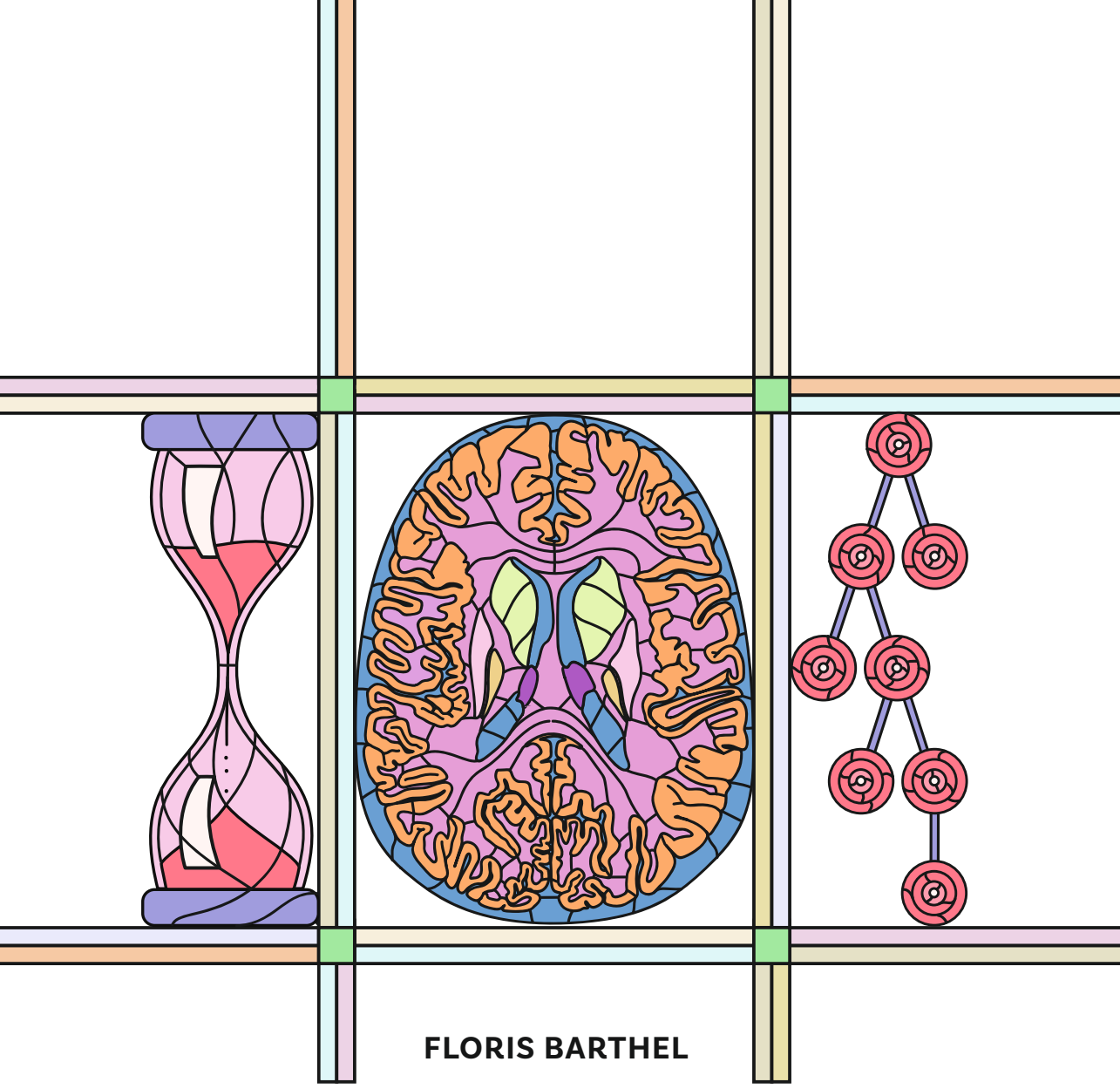


FLORIS BARTHEL

The Molecular Anatomy of Diffuse Gliomas in Adults

The Molecular Anatomy of Diffuse Gliomas in Adults

Floris Pieter Barthel

Financial support for the publication of this thesis is gratefully acknowledged and provided by Hersen Tumor Fonds and Stichting STOPhersentumoren.nl

Cover art: Matt Wimsatt
Layout and design: Erwin Timmerman, persoonlijkproefschrift.nl
Provided by thesis specialist Ridderprint, ridderprint.nl

ISBN: 978-94-6458-444-8

VRIJE UNIVERSITEIT

The Molecular Anatomy of Diffuse Gliomas in Adults

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op donderdag 15 september 2022 om 13.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door

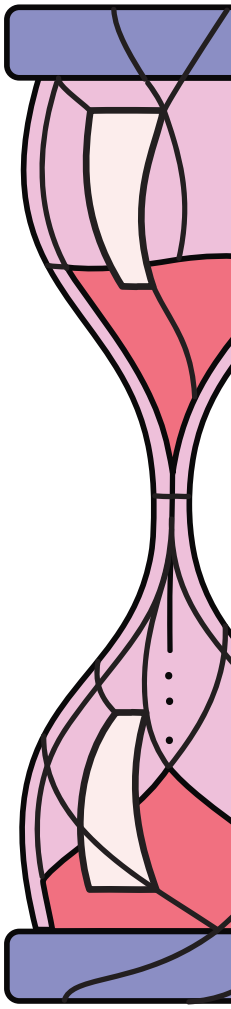
Floris Pieter Barthel

geboren te Naarden

promotor:	prof.dr. R.G.W. Verhaak
copromotor:	prof.dr. P. Wesseling
promotiecommissie:	prof.dr. P.C. de Witt Hamer dr. D. Brandsma prof.dr.ir. E. Cuppen dr. P.J. French dr. S.P. Niclou prof.dr. F.C.P. Holstege

For Nina

Table of Contents



CHAPTER 1

Introduction

This chapter is based on the following review papers

Barthel FP, Johnson KC, Wesseling P, Verhaak RGW. *Evolving Insights into the Molecular Neuropathology of Diffuse Gliomas in Adults.* Neurol Clin. 2018 Aug;36(3):421-437.

Barthel FP, Wesseling P, Verhaak RGW. *Reconstructing the molecular life history of gliomas.* Acta Neuropathol. 2018 May;135(5):649-670.

Gliomas encompass a very diverse group and account for the great majority of tumors originating in the parenchyma of the central nervous system (CNS). Every year in the United States, 6.56 in 100.000 people are newly diagnosed with a glioma, and 4.2 in 100.000 succumb to the disease [1]. Two larger glioma groups are recognized: so-called diffuse gliomas, characterized by extensive infiltrative growth into the surrounding CNS parenchyma, and more circumscribed (non-diffuse) gliomas such as pilocytic astrocytoma and ependymomas. Diffuse gliomas, by far the most frequent gliomas in adult patients, are traditionally classified according to their microscopic similarities with (precursors of) glial cells and then designated as diffuse astrocytomas, oligodendrogliomas or mixed gliomas/oligoastrocytomas. Additionally, a malignancy grade is assigned to these tumors based on presence/absence of especially marked mitotic activity, florid microvascular proliferation (MVP), and necrosis [2, 3]. For over a century, such microscopic evaluation has provided the gold standard for the diagnosis of gliomas, assessment of prognosis and formed the basis for therapeutic management. However, multiple studies showed that a purely histopathologic classification suffers from considerable inter- and intraobserver variability [4-6].

