomas after temozolomide treatment. *Acta Neuropathol* **129**, 597-607 (2015).
19. V. W. Zhou, A. Goren, B. E. Bernstein, Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* **12**, 7-18 (2011).
20. T. B. Miranda, P. A. Jones, DNA methylation: the nuts and bolts of repression. *J Cell Physiol* **213**, 384-390 (2007).

21. R. J. Klose, A. P. Bird, Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* **31**, 89-97 (2006).
22. M. P. Ball *et al.*, Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* **27**, 361-368 (2009).
23. R. Lister *et al.*, Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322 (2009).
24. A. K. Maunakea, I. Chepelev, K. Cui, K. Zhao, Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res* **23**, 1256-1269 (2013).
25. S. Shukla *et al.*, CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74-79 (2011).
26. B. Jin, K. D. Robertson, DNA methyltransferases, DNA damage repair, and cancer. *Adv Exp Med Biol* **754**, 3-29 (2013).
27. Y. F. He *et al.*, Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307 (2011).
28. S. Ito *et al.*, Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-1303 (2011).
29. M. Tahiliani *et al.*, Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935 (2009).
30. D. Hanahan, R. A. Weinberg, The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
31. Nature Milestones in Cancer. *Nat Rev Cancer* **6**, S7-S23 (2006).
32. A. Balmain, Cancer genetics: from Boveri and Mendel to microarrays. *Nat Rev Cancer* **1**, 77-82 (2001).
33. A. G. Knudson, Jr., Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**, 820-823 (1971).
34. C. J. Tabin *et al.*, Mechanism of activation of a human oncogene. *Nature* **300**, 143-149 (1982).
35. E. P. Reddy, R. K. Reynolds, E. Santos, M. Barbacid, A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. *Nature* **300**, 149-152 (1982).
36. E. Taparowsky *et al.*, Activation of the T24 bladder carcinoma transforming gene is linked to a single amino acid change. *Nature* **300**, 762-765 (1982).
37. C. R. Lin *et al.*, Expression cloning of human EGF receptor complementary DNA: gene amplification and three related messenger RNA products in A431 cells. *Science* **224**, 843-848 (1984).
38. G. T. Merlino *et al.*, Amplification and enhanced expression of the epidermal growth factor receptor gene in A431 human carcinoma cells. *Science* **224**, 417-419 (1984).
39. Y. H. Xu *et al.*, Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs overproduced in A431 carcinoma cells. *Nature* **309**, 806-810 (1984).
40. T. A. Libermann *et al.*, Amplification, enhanced expression and possible rearrangement of EGF receptor gene in primary human brain tumours of glial origin. *Nature* **313**, 144-147 (1985).
41. W. K. Cavenee *et al.*, Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* **305**, 779-784 (1983).
42. S. H. Friend *et al.*, A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323**, 643-646 (1986).