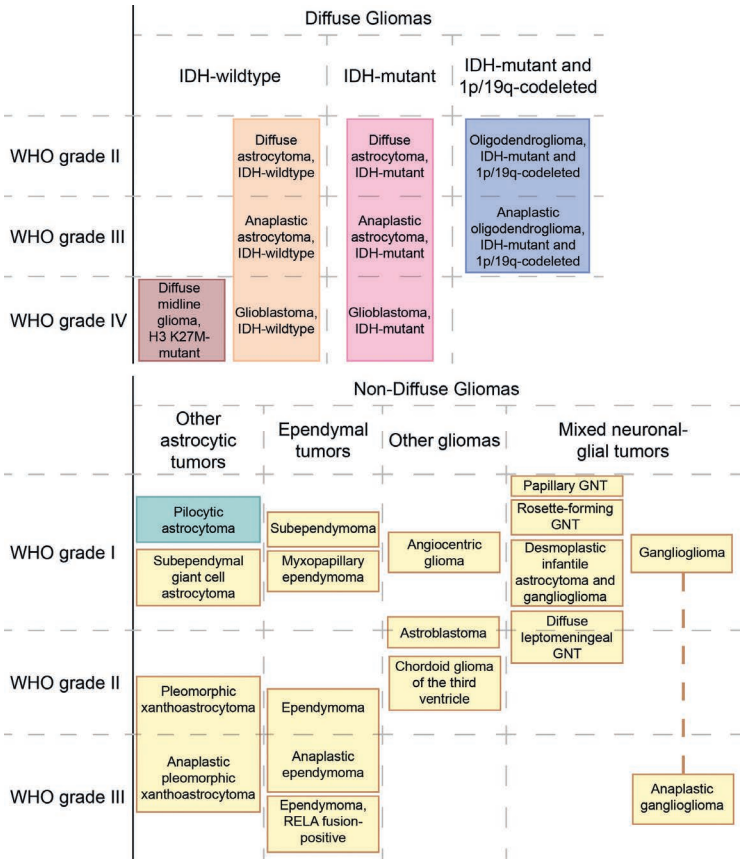
As with other human cancers, the pathogenesis and molecular evolution of gliomas is often characterized by somatic chromosomal aberrations, widespread or focal copy number changes and targeted gain and loss of function events in oncogenes and tumor suppressor genes [7-9]. Various permutations of somatic alterations were shown to be associated with distinct tumor entities and differential sensitivities to treatment, such as a chromosome 1p/19q-codeletion in oligodendrogliomas conferring increased sensitivity to chemotherapy [10, 11]. In the course of the last two decades it became increasingly clear that such molecular characteristics may provide a more robust and objective basis for subtyping of diffuse gliomas and both scientists and clinicians turned towards molecular markers to aid diagnosis [12-16].

In adult patients three major subgroups of diffuse glioma are now defined based on the presence or absence of mutations in the isocitrate dehydrogenase 1 (*IDH1*) or *IDH2* gene and of complete, combined loss of the short arm of chromosome 1 and of the long arm of chromosome 19 (complete 1p/19q-codeletion):

- IDH-wildtype: most of these histologically represent astrocytic tumors, a large percentage belonging to the highest malignancy grade, i.e. glioblastomas
- IDH-mutant and 1p/19q-non-codeleted: these tumors also generally have an astrocytic phenotype, but a much larger percentage is at first diagnosis histologically lower grade/WHO grade II or III

- IDH-mutant and 1p/19-codeleted: most of these are characterized by a prominent oligodendroglial phenotype of the tumor cells

The focus of this thesis is on these diffuse gliomas, however in some cases tumors from the group of non-diffuse gliomas are referenced (**Figure 1-1**). The most common non-diffuse gliomas are pilocytic astrocytomas that predominantly occur in children, generally are more circumscribed (therefore grouped under non-diffuse gliomas) and show an indolent behavior [17-19]. The transition from a purely histological to a histo-molecular classification of especially diffuse gliomas represents a paradigm shift and necessitates re-evaluation of histologic criteria used for grading and guidance of therapeutic decisions [20, 21].



↑ **Figure 1-1. The WHO classification of diffuse and non-diffuse gliomas. Tumors are grouped by WHO grading and putative cell-of-origin.**

Summary of overall aims and outline of this thesis

The clinical classification of CNS tumors including glioma is determined by standards put forward by the World Health Organization (WHO, **Figure 1-1**). There are high rates of inter- and intra-observer diagnostic disagreement using classic histopathology, which was considered the diagnostic standard at the onset of starting this PhD trajectory [4-6]. Large consortium characterization efforts have used molecular profiling to define clinically relevant glioma subtypes in an unbiased manner [22-24]. These efforts provide a basis for **Chapter 2** of this thesis, which aims to further refine the molecular disease classification using patterns in DNA methylation, copy number and mutations across the spectrum of histological types.

This work indicated that glioma subtypes are hallmarked by common sets of driver alterations. I reasoned that effective treatment should move a tumor in a predictable manner due to the natural selection of features enabling treatment resistance. Using a large set of pre- and post-treatment tumor samples I set out to determine the longitudinal landscape of molecular changes over time and identify selection pressures exerted by treatment in **Chapter 3**. As described in more detail in this introduction, the neuro-oncology community is increasingly adopting DNA methylation profiling for diagnostic purposes. Tumor heterogeneity is thought to bias diagnostic accuracy and in **Chapter 4** I aim to dissect the extent of heterogeneity in DNA methylation-based tumor classification when considering multiple spatially separated samples from the same tumor.

While it is well known that alkylating chemotherapies may in some cases lead to the accrual of a massive amount of point mutations, called hypermutation, the genomic effects of ionizing radiation are less well understood. Using the same set of pre- and post-treatment tumor samples introduced in **Chapter 3**, I take a deep-dive to investigate the genomic effects of radiotherapy in **Chapter 5**. The work described in **Chapter 2** suggested that glioma cells from different disease subtypes follow distinct strategies for maintaining the ends of chromosomes called telomeres. In **Chapter 6** I follow up on these findings by mapping strategies for telomere maintenance across a large assortment of tumor types.

In **Chapter 7**, I integrate the main observations of previous chapters into a model for tumor development. Finally, in **Chapter 8**, I summarize the main findings of this thesis and discuss potential for translation and relevance to clinical practice. To provide context to the ensuing chapters, this introductory **Chapter 1** first elaborates on the history and background of diffuse gliomas, molecular profiling and telomere maintenance.

A brief history of the (molecular) pathology of gliomas

The term "glioma" was first introduced by the German pathologist Rudolf Virchow in the 1850's (**Figure 1-2**), and in 1925 Cushing and Bailey introduced the entity "glioblastoma multiforme" for a high-grade malignant glioma showing a wide range of histological features (hence 'multiforme') [25, 26]. Later on, (neuro) pathologists like Ringertz, Scherer, Broders and Kernohan provided important next building blocks for systematic histopathological classification (typing and grading) of gliomas [27-31].

Even after the publication of the first edition of the World Health Organization (WHO) classification of tumors of the central nervous system (CNS) in 1979, different schemes for typing and grading of diffuse gliomas were used in parallel [32]. However, the second edition of the WHO classification (published in 1993) was much more universally accepted as the standard for glioma classification [33, 34]. For grading of astrocytomas this latter classification incorporated elements of the St. Anne-Mayo grading approach in which absence or presence of mitotic activity, microvascular proliferation and necrosis was used to assign a malignancy grade [35].

The third and fourth edition of the WHO classification of CNS tumors (published in respectively 2000 and 2007) were built on essentially the same approach of histopathology-based diagnosis of diffuse gliomas, in some situations supported by the use of immunohistochemical markers [36-40]. However, despite being the time-honored diagnostic gold standard, it was increasingly clear that histopathologic classification of diffuse gliomas suffers from considerable inter- and intraobserver variability, even amongst expert-(neuro)pathologists, and that the use of molecular markers had great potential to substantially improve the unequivocal discrimination of clinically relevant diffuse glioma subgroups [4-6, 41-43].

The findings in 1994 and 1998 that gliomas with a combined deletion of chromosome arms 1p and 19q (1p/19q codeletion) were associated with significantly improved survival and increased sensitivity to combinatorial chemotherapy with procarbazine, lomustine and vincristine (PCV) were the first discoveries to pave the way for molecular neuropathology of CNS tumors [10, 11, 44-46]. Typically, 1p/19q-codeleted tumors showed oligodendroglial histology, but the codeletion was reported to occur across different types and grades of diffuse gliomas and still carry its prognostic and predictive impact [11]. Of note, the apparent chemosensitivity of oligodendrogliomas was already reported in 1988, years before a connection was made to the 1p/19q codeletion in 1998 [47]. The fact that (com-

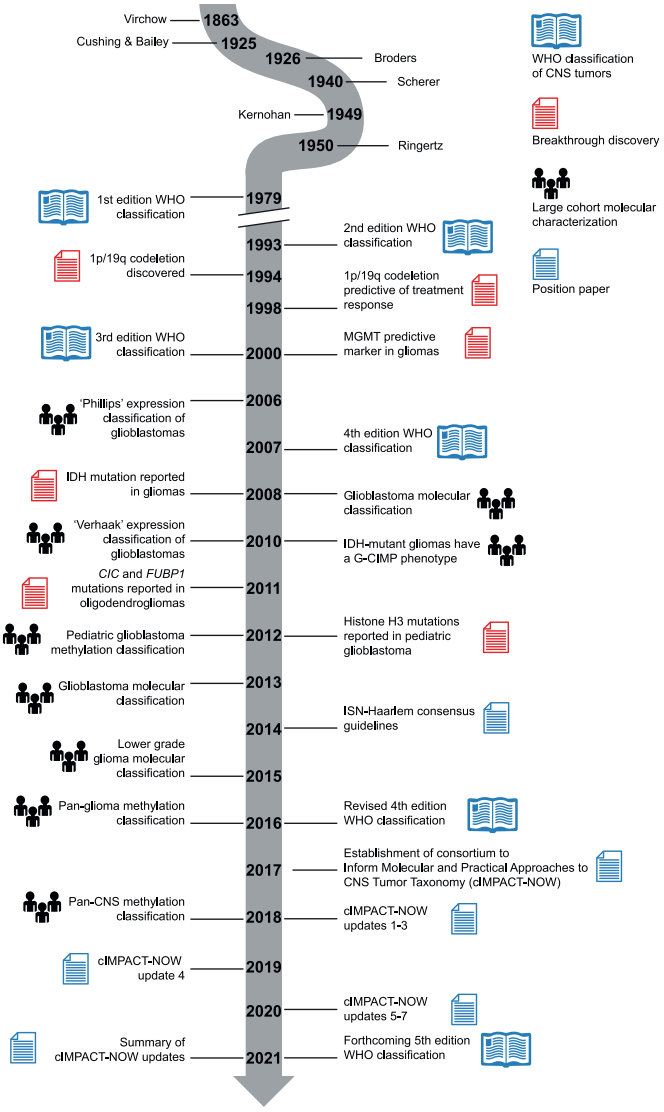
plete) 1p/19q-codeleted tumors responded relatively well to PCV treatment led to introduction of 1p/19q-testing in clinical practice before this molecular marker became a part of the WHO diagnostic criteria for a subset of diffuse gliomas [10].

Another finding with a major impact on the neuropathology of diffuse gliomas is the discovery of mutations of the isocitrate dehydrogenase 1 (*IDH1*) gene and the related *IDH2* gene [16, 48, 49]. IDH-mutant and IDH-wildtype astrocytomas are clinically different tumors despite overlapping histological appearances. IDH mutations were first reported in glioblastomas and later on discovered to be much more prevalent amongst lower grade (WHO grade II and III) gliomas [49]. IDH mutations contribute to gliomagenesis by activating alternative transcriptional programs via the genome-wide disruption of DNA methylation [50]. It has been proposed that this stochastic process is subject to careful selection in an effort to prioritize epigenetic changes that promote tumorigenesis [51]. Using modern genomic sequencing technologies, several groups independently reported that histologically similar gliomas could be subdivided into distinct tumor entities based on mutations in IDH and the 1p/19q codeletion, providing further fuel to the notion that a reclassification of these tumors is needed [12, 13, 52].

Although the methylation status of the DNA repair enzyme O(6)-methylguanine-DNA methyltransferase (*MGMT)* does not distinguish between particular molecular subtypes of glioma, it is an important predictive biomarker. As was first reported in 2000, epigenetic silencing of the *MGMT* gene by methylation of the promoter appears to compromise conventional DNA repair mechanisms and thereby increases sensitivity to conventional chemotherapy using alkylating agents such as the widely used temozolomide [53-56]. Especially in elderly patients with glioblastoma lacking *MGMT* promoter methylation the benefit of temozolomide treatment often does not seem to outweigh the negative side-effects [57, 58].

Because of the rapidly increasing insights in the diagnostic potential of molecular markers in glial (and other) CNS tumors, in 2014 a meeting was organized in Haarlem, The Netherlands under the sponsorship of the International Society of Neuropathology (ISN) and focusing on how molecular information could be optimally incorporated into a next WHO classification. The white paper that resulted from this meeting provided the basis for the design of an integrated histo-molecular classification of especially glial and embryonal CNS tumors as represented in the 2016 (i.e. revised fourth edition) of the WHO classification [59-61]. Indeed, state-of-the-art diagnosis of diffuse gliomas now requires assessment of presence or absence of IDH mutation and 1p/19q codeletion. Meanwhile, a “Not

Otherwise Specified" (NOS) category was created for cases were information on these defining molecular features is lacking, e.g. because molecular testing was not available, not informative or not performed.



↑ **Figure 1-2. A timeline of events in the history of molecular neuropathology.**

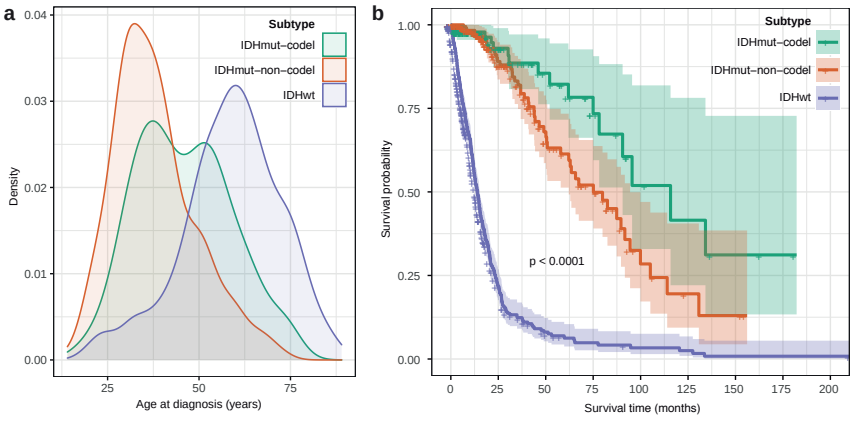
This selection of events is categorized into breakthrough discoveries, large cohort molecular characterizations, position papers and different editions of the WHO classification of CNS tumors. ISN = International Society of Neuropathology.

Clinical and molecular features of diffuse gliomas

Per the introduction of the 2016 revised 4th edition WHO classification IDH mutation and 1p/19q codeletion status have become so-called "subtype-defining" features. Diffuse gliomas that are both IDH-mutant and 1p/19q-codeleted are classified as oligodendroglioma, whereas tumors lacking (complete) codeletion of these chromosome arms are classified as astrocytoma and can be further separated based on IDH-status. Because of this, the ambivalent oligoastrocytoma diagnosis has largely disappeared in favor of molecularly defined subtypes, barring rare cases that contain a mixture of glioma cells carrying the codeletion and of oligodendroglial appearance, and including astrocytic cells with intact 1p/19q [62].

Clinical features such as age at diagnosis and expected survival vary widely by subtype, contributing greatly to its perceived clinical and biological relevance (**Figure 1-3**). On average, IDH-mutant astrocytomas and glioblastomas are diagnosed at a median age of 38 years and demonstrate a median survival after diagnosis of 75 months overall, although this further varies by WHO grade. With a median survival time of 116 months, IDH-mutant and 1p/19q-codeleted oligodendrogliomas demonstrate the highest survival rate amongst diffuse gliomas. With a median age at diagnosis of 46 years, patients with these tumors are generally older compared to those with IDH-mutant astrocytomas. In adults, IDH-wildtype diffuse gliomas are more frequently diagnosed in older patients (median age at diagnosis 59 years) and show the least favorable prognosis with a median survival of 14.0 months.

In addition to subtype defining molecular changes, glioma subtypes can be characterized by additional somatic mutations or copy number changes in a number of genes across various pathways, substantiating the hypothesis that the subtypes follow unique gliomagenic trajectories and represent different biologies (**Figure 1-4**). The landscape of molecular alterations by WHO diagnosis and tumor subtype are discussed at length in **Chapter 2**, and the evolution of molecular alterations over time in recurrent tumors is the focus of **Chapter 3**.



↑ Figure 1-3. Clinical characteristics of diffuse gliomas separated by subtype.

a. Density plot showing the distributin of age at diagnosis by subtype. **b.** Kaplan-Meier survival curves showing survival by subtype. Log-rank P-value is indicated. Tick marks show censoring.