43. S. J. Baker *et al.*, Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* **244**, 217-221 (1989).
44. C. A. Finlay, P. W. Hinds, A. J. Levine, The p53 proto-oncogene can act as a suppressor of transformation. *Cell* **57**, 1083-1093 (1989).
45. A. Kamb *et al.*, A cell cycle regulator potentially involved in genesis of many tumor types. *Science* **264**, 436-440 (1994).
46. P. C. Nowell, *A minute chromosome in human granulocytic leukemia*. (Science, 1960).
47. J. D. Rowley, A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290-293 (1973).
48. A. P. Feinberg, B. Vogelstein, Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301**, 89-92 (1983).
49. M. A. Gama-Sosa *et al.*, The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res* **11**, 6883-6894 (1983).
50. S. E. Goetz, B. Vogelstein, S. R. Hamilton, A. P. Feinberg, Hypomethylation of DNA from benign and malignant human colon neoplasms. *Science* **228**, 187-190 (1985).
51. V. Greger, E. Passarge, W. Hopping, E. Messmer, B. Horsthemke, Epigenetic Changes May Contribute to the Formation and Spontaneous Regression of Retinoblastoma. *Hum. Genet.* **83**, 155-158 (1989).
52. T. Sakai *et al.*, Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *Am J Hum Genet* **48**, 880-888 (1991).
53. N. Ohtani-Fujita *et al.*, CpG methylation inactivates the promoter activity of the human retinoblastoma tumor-suppressor gene. *Oncogene* **8**, 1063-1067 (1993).
54. J. G. Herman *et al.*, Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A* **91**, 9700-9704 (1994).
55. A. Merlo *et al.*, 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nat Med* **1**, 686-692 (1995).
56. J. F. Costello, M. S. Berger, H. S. Huang, W. K. Cavenee, Silencing of p16/CDKN2 expression in human gliomas by methylation and chromatin condensation. *Cancer Res* **56**, 2405-2410 (1996).
57. M. Gonzalez-Zulueta *et al.*, Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. *Cancer Res* **55**, 4531-4535 (1995).
58. J. F. Costello, B. W. Futscher, R. A. Kroes, R. O. Pieper, Methylation-related chromatin structure is associated with exclusion of transcription factors from and suppressed expression of the O-6-methylguanine DNA methyltransferase gene in human glioma cell lines. *Mol. Cell. Biol.* **14**, 6515-6521 (1994).
59. R. O. Pieper *et al.*, Direct correlation between methylation status and expression of the human O-6-methylguanine DNA methyltransferase gene. *Cancer Commun* **3**, 241-253 (1991).
60. G. Zardo *et al.*, Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors. *Nat Genet* **32**, 453-458 (2002).
61. C. Hong, A. W. Bollen, J. F. Costello, The contribution of genetic and epigenetic mechanisms to gene silencing in oligodendrogliomas. *Cancer Res* **63**, 7600-7605 (2003).

62. C. Hong *et al.*, Shared epigenetic mechanisms in human and mouse gliomas inactivate expression of the growth suppressor SLC5A8. *Cancer Res* **65**, 3617-3623 (2005).
63. C. Hong *et al.*, Epigenome scans and cancer genome sequencing converge on WNK2, a kinase-independent suppressor of cell growth. *Proc Natl Acad Sci U S A* **104**, 10974-10979 (2007).
64. F. Gaudet *et al.*, Induction of tumors in mice by genomic hypomethylation. *Science* **300**, 489-492 (2003).
65. D. J. Weisenberger *et al.*, CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* **38**, 787-793 (2006).
66. T. Hinoue *et al.*, Analysis of the association between CIMP and BRAF in colorectal cancer by DNA methylation profiling. *PLoS ONE* **4**, e8357 (2009).
67. C. K. Kaufman *et al.*, A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. *Science* **351**, aad2197 (2016).
68. M. Toyota *et al.*, CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A* **96**, 8681-8686 (1999).
69. H. Shen, P. W. Laird, Interplay between the cancer genome and epigenome. *Cell* **153**, 38-55 (2013).
70. S. C. Mack, C. G. Hubert, T. E. Miller, M. D. Taylor, J. N. Rich, An epigenetic gateway to brain tumor cell identity. *Nat Neurosci* **19**, 10-19 (2015).
71. S. Behjati *et al.*, Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nature Genetics* **45**, 1479-1482 (2013).
72. J. Schwartzenuber *et al.*, Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482**, 226-231 (2012).
73. J. Zhang *et al.*, Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. *Nat Genet* **45**, 602-612 (2013).
74. G. Wu *et al.*, Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nature Genetics* **44**, 251-253 (2012).
75. J. A. Losman, W. G. Kaelin, Jr., What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes Dev* **27**, 836-852 (2013).
76. L. Dang *et al.*, Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739-744 (2009).
77. P. S. Ward *et al.*, The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **17**, 225-234 (2010).
78. H. Nouchmehr *et al.*, Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510-522 (2010).
79. M. E. Figueroa *et al.*, Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553-567 (2010).
80. W. Xu *et al.*, Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17-30 (2011).
81. R. Chowdhury *et al.*, The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep* **12**, 463-469 (2011).
82. S. Turcan *et al.*, IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* **483**, 479-483 (2012).