Molecular category		IDH-wildtype	IDH-mutant & 1p/19q-non-codeleted	IDH-mutant & 1p/19q-codeleted
Histological type and WHO grade	Grade II	Diffuse astrocytoma	Diffuse astrocytoma	Oligodendroglioma
	Grade III	Anaplastic astrocytoma	Anaplastic astrocytoma	Anaplastic oligodendroglioma
	Grade IV	Glioblastoma	Glioblastoma	
Additional molecular features	Cell Cycle	*CDKN2A/B* 60% *TP53* 23% *CDK4* 15% *MDM4* 12% *RB1* 9% *MDM2* 8%	*TP53* 90% *CCND2* 9% *CDKN2A/B* 8%	
	RTK/PI3K	*EGFR* 54% *PTEN* 28% *PDGFRA* 13% *NF1* 13% *PIK3CA* 11% *PIK3R1* 7%		*PIK3CA* 12% *PIK3R1* 5%
	TMM	TPM 78%	*ATRX* 75%	TPM 98%
	Other			*CIC* 46% *FUBP1* 27% *NOTCH1* 14% *NIPBL* 7%

Oncogene
Tumor suppressor

↑ Figure 1-4. Groups of diffuse gliomas classified according to IDH mutation and 1p/19q codeletion status demonstrate a distinct landscape of molecular features.

Gene pathways are indicated. Genes in orange indicate genes preferentially targeted by gain-of-function (such as hotspot) mutational events or amplification events. Genes marked in blue indicate genes commonly affected by loss-of-function mutation (such as frameshift) or deletion events. Only features with frequencies greater than five percent are shown. Frequencies were derived from a recent publication[15]. TPM = TERT promoter mutation, RTK = receptor tyrosine kinase, PI3K = phosphoinositide-3 kinase, TMM = telomere maintenance mechanism.

Clinical management of diffuse glioma

While there has been significant progress in disease classification and our understanding of its biology, this has not translated into substantial improvements in treatment or patient outcomes. In fact, the optimal management of glioma remains controversial, and differences in the clinical management of these tumors is widespread. Surgery, radiotherapy and chemotherapy remain at the cornerstone of glioma treatment and combinations thereof form the basis of treatment regardless of tumor subtype.

The first step in the clinical management of a suspected glioma consists of maximal surgical resection, irrespective of tumor subtype. Historically, suspected lower grade tumors were in some cases withheld immediate surgical treatment or limited to a diagnostic biopsy, a treatment strategy intended to delay treatment resistance, although this practice has largely been done away with due to contradictory evidence [63, 64].

Postoperative treatment in glioma patients consists of varying combinations of radio- and chemotherapy and is dependent on tumor subtype, grade and extent of resection. Newly diagnosed grade IV glioblastoma is traditionally managed using a combination of radiotherapy and concomitant temozolomide (an alkylating chemotherapeutic agent) followed by at least six cycles of adjuvant temozolomide according to results from a large trial on this regiment [65, 66]. Recommendations for the treatment of lower grade glioma are less well defined. Lower grade oligodendrogliomas tend to respond well to chemotherapy with PCV, making chemotherapy a first-line treatment modality for the management of these tumors [47]. Nevertheless, in patients with grade II tumors that received a gross total resection, delaying postoperative treatment can postpone the toxicities of therapy while not having an adverse effect on survival [67-69].

The majority of practice defining studies were done prior to the introduction of the revised 4th edition WHO classification, necessitating retrospective reanalysis of prior results in light of molecular tumor subtypes and requiring future studies to separate patient accordingly. Surprisingly, results from the ongoing CATNON trial suggest that IDH wild-type tumors do not benefit from adjuvant temozolomide, bringing into question a decades old paradigm in glioma management. These concerning new data, combined with already dismal outcomes despite optimized combinations of conventional therapies, suggest that new treatment options for glioma are urgently needed.

In **Chapter 3** we investigated whether we could use the genomics of paired primary (untreated) and recurrent (post-treatment) tumors to gain insights into

treatment failure and tumor evolution. In **Chapter 5** we specifically looked at effects of radiotherapy on the genomes of these tumors in order to identify novel therapeutic vulnerabilities. Finally, in **Chapter 8** we speculate on the role of novel therapies, such as targeted molecular therapies, in providing new avenues for treatment and new hope for improved disease outcomes.

Transcriptome and methylome profiling to identify tumor subtypes

Tumor subtyping is a field that is of interest to clinicians and basic scientists alike because it has enabled the discovery of novel tumor entities that are biologically and clinically distinct. For basic scientists, it is of interest to discern biologically distinct tumor subgroups in order to identify new potential therapeutic targets. For clinicians, subtyping tumors can help inform patient prognosis and treatment strategy.

By studying large gene expression datasets, researchers were able to identify subtypes characterized by the activation of distinct transcriptional pathways, such as a "mesenchymal" subtype enriched in angiogenesis and inflammatory genes, a "classical" subtype enriched in stem cell and cell cycle genes, a "proneural" subtype enriched in neurodevelopmental genes, and a "neural" subtype enriched in adult neural markers [22, 70]. With the growing popularity of expression subtyping, tumor heterogeneity has become an increasingly important concern. Several studies demonstrated that multiple samples from the same tumor can be classified according to different transcriptomic subtypes [71-73]. In a recent revision to the glioma subtypes by our group we classified IDH-wildtype tumors after discarding a set of genes enriched in non-tumor tissue and were no longer able to detect the previously described neural subtype, suggesting this subtype was not a glioma-intrinsic subtype and reflected contamination from non-tumor tissue [74]. Indeed, because of their diffusely infiltrative growth pattern one can expect that diffuse gliomas contain a variable amount of 'contamination' of non-neoplastic parenchymal cells [75-77]. Altogether, these studies highlight the importance of tumor heterogeneity in the transcriptional profiling in gliomas, and the need to consider contributions from individual cell types.

DNA methylation profiling has emerged as another method capable of distinguishing gliomas by tumor type and provides complementary molecular information to transcriptional profiles. DNA methylation refers to the covalent addition of a methyl group to the fifth position carbon of a cytosine nucleotide resulting in 5-methylcytosine. This epigenetic modification occurs primarily in the context of nucleotide-pairs in which cytosine is followed by guanine (i.e., CpG dinucleotides)

and is a critical regulator of gene expression [78]. Genome-wide assessment of glioblastomas identified that a subset of tumors harbor a distinctive Glioma CpG Island Methylator Phenotype (G-CIMP), which reflected genome-wide patterns of promoter hypermethylation [79] [23] [80]. This turned out to be a seminal finding, eventually establishing that *IDH1* mutations are causative of the G-CIMP phenotype through the production of the oncometabolite 2-hydroxyglutarate and its inhibitory effect on the DNA demethylating enzyme tet Methylcytosine Dioxygenase 1 (*TET1*) [81-84]. Methylation has since been used to identify tumor subtypes beyond the G-CIMP divide, identifying a number of glioblastoma subtypes in both adults and children [85, 86]. Nevertheless, whether DNA methylation can be used to classify the spectrum of adult diffuse gliomas was unknown at the onset of this PhD-thesis and is discussed in **Chapter 2**. Although tumor heterogeneity is pervasive in gene expression analysis, the impact on DNA methylation-based subtyping was unclear at the start of this thesis and is discussed in more detail in **Chapter 4**.

Telomere maintenance in tumorigenesis

The ends of linear eukaryotic chromosomal DNA consist of several kilobases of repetitive “telomeric” sequence [87]. Due to the directional limitations of DNA polymerase, these ends shorten with every cell division [88]. In scenarios consistent with significant increases in cellular turnover, as in cancer, these telomeric repeat sequences are in danger of diminishing. One of the primary functions of telomeric sequences is binding the protective shelterin protein complex that prevent chromosome ends from being recognized as a DNA double strand breaks [89]. Diminished telomeric sequences inadvertently activate DNA double strand break repair mechanisms and cycles of genomic instability ensue [90].

Extensive genomic instability is counterproductive for tumor growth and cancers have broadly adapted mechanisms for maintaining telomeres at a viable length to preserve genome integrity. Two primary mechanisms have been extensively described. The first mechanism aberrantly activates the somatically silent reverse transcriptase Telomerase Reverse Transcriptase (*TERT*), that is able to synthesize telomeric repeats on-demand [91]. The other, less common, alternative lengthening of telomeres (ALT) mechanism is thought to utilize recombination and sister-chromatid exchange to transiently alleviate diminished telomeres as-needed [92].

We first noticed an association between telomere maintenance and glioma subtype in **Chapter 2**. In **Chapter 6** we comprehensively analyze telomere length

and maintenance mechanisms across an unprecedented number of tumor samples and cancer types. In **Chapter 7** we reflect back on what we have learned in previous chapters and speculate on the biological processes that drive gliomagenesis.

Concluding remarks

The integration of molecular features with neuropathological assessment has yielded invaluable insights into the underpinnings of diffuse gliomas. Past studies have helped to identify glioma-initiating events, revealed potentially targetable oncogenic proteins, defined co-occurrences of molecular alterations in specific tumor subgroups, demonstrated that epigenetic inactivation of genes is predictive of therapeutic response, and that there exists a dynamic molecular landscape in glioma progression. The resulting molecular evidence has enabled the further refinement of tumor subgroups that would have previously been indiscernible using histopathological assessment alone.

The ability to separate tumors based on molecular alterations possesses clear clinical relevance. The unsupervised learning approaches applied to sequencing data do not suffer from the inter-observer variability that can hinder histologically defined diagnostic entities. Indeed, diagnostic entities that were once histologically ambiguous can now often be accurately classified into molecular subgroups that display different clinical behavior. Tumor classification improvements can be made even to well-understood diagnostic entities such IDH-mutant tumors. For example, combining multiple molecular data, such as mutations and DNA methylation, further parsed IDH-mutant tumors into subgroups with unique clinical and molecular characteristics. As sequencing costs drop and more tumors are comprehensively profiled, the sensitivity to detect new molecular subtypes increases. In response to the rapid changes to tumor classifications the Consortium to Inform Molecular and Practical Approaches to CNS Tumor Taxonomy (cIMPACT-NOW) has been launched to evaluate the value of molecular markers for CNS tumor classification with greater frequency than the current WHO timeline allows for (**Figure 1-2**) [93]. Efforts utilizing DNA methylation profiling to characterize all CNS tumors are at the forefront of innovation in molecular neuropathology [94]. Molecular markers are additionally gaining traction as measures of tumor grade [95]. The forthcoming 5th edition of the WHO guidelines closely follows recommendations from cIMPACT-NOW and is scheduled for publication in 2021 [96].

This thesis comprehensively details the "molecular anatomy" of diffuse gliomas to our best knowledge. Nevertheless, a number of critical questions remain. For example, there is no question that glioma is a heterogeneous disease. However, the extent to which this heterogeneity reflects the normal variability between highly proliferative and often DNA repair deficient tumor cells or the pathogenic dedifferentiation and progression of the disease remains to be understood. Moreover, while we can identify differences in histological cell types and molecular subtypes it is still unclear whether these differences implicate distinct a cell-of-origin or simply a unique developmental trajectory. In **Chapter 7** we speculated on the origins and molecular changes that underlie gliomagenesis starting from the first aberrant cell. Future advancements in glioma efforts need to bring findings from large molecular profiling studies back to the basic science laboratory to further our mechanistic understanding of the life history of disease. Such endeavors necessitate a bridge between bench, bed and computer. Finally, it is imperative to keep the end in mind, which is the notion that new findings should drive advancements in cancer therapy. It is critical to remain vigilant in identifying putative therapeutic targets in order to realize the promise of precision medicine.

CHAPTER 2

Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma

Published in print

Ceccarelli M, **Barthel FP** (co-first author), Malta TM, Sabedot TS, TCGA Research Network, Noushmehr H, Iavarone A, Verhaak RG. *Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma.* Cell. 2016 Jan 28;164(3):550-63.

Abstract

Therapy development for adult diffuse glioma is hindered by incomplete knowledge of somatic glioma driving alterations and suboptimal disease classification. We defined the complete set of genes associated with 1,122 diffuse grade II-III-IV gliomas from The Cancer Genome Atlas and used molecular profiles to improve disease classification, identify molecular correlations, and provide insights into the progression from low- to high-grade disease. Whole genome sequencing data analysis determined that *ATRX* but not *TERT* promoter mutations are associated with increased telomere length. Recent advances in glioma classification based on *IDH* mutation and 1p/19q co-deletion status were recapitulated through analysis of DNA methylation profiles, which identified clinically relevant molecular subsets. A subtype of IDH-mutant glioma was associated with DNA demethylation and poor outcome; a group of IDH-wildtype diffuse glioma showed molecular similarity to pilocytic astrocytoma and relatively favorable survival. Understanding of cohesive disease groups may aid improved clinical outcomes.

Introduction

Diffuse gliomas represent 80% of malignant brain tumors [97]. Adult diffuse gliomas are classified and graded according to histological criteria (oligodendroglioma, oligoastrocytoma, astrocytoma and glioblastoma; grade II to IV). Although histopathologic classification is well established and is the basis of the WHO classification of CNS tumors [39], it suffers from high intra- and inter-observer variability, particularly amongst grade II-III tumors [4]. Recent molecular characterization studies have benefited from the availability of the datasets generated by The Cancer Genome Atlas (TCGA) [12, 24, 73, 98-100], and have related genetic, gene expression and DNA methylation signatures with prognosis [22, 23, 86]. For example, mutations in the isocitrate dehydrogenase genes 1 and 2 (*IDH1*/*IDH2*) define a distinct subset of glioblastoma (GBM) with a hypermethylation phenotype (G-CIMP) with favorable outcome [16, 23]. Conversely, the absence of *IDH* mutations in LGG marks a distinct IDH-wildtype subgroup characterized by poor, GBM-like prognosis [12, 100]. Recent work by us and others has proposed classification of glioma into *IDH* wildtype cases, *IDH* mutant samples (IDH-mutant-codel), and an *IDH* mutant group additionally carrying codeletion of chromosome arm 1p and 19q and samples with euploid 1p/19q (IDH-mutant-non-codel), regardless of grade and histology [12, 100]. Mutation of the *TERT* promoter, which has been reported with high frequency across glioma, may be an additional defining feature. Current analyses have not yet clarified the relationships between LGGs and GBMs that share common genetic hallmarks like *IDH* mutation or *TERT* promoter mutation status. An improved understanding of these relationships will be necessary as we evolve toward an objective genome-based clinical classification.

To address the above issues, we assembled a dataset comprising all TCGA newly diagnosed diffuse glioma consisting of 1,122 patients, and comprehensively analyzed using sequencing and array based molecular profiling approaches. We have addressed crucial technical challenges in analyzing this comprehensive dataset, including the integration of multiple platforms and data sources (e.g. multiple methylation and gene expression platforms). We identified new diffuse glioma subgroups with distinct molecular and clinical features and shed light on the mechanisms driving progression of LGG (WHO grades II and III) into full-blown GBM (WHO grade IV).

Results

Patient cohort characteristics

The TCGA LGG and GBM cohorts consist of 516 and 606 patients, respectively. Independent analysis of the GBM dataset was previously described, as was analysis of 290 LGG samples [24, 100]. Two-hundred and twenty-six LGG samples were added to our current cohort (**Table 2-1**). Clinical data including age, tumor grade, tumor histology and survival was available for 93% (1046/1122) of cases (***Table S1****, available online with the full-text of this article at www.cell.com*). The majority of samples were grade IV tumors (n=590, 56%), while 216 (21%) and 241 (23%) were grade II and III tumors, respectively. Similarly, 590 (56%) samples were classified as GBM, 174 (17%) as oligodendroglioma, 169 (16%) as astrocytoma and 114 (11%) as oligoastrocytoma.

Amongst the data sources considered in our analysis were gene expression (n = 1,045), DNA copy number (n = 1,084), DNA methylation (n = 932), exome sequencing (n = 820) and protein expression (n = 473). Multiple and overlapping characterization assays were employed. All data files that were used in our analysis can be found at *https://tcga-data.nci.nih.gov/docs/publications/lgggbm_2015/*.

Table 2-1. Cohort demographics grouped by IDH and 1p/19q codeletion status.