83. C. Lu *et al.*, IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474-478 (2012).
84. D. Rohle *et al.*, An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science* **340**, 626-630 (2013).
85. F. Wang *et al.*, Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. *Science* **340**, 622-626 (2013).
86. <http://www.clinicaltrials.gov>, accessed April 3, 2016.
87. C. Darwin, *On the origin of species by means of natural selection*. (J. Murray, London,, 1859), pp. ix, 1, 502 p.
88. J. G. Mendel, *Versuche über Pflanzenhybriden. Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865, Abhandlungen, 3–47*. (English translation at <http://www.mendelweb.org/>, 1866).
89. A. D. Hershey, M. Chase, Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* **36**, 39-56 (1952).
90. P. C. Nowell, The clonal evolution of tumor cell populations. *Science* **194**, 23-28 (1976).
91. M. Greaves, C. C. Maley, Clonal evolution in cancer. *Nature* **481**, 306-313 (2012).
92. L. M. Merlo, J. W. Pepper, B. J. Reid, C. C. Maley, Cancer as an evolutionary and ecological process. *Nat Rev Cancer* **6**, 924-935 (2006).
93. M. Gerlinger *et al.*, Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883-892 (2012).
94. M. J. Aryee *et al.*, DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. *Sci Transl Med* **5**, 169ra110 (2013).
95. D. Brocks *et al.*, Intratumor DNA methylation heterogeneity reflects clonal evolution in aggressive prostate cancer. *Cell Rep* **8**, 798-806 (2014).
96. M. Gerlinger, C. Swanton, How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *Br. J. Cancer* **103**, 1139-1143 (2010).
97. D. A. Landau *et al.*, Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714-726 (2013).
98. L. Ding *et al.*, Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506-510 (2012).
99. S. Yachida *et al.*, Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114-1117 (2010).
100. X. Wu *et al.*, Clonal selection drives genetic divergence of metastatic medulloblastoma. *Nature* **482**, 529-533 (2012).
101. M. Westphal, K. Lamszus, The neurobiology of gliomas: from cell biology to the development of therapeutic approaches. *Nat Rev Neurosci* **12**, 495-508 (2011).
102. See supplementary tables available online at <http://science.sciencemag.org/content/343/6167/189/suppl/DC1>.
103. Y. Jiao *et al.*, Frequent ATRX, CIC, FUBP1 and IDH1 mutations refine the classification of malignant gliomas. *Oncotarget* **3**, 709-722 (2012).
104. T. Watanabe, S. Nobusawa, P. Kleihues, H. Ohgaki, IDH1 mutations are early events in the development of astrocytomas and oligodendrogliomas. *Am J Pathol* **174**, 1149-1153 (2009).
105. K. Watanabe *et al.*, Incidence and timing of p53 mutations during astrocytoma progression in patients with multiple biopsies. *Clin Cancer Res* **3**, 523-530 (1997).