Feature	**IDH wt** (*n* = 520)	**IDH mut non-codel** (*n* = 283)	**IDH mut codel** (*n* = 171)	**Unknown** (*n* = 148)
Clinical				
Histology (n)				
Astrocytoma	52 (10.0%)	112 (39.6%)	4 (2.3%)	1 (0.7%)
Glioblastoma	419 (80.6%)	32 (11.3%)	2 (1.2%)	137 (92.6%)
Oligoastrocytoma	15 (2.9%)	69 (24.4%)	30 (17.5%)	0 (0%)
Oligodendroglioma	19 (3.7%)	37 (13.1%)	117 (68.4%)	1 (0.7%)
Unknown	15 (2.9%)	33 (11.7%)	18 (10.5%)	9 (6.1%)
Grade (n)				
G2	19 (3.7%)	114 (40.3%)	81 (47.4%)	2 (1.4%)
G3	67 (12.9%)	104 (36.7%)	70 (40.9%)	0 (0%)
G4	419 (80.6%)	32 (11.3%)	2 (1.2%)	137 (92.6%)
Unknown	15 (2.9%)	33 (11.7%)	18 (10.5%)	9 (6.1%)

Table 2-1. Cohort demographics grouped by IDH and 1p/19q codeletion status.
Continued

Feature	IDH wt (*n* = 520)	IDH mut non-codel (*n* = 283)	IDH mut codel (*n* = 171)	Unknown (*n* = 148)
Age				
Median (LQ-UQ)	59 (51-68)	38 (30-44)	46 (35-54)	55 (48-68)
Unknown (n)	16	33	18	9
Survival				
Median (CI)	14.0 (12.6-15.3)	75.1 (62.1-94.5)	115.8 (90.5-Inf)	12.6 (11.3-14.9)
Unknown (n)	14	32	18	12
KPS				
<70	85 (16.3%)	8 (2.8%)	5 (2.9%)	21 (14.2%)
70-80	196 (37.7%)	41 (14.5%)	18 (10.5%)	60 (40.5%)
90	29 (5.6%)	60 (21.2%)	32 (18.7%)	2 (1.4%)
100	51 (9.8%)	44 (15.9%)	30 (17.5%)	14 (9.5%)
Unknown	159 (30.6%)	129 (45.6%)	86 (50.3%)	51 (34.5%)
Molecular				
MGMT promoter				
Methylated	170 (32.7%)	242 (85.5%)	169 (98.8%)	32 (21.6%)
Unmethylated	248 (47.7%)	36 (12.7%)	1 (0.6%)	34 (23.0%)
Unknown	102 (19.6%)	5 (1.8%)	1 (0.6%)	82 (55.4%)
TERT promoter				
Mutant	67 (12.9%)	8 (2.8%)	86 (50.3%)	1 (0.7%)
Wild-type	19 (9.8%)	146 (51.6%)	2 (1.2%)	0 (0%)
Unknown	434 (83.5%)	129 (45.6%)	83 (48.5%)	135 (99.3%)
TERT expression				
Expressed	178 (34.2%)	14 (4.9%)	153 (89.5%)	6 (4.1%)
Not expressed	51 (9.8%)	242 (85.5%)	16 (9.4%)	7 (4.7%)
Unknown	291 (56.0%)	27 (9.5%)	2 (1.2%)	135 (91.2%)

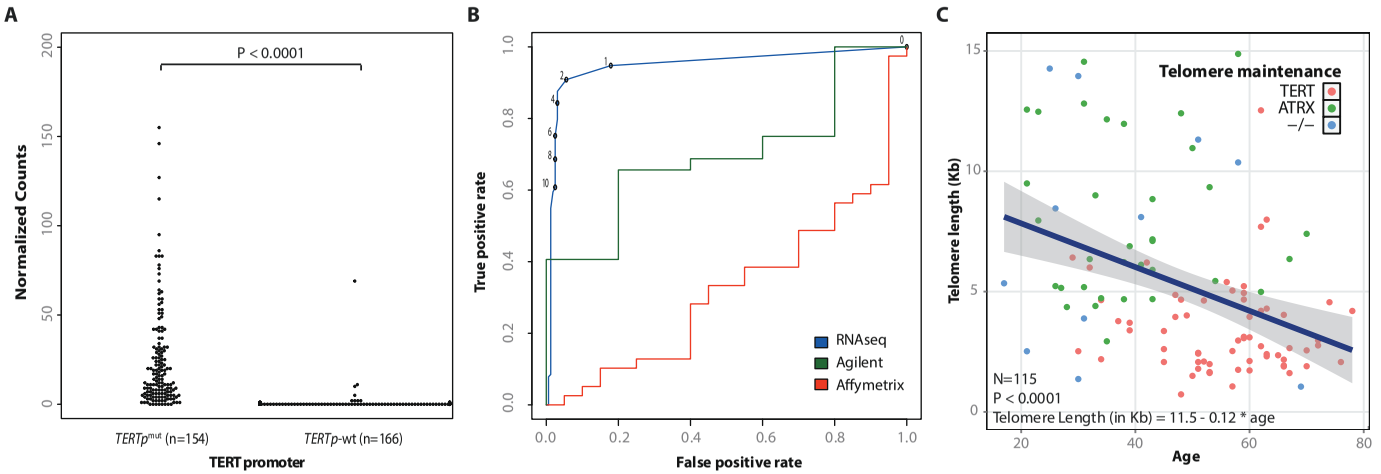
Identification of novel glioma-associated genomic alterations

To establish the set of genomic alterations that drive gliomagenesis, we called point mutations and indels on the exomes of 513 LGG and 307 GBM using the Mutect, Indelocator, Varscan2 and RADIA algorithms, and considered all mutations identified by at least two callers. Significantly mutated genes (SMGs) were determined using MutSigCV. This led to the identification of seventy-five SMGs,

of which ten had been previously reported in GBM [24], twelve had been reported in LGG [100], and eight had been identified in both GBM and LGG studies. Forty-five SMGs have not been previously associated with glioma and ranged in mutation frequency from 0.5% to 2.6% (***Table S2A**, available online with the full-text of this article at www.cell.com*). We used GISTIC2 to analyze the DNA copy number profiles of 1,084 samples, including 513 LGG and 571 GBM, and identified 162 significantly altered DNA copy number segments (***Table S2B**, available online with the full-text of this article at www.cell.com*). We employed PRADA and deFuse to detect 1,144 gene fusion events in the RNA-seq profiles available for 154 GBM and 513 LGG samples, of which 37 in-frame fusions involved receptor tyrosine kinases (***Table S2C**, available online with the full-text of this article at www.cell.com*). Collectively, these analyses recovered all known glioma driving events, including in *IDH1* (*n* = 457), *TP53* (*n* = 328), *ATRX* (*n* = 220), *EGFR* (*n* = 314), *PTEN* (*n* = 168), *CIC* (*n* = 80), *FUBP1* (*n* = 45). Notable newly predicted glioma drivers relative to the earlier TCGA analyses were genes associated with chromatin organization such as *SETD2* (*n* = 24), *ARID2* (*n* = 20), *DNMT3A (n* = 11), and the *KRAS/NRAS* oncogenes (*n* = 25 and *n* = 5, respectively).

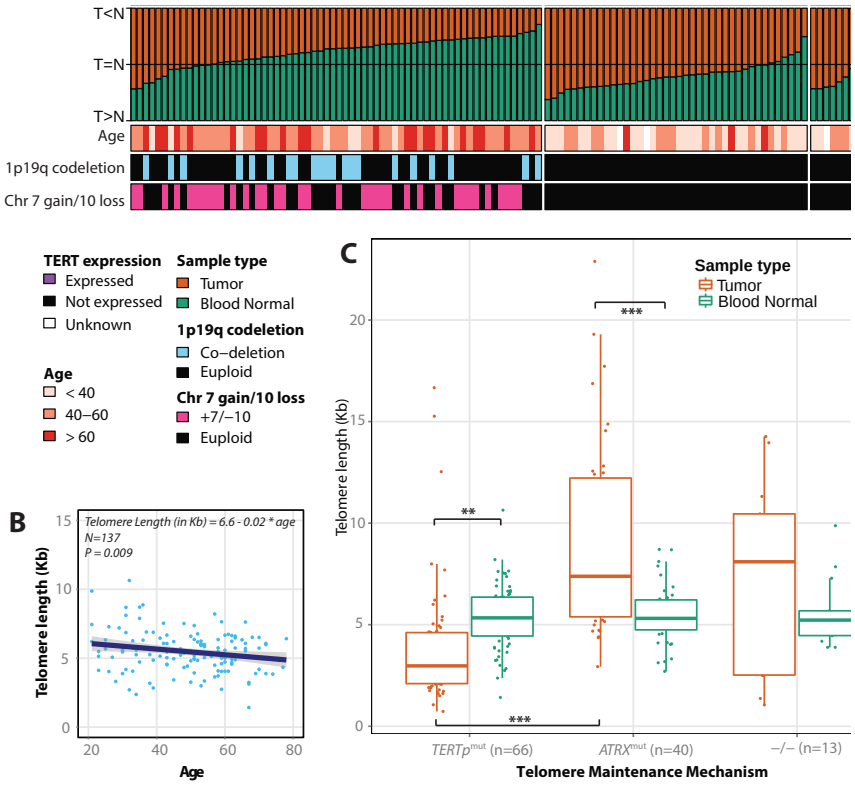
We overlapped copy number, mutation (*n* = 793) and fusion transcript (*n* = 649) profiles and confirmed the convergence of genetic drivers of glioma into pathways including the Ras-Raf-MEK-ERK, p53/apoptosis, PI3K/AKT/mTOR, chromatin modification and cell cycle pathways. The Ras-Raf-MEK-ERK signaling cascade showed alterations in 106 of 119 members detected across 578 cases (73%), mostly occurring in IDH-wildtype samples (*n* = 327 of 357, 92%). Conversely, we found that a set of 36 genes involved in chromatin modification was targeted by genetic alterations in 423 tumors (54%, *n* = 36 genes), most of which belonged to the IDH-mutant-non-codel group (n = 230, 87%).

In order to identify new somatically altered glioma genes, we used MutComFocal to nominate candidates altered by mutation as well as copy number alteration. Prominent among these genes was *NIPBL,* a crucial adherin subunit that is essential for loading cohesins on chromatin (***Table S2D**, available online with the full-text of this article at www.cell.com*) [101]. The cohesin complex is responsible for the adhesion of sister chromatids following DNA replication and is essential to prevent premature chromatid separation and faithful chromosome segregation during mitosis [101]. Alterations in the cohesin pathway have been reported in 12% of acute myeloid leukemias [102]. Mutations of the cohesin complex gene *STAG2* had been previously reported in GBM [24]. Taken together, 16% of the LGG/GBM showed mutations and/or CNAs in multiple genes involved in the cohesin complex, thus nominating this process as a prominent pathway involved in gliomagenesis.



↑ **Figure 2-1. Telomere length quantification of 120 gliomas with known *TERT* promoter and *ATRX* mutational status.**

A. RNAseq *TERT* expression is upregulated in *TERTp* mutant cases, but not in ATRX and double negative cases (P<0.0001). **B.** *TERT* expression as quantified by RNA sequencing is a highly sensitive and specific marker for the absence or presence of the *TERTp* mutation (AUC 0.95). Using a cutoff value of 2, sensitivity and specificity are 91% and 95%, respectively. Interestingly, microarray data is poor substitute for the *TERT* promoter mutation with an AUC of 0.70 and 0.32 for the Agilent and Affymetrix microarray respectively. **C.** Telomeres gradually shorten with increasing age in tumor samples (P<0.0001). Note the steeper decline relative to **Figure 2-1**b and that *ATRX* mutant patients are in the younger age range whereas *TERTp* mutant patients are in the older age range. This suggests an independent contribution of telomere maintenance to telomere length.



↑ **Figure 2-2. Telomere length associations in glioma.**

A. Heatmap of relative tumor/normal telomere lengths of 119 gliomas, grouped by *TERTp* and *ATRX* mutation status. **B.** Telomere length decreases with increasing age (measured in years at diagnosis) in blood normal control samples (n=137). **C.** Quantitative telomere length estimates of tumors and blood normal, grouped by *TERTp* mutant (n=67, 56%), *ATRX* mutant (n=40, 33%) and double negative (n=13, 11%) status. *** = P<0.0001; ** = P<0.001.

Telomere length is positively correlated with *ATRX* but not *TERT* promoter mutations

Mutations in the *TERT* promoter (*TERTp*) have been reported in 80% of GBM [103]. We used *TERTp* mutation calls from targeted sequencing (*n* = 287) and complemented them with *TERTp* mutations inferred from whole genome sequencing (WGS) data (*n* = 42). *TERTp* mutations are nearly mutually exclusive with mutations in *ATRX* [12], which was confirmed in our cohort. Overall, 85% of diffuse gliomas harbored mutations of *TERTp* (*n* = 157, 48%) or *ATRX* (*n* = 120, 37%). *TERTp* mutations activate *TERT* mRNA expression through the creation of a *de novo* ETS transcription factor-binding site [104] and we observed significant *TERT* upreg-

ulation in *TERTp* mutant cases (p-value < 0.0001, **Figure 2-1a**). *TERT* expression measured by RNA-seq was a highly sensitive (91%) and specific (95%) surrogate for the presence of *TERTp* mutation (**Figure 2-1b**). We correlated *TERTp* status with glioma driving alterations and observed that nearly all IDH-wildtype cases with chromosome 7 gain and chromosome 10 loss harbored *TERTp* mutations or upregulated *TERT* expression (*n* = 52/53 and *n* = 134/147, respectively; **Figure 2-2a**). Conversely, only 45% of IDH-wildtype samples lacking chromosome 7/chromosome 10 events showed *TERTp* mutations or elevated *TERT* expression (n=15/33 and n = 43/82, respectively). Thus, *TERTp* mutations may precede the chr 7/chr 10 alterations which have been implicated in glioma initiation [105].

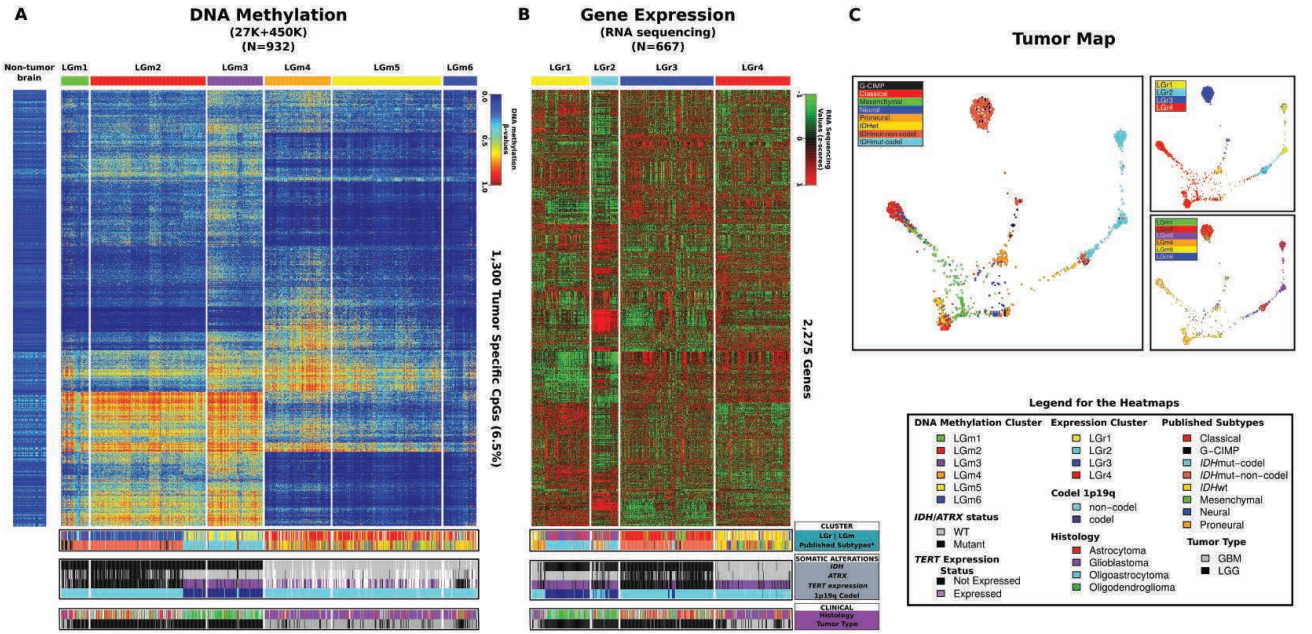
To correlate *TERTp* mutations to telomere length, we used whole genome sequencing and low pass whole genome sequencing data to estimate telomere length in 141 pairs of matched tumor and normal samples. As expected, we observed an inverse correlation of telomere length with age at diagnosis in matching blood normal samples (**Figure 2-2b**) and tumor samples (**Figure 2-1c**). Glioma samples harboring ATRX mutations showed significantly longer telomeres compared to *TERTp* mutant samples (t-test p-value < 0.0001; **Figure 2-2c**). Among *TERTp* mutation gliomas, there was no difference in telomere length between samples with and without additional *IDH1/IDH2* mutations, despite a difference in age. ATRX forms a complex with DAXX and H3.3, and the genes encoding these proteins are frequently mutated in pediatric gliomas [86]. Mutations in DAXX and H3F3A were identified in only two samples in our WGS dataset. The ATRX-DAXX-H3.3 complex is associated with the alternative lengthening of telomeres (ALT) and our observations confirm previously hypothesized fundamental differences between the telomere control exerted by telomerase and ALT [106].