106. P. P. Medina *et al.*, Frequent BRG1/SMARCA4-inactivating mutations in human lung cancer cell lines. *Hum Mutat* **29**, 617-622 (2008).
107. S. Glaros, G. M. Cirrincione, A. Palanca, D. Metzger, D. Reisman, Targeted knockout of BRG1 potentiates lung cancer development. *Cancer Res* **68**, 3689-3696 (2008).
108. N. J. Szerlip *et al.*, Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response. *Proc Natl Acad Sci U S A* **109**, 3041-3046 (2012).
109. A. Sottoriva *et al.*, Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A* **110**, 4009-4014 (2013).
110. C. Greenman *et al.*, Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-158 (2007).
111. G. Reifenberger, J. Reifenberger, K. Ichimura, P. S. Meltzer, V. P. Collins, Amplification of multiple genes from chromosomal region 12q13-14 in human malignant gliomas: preliminary mapping of the amplicons shows preferential involvement of CDK4, SAS, and MDM2. *Cancer Res* **54**, 4299-4303 (1994).
112. H. Wang *et al.*, Analysis of the activation status of Akt, NFkappaB, and Stat3 in human diffuse gliomas. *Lab Invest* **84**, 941-951 (2004).
113. D. N. Louis, Molecular pathology of malignant gliomas. *Annu Rev Pathol* **1**, 97-117 (2006).
114. T. Tsai *et al.*, Rapid identification of germline mutations in retinoblastoma by protein truncation testing. *Arch Ophthalmol* **122**, 239-248 (2004).
115. C. Houdayer *et al.*, Comprehensive screening for constitutional RB1 mutations by DHPLC and QMPSF. *Hum Mutat* **23**, 193-202 (2004).
116. X. Q. Qin, T. Chittenden, D. M. Livingston, W. G. Kaelin, Jr., Identification of a growth suppression domain within the retinoblastoma gene product. *Genes Dev* **6**, 953-964 (1992).
117. J. Koh, G. H. Enders, B. D. Dynlacht, E. Harlow, Tumour-derived p16 alleles encoding proteins defective in cell-cycle inhibition. *Nature* **375**, 506-510 (1995).
118. S. A. Forbes *et al.*, The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet* **Chapter 10**, Unit 10 11 (2008).
119. M. C. Fargnoli *et al.*, CDKN2a/p16INK4a mutations and lack of p19ARF involvement in familial melanoma kindreds. *J Invest Dermatol* **111**, 1202-1206 (1998).
120. S. Kang, A. G. Bader, P. K. Vogt, Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic. *Proc Natl Acad Sci U S A* **102**, 802-807 (2005).
121. S. Y. Han *et al.*, Functional evaluation of PTEN missense mutations using in vitro phosphoinositide phosphatase assay. *Cancer Res* **60**, 3147-3151 (2000).
122. T. Sato, A. Nakashima, L. Guo, K. Coffman, F. Tamanoi, Single amino-acid changes that confer constitutive activation of mTOR are discovered in human cancer. *Oncogene* **29**, 2746-2752 (2010).
123. J. Okosun *et al.*, Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. *Nat Genet* **46**, 176-181 (2014).
124. M. Gerlinger *et al.*, Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet* **46**, 225-233 (2014).
125. C. C. Oakes *et al.*, Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia. *Cancer Discov* **4**, 348-361 (2014).

126. B. E. Johnson *et al.*, Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* **343**, 189-193 (2014).
127. S. D. Fouse, J. F. Costello, Epigenetics of neurological cancers. *Future Oncol* **5**, 1615-1629 (2009).
128. N. Baeza, M. Weller, Y. Yonekawa, P. Kleihues, H. Ohgaki, PTEN methylation and expression in glioblastomas. *Acta Neuropathol* **106**, 479-485 (2003).
129. T. Y. Kim, S. Zhong, C. R. Fields, J. H. Kim, K. D. Robertson, Epigenomic profiling reveals novel and frequent targets of aberrant DNA methylation-mediated silencing in malignant glioma. *Cancer Res* **66**, 7490-7501 (2006).
130. R. P. Nagarajan *et al.*, Recurrent epimutations activate gene body promoters in primary glioblastoma. *Genome Res* **24**, 761-774 (2014).
131. M. Nakamura, Y. Yonekawa, P. Kleihues, H. Ohgaki, Promoter hypermethylation of the RB1 gene in glioblastomas. *Lab Invest* **81**, 77-82 (2001).
132. J. K. Wiencke *et al.*, Methylation of the PTEN promoter defines low-grade gliomas and secondary glioblastoma. *Neuro Oncol* **9**, 271-279 (2007).
133. A. Lai *et al.*, Evidence for sequenced molecular evolution of IDH1 mutant glioblastoma from a distinct cell of origin. *J Clin Oncol* **29**, 4482-4490 (2011).
134. V. K. Hill *et al.*, Stability of the CpG island methylator phenotype during glioma progression and identification of methylated loci in secondary glioblastomas. *BMC Cancer* **14**, 506 (2014).
135. F. Banine *et al.*, SWI/SNF chromatin-remodeling factors induce changes in DNA methylation to promote transcriptional activation. *Cancer Res* **65**, 3542-3547 (2005).
136. R. J. Gibbons *et al.*, Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat Genet* **24**, 368-371 (2000).
137. J. F. Costello, B. W. Futscher, K. Tano, D. M. Graunke, R. O. Pieper, Graded methylation in the promoter and body of the O6-methylguanine DNA methyltransferase (MGMT) gene correlates with MGMT expression in human glioma cells. *J Biol Chem* **269**, 17228-17237 (1994).
138. S. Everhard *et al.*, MGMT methylation: a marker of response to temozolomide in low-grade gliomas. *Ann Neurol* **60**, 740-743 (2006).
139. S. Kesari *et al.*, Phase II study of protracted daily temozolomide for low-grade gliomas in adults. *Clin Cancer Res* **15**, 330-337 (2009).
140. W. Taal *et al.*, First-line temozolomide chemotherapy in progressive low-grade astrocytomas after radiotherapy: molecular characteristics in relation to response. *Neuro Oncol* **13**, 235-241 (2011).
141. W. Wick *et al.*, Prognostic or predictive value of MGMT promoter methylation in gliomas depends on IDH1 mutation. *Neurology* **81**, 1515-1522 (2013).
142. J. Laffaire *et al.*, Methylation profiling identifies 2 groups of gliomas according to their tumorigenesis. *Neuro Oncol* **13**, 84-98 (2011).
143. D. Sproul *et al.*, Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns. *Genome Biol* **13**, R84 (2012).
144. T. Witte, C. Plass, C. Gerhauser, Pan-cancer patterns of DNA methylation. *Genome Med* **6**, 66 (2014).
145. See supplementary tables available online at <http://www.sciencedirect.com/science/article/pii/S1535610815002640> - mmc2.

146. J. P. Issa *et al.*, Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat Genet* **7**, 536-540 (1994).
147. F. Eckhardt *et al.*, DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* **38**, 1378-1385 (2006).
148. K. Kerkerl *et al.*, Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet* **40**, 904-908 (2008).
149. D. A. Landau *et al.*, Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell* **26**, 813-825 (2014).
150. H. Heyn *et al.*, DNA methylation contributes to natural human variation. *Genome Res* **23**, 1363-1372 (2013).
151. J. Shi *et al.*, Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. *Nat Commun* **5**, 3365 (2014).
152. A. Rufini *et al.*, p73 in Cancer. *Genes Cancer* **2**, 491-502 (2011).
153. H. Z. Chen, S. Y. Tsai, G. Leone, Emerging roles of E2Fs in cancer: an exit from cell cycle control. *Nat Rev Cancer* **9**, 785-797 (2009).
154. H. Suzuki *et al.*, Mutational landscape and clonal architecture in grade II and III gliomas. *Nat Genet* **47**, 458-468 (2015).
155. G. Jin *et al.*, 2-hydroxyglutarate production, but not dominant negative function, is conferred by glioma-derived NADP-dependent isocitrate dehydrogenase mutations. *PLoS ONE* **6**, e16812 (2011).
156. B. Pietrak *et al.*, A tale of two subunits: how the neomorphic R132H IDH1 mutation enhances production of alphaHG. *Biochemistry* **50**, 4804-4812 (2011).
157. L. B. Bralten *et al.*, IDH1 R132H decreases proliferation of glioma cell lines in vitro and in vivo. *Ann Neurol* **69**, 455-463 (2011).
158. P. S. Ward *et al.*, The potential for isocitrate dehydrogenase mutations to produce 2-hydroxyglutarate depends on allele specificity and subcellular compartmentalization. *J Biol Chem* **288**, 3804-3815 (2013).
159. H. Yan *et al.*, IDH1 and IDH2 mutations in gliomas. *N Engl J Med* **360**, 765-773 (2009).
160. T. Schumacher *et al.*, A vaccine targeting mutant IDH1 induces antitumour immunity. *Nature* **512**, 324-327 (2014).
161. S. Pusch *et al.*, Glioma IDH1 mutation patterns off the beaten track. *Neuropathol Appl Neurobiol* **37**, 428-430 (2011).
162. G. Jin *et al.*, Disruption of wild-type IDH1 suppresses D-2-hydroxyglutarate production in IDH1-mutated gliomas. *Cancer Res.* **73**, 496-501 (2013).
163. H. A. Luchman, C. Chesnelong, J. G. Cairncross, S. Weiss, Spontaneous loss of heterozygosity leading to homozygous R132H in a patient-derived IDH1 mutant cell line. *Neuro Oncol* **15**, 979-980 (2013).
164. F. Favero *et al.*, Glioblastoma adaptation traced through decline of an IDH1 clonal driver and macro-evolution of a double-minute chromosome. *Annals of Oncology* **26**, 880-887 (2015).
165. T. Mazar *et al.*, DNA methylation and somatic mutations converge on the cell cycle and define similar evolutionary histories in brain tumors. *Cancer Cell* **28**, 307-317 (2015).
166. S. Pusch *et al.*, D-2-Hydroxyglutarate producing neo-enzymatic activity inversely correlates with frequency of the type of isocitrate dehydrogenase 1 mutations found in glioma. *Acta Neuropathol Commun* **2**, 19 (2014).

167. C. W. Brennan *et al.*, The Somatic Genomic Landscape of Glioblastoma. *Cell* **155**, 462-477 (2013).
168. U. C. Okoye-Okafor *et al.*, New IDH1 mutant inhibitors for treatment of acute myeloid leukemia. *Nat Chem Biol* **11**, 878-886 (2015).
169. C. L. Kleinman *et al.*, Fusion of TTYH1 with the C19MC microRNA cluster drives expression of a brain-specific DNMT3B isoform in the embryonal brain tumor ETMR. *Nat Genet* **46**, 39-44 (2014).
170. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
171. A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303 (2010).
172. M. A. DePristo *et al.*, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-498 (2011).
173. K. Cibulskis *et al.*, Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-219 (2013).
174. K. Ye, M. H. Schulz, Q. Long, R. Apweiler, Z. Ning, Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865-2871 (2009).
175. K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
176. 1000 Genomes Project Consortium *et al.*, A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-1073 (2010).
177. S. Rozen, H. Skaletsky, Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365-386 (2000).
178. B. C. Christensen *et al.*, DNA methylation, isocitrate dehydrogenase mutation, and survival in glioma. *J Natl Cancer Inst* **103**, 143-153 (2011).
179. M. F. Berger *et al.*, The genomic complexity of primary human prostate cancer. *Nature* **470**, 214-220 (2011).
180. M. A. Chapman *et al.*, Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467-472 (2011).
181. E. S. Venkatraman, A. B. Olshen, A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657-663 (2007).
182. B. S. Taylor *et al.*, Functional copy-number alterations in cancer. *PLoS ONE* **3**, e3179 (2008).
183. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
184. F. Favero *et al.*, Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol* **26**, 64-70 (2015).
185. A. B. Olshen *et al.*, Parent-specific copy number in paired tumor-normal studies using circular binary segmentation. *Bioinformatics* **27**, 2038-2046 (2011).
186. P. Van Loo *et al.*, Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A* **107**, 16910-16915 (2010).
187. S. Nik-Zainal *et al.*, The life history of 21 breast cancers. *Cell* **149**, 994-1007 (2012).
188. T. J. Triche, Jr., D. J. Weisenberger, D. Van Den Berg, P. W. Laird, K. D. Siegmund, Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res* **41**, e90 (2013).

189. M. E. Price *et al.*, Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin* **6**, 4 (2013).
190. G. K. Smyth, Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3 (2004).
191. G. Yu, L. G. Wang, Y. Han, Q. Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287 (2012).
192. F. Krueger, S. R. Andrews, Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572 (2011).
193. A. K. Maunakea *et al.*, Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253-257 (2010).
194. M. Esteller, S. R. Hamilton, P. C. Burger, S. B. Baylin, J. G. Herman, Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia. *Cancer Res* **59**, 793-797 (1999).
195. C. Rohde, Y. Zhang, R. Reinhardt, A. Jeltsch, BISMAR--fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. *BMC Bioinformatics* **11**, 230 (2010).
196. K. C. Wiegand *et al.*, ARID1A mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med* **363**, 1532-1543 (2010).
197. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).
198. C. Trapnell *et al.*, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-515 (2010).
199. M. Griffith *et al.*, Alternative expression analysis by RNA sequencing. *Nat Methods* **7**, 843-847 (2010).
200. A. Subramanian *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
201. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**, 9440-9445 (2003).
202. P. J. Campbell *et al.*, Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc Natl Acad Sci U S A* **105**, 13081-13086 (2008).
203. R. Desper, O. Gascuel, Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J Comput Biol* **9**, 687-705 (2002).
204. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289-290 (2004).
205. S. Meiboom, D. Gill, Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.* **29**, 688 (1958).
206. A. M. Weljie, J. Newton, P. Mercier, E. Carlson, C. M. Slupsky, Targeted profiling: quantitative analysis of ¹H NMR metabolomics data. *Anal Chem* **78**, 4430-4442 (2006).
207. J. Ernst *et al.*, Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49 (2011).
208. N. A. Rizvi *et al.*, Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124-128 (2015).

209. A. Snyder *et al.*, Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* **371**, 2189-2199 (2014).
210. D. T. Le *et al.*, PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* **372**, 2509-2520 (2015).
211. E. M. Van Allen *et al.*, Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207-211 (2015).
212. M. M. Gubin, M. N. Artyomov, E. R. Mardis, R. D. Schreiber, Tumor neoantigens: building a framework for personalized cancer immunotherapy. *J Clin Invest* **125**, 3413-3421 (2015).
213. M. M. Gubin, R. D. Schreiber, The odds of immunotherapy success. *Science* **350**, 158-159 (2015).
214. E. Bouffet *et al.*, Immune Checkpoint Inhibition for Hypermutant Glioblastoma Multiforme Resulting From Germline Biallelic Mismatch Repair Deficiency. *J Clin Oncol*, (2016).
215. R. M. Brena, J. F. Costello, Genome-epigenome interactions in cancer. *Hum Mol Genet* **16 Spec No 1**, R96-105 (2007).
216. S. Aumann, O. Abdel-Wahab, Somatic alterations and dysregulation of epigenetic modifiers in cancers. *Biochem Biophys Res Commun* **455**, 24-34 (2014).
217. E. G. Shaw *et al.*, Randomized trial of radiation therapy plus procarbazine, lomustine, and vincristine chemotherapy for supratentorial adult low-grade glioma: initial results of RTOG 9802. *J Clin Oncol* **30**, 3065-3070 (2012).
218. D. W. Parsons *et al.*, An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807-1812 (2008).
219. <http://www.clinicaltrials.gov>, accessed May 2, 2016.
220. J. H. Sampson *et al.*, Immunologic escape after prolonged progression-free survival with epidermal growth factor receptor variant III peptide vaccination in patients with newly diagnosed glioblastoma. *J Clin Oncol* **28**, 4722-4729 (2010).

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

5/25/16

Date