As demonstrated by the identification of *TERTp* mutations, somatic variants affecting regulatory regions may play a role in gliomagenesis. Using 67 matched whole-genome and RNA-seq expression pairs, we similarly sought to identify mutations located within 2kb upstream of transcription start sites and associated with a gene expression change. Using strict filtering methods, we identified twelve promoter regions with mutations in at least six samples. Three of twelve regions related to a significant difference in the expression of the associated gene expression, suggesting possible functional consequences. Other than *TERT* (*n* = 37), promoter mutations of the ubiquitin ligase *TRIM28* (*n* = 8) and the calcium channel gamma subunit *CACNG6* (*n* = 7) correlated with respectively upregulation and downregulation of these genes, respectively (***Table S2E**, available online with the full-text of this article at www.cell.com*). TRIM28 has been reported to mediate

the ubiquitin-dependent degradation of AMP-activated protein kinase (AMPK) leading to activation of mTOR signaling and hypersensitization to AMPK agonists, such as metformin [107].

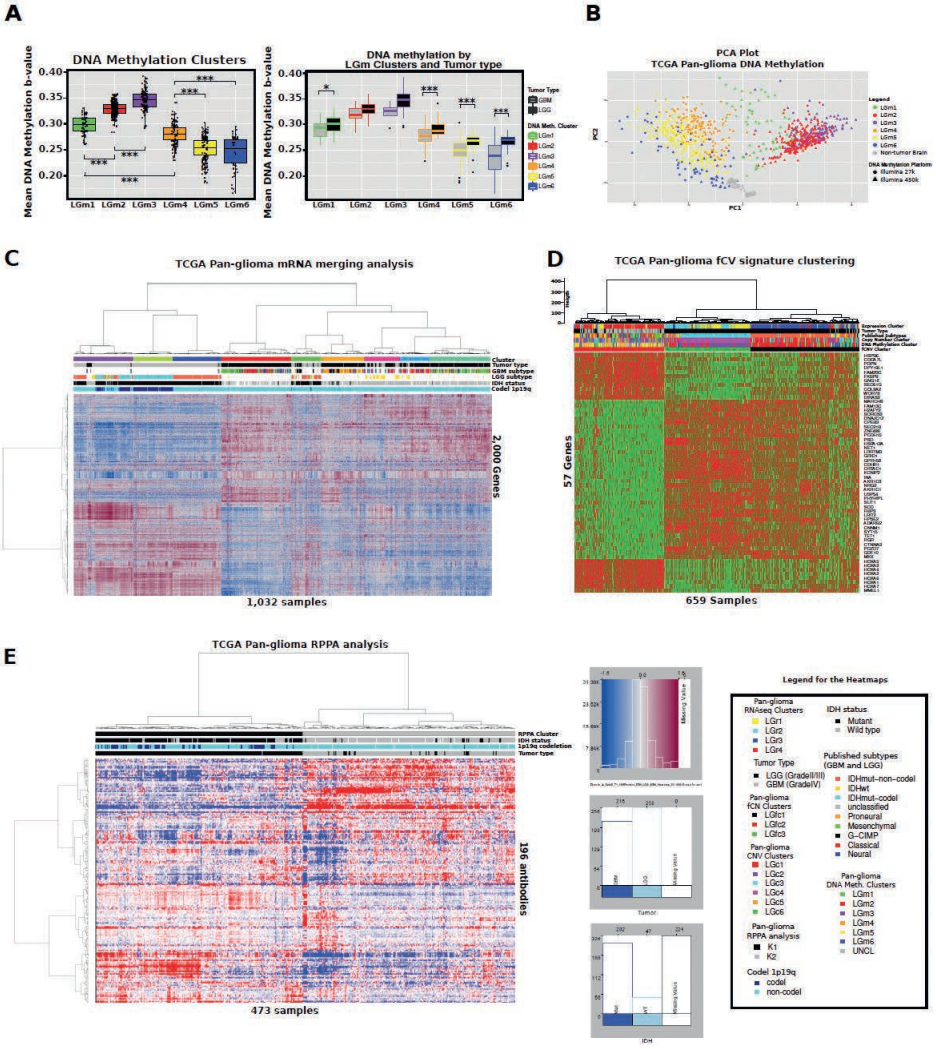
Unsupervised clustering of gliomas identifies six methylation groups and four RNA expression groups associated with IDH status

To segregate the DNA methylation subtypes across the pan-glioma dataset, we analyzed 932 glioma samples profiled on the HumanMethylation450 platform (516 LGG and 142 GBM) and the HumanMethylation27 platform (287 GBM). In order to incorporate the maximum number of samples, we merged datasets from both methylation platforms yielding a core set of 25,978 CpG probes. To reduce computational requirements to cluster this large dataset, we eliminated sites that were methylated (mean β-value ≥ 0.3) in non-tumor brain tissues and selected 1,300 tumor specific methylated probes (1,300/25,978, 5%) to perform unsupervised k-means consensus clustering. This identified six distinct clusters, labeled LGm1-6 (**Figure 2-3a**; ***Table S1****, available online with the full-text of this article at www.cell.com*; ***Table S3A****, available online with the full-text of this article at www.cell.com*). Next, we sought to determine pan-glioma expression subtypes through unsupervised clustering analysis of 667 RNA-seq profiles (513 LGG and 154 GBM) which resulted in four main clusters labeled LGr1-4 (**Figure 2-3**b; ***Table S1****, available online with the full-text of this article at www.cell.com*; ***Table S3A****, available online with the full-text of this article at www.cell.com*). An additional 378 GBM samples with Affymetrix HT-HG-U133A profiles (but lacking RNA-seq data) were classified into the four clusters using a k-nearest neighbor classification procedure. *IDH* mutation status was the primary driver of methylome and transcriptome clustering and separated the cohort into two macro-groups. The LGm1/LGm2/LGm3 DNA methylation macro-group carried *IDH1* or *IDH2* mutations (446 of 448, 99%) and was enriched for LGG (421/454, 93%) while LGm4/LGm5/LGm6 were *IDH*-wildtype (414/415, 99%) and enriched for GBM (383/478, 80%). LGm1-3 showed genome wide hypermethylation compared to LGm4-6 clusters (**Figure 2-4a**), documenting the association between *IDH* mutation and increased DNA methylation [23, 84]. Principal component analysis using 19,520 probes yielded similar results, thus emphasizing that our probe selection method did not introduce unwanted bias (**Figure 2-4b**). The gene expression clusters LGr1-3 harbored *IDH1* or *IDH2* mutations (437 of 533, 82%) and were enriched for LGG (436/563, 77%) while the LGr4 was exclusively IDH-wildtype (376 of 387, 97%) and enriched for GBM (399/476, 84%).



↑ **Figure 2-3. Pan-glioma DNA methylation and transcriptome subtypes.**

A. Heatmap of DNA methylation data. Columns represent 932 TCGA glioma samples grouped according to unsupervised cluster analysis, rows represent DNA methylation probes sorted by hierarchical clustering. Non-neoplastic samples are represented on the left of the heatmap (n = 110). **B.** Heatmap of RNA sequencing data. Unsupervised clustering analysis for 667 TCGA glioma samples profiled using RNA sequencing are plotted in the heatmap using 2,275 most variant genes. Previously published subtypes were derived from Brennan *et al.* Cell, 2013 and TCGA Research Network, NEJM, 2015. **C.** Tumor Map based on mRNA expression and DNA methylation data. Each data point is a TCGA sample colored coded according to their identified status. A live interactive version of this map is available at http://tumormap.ucsc.edu/?p=ynewton.gliomas-paper.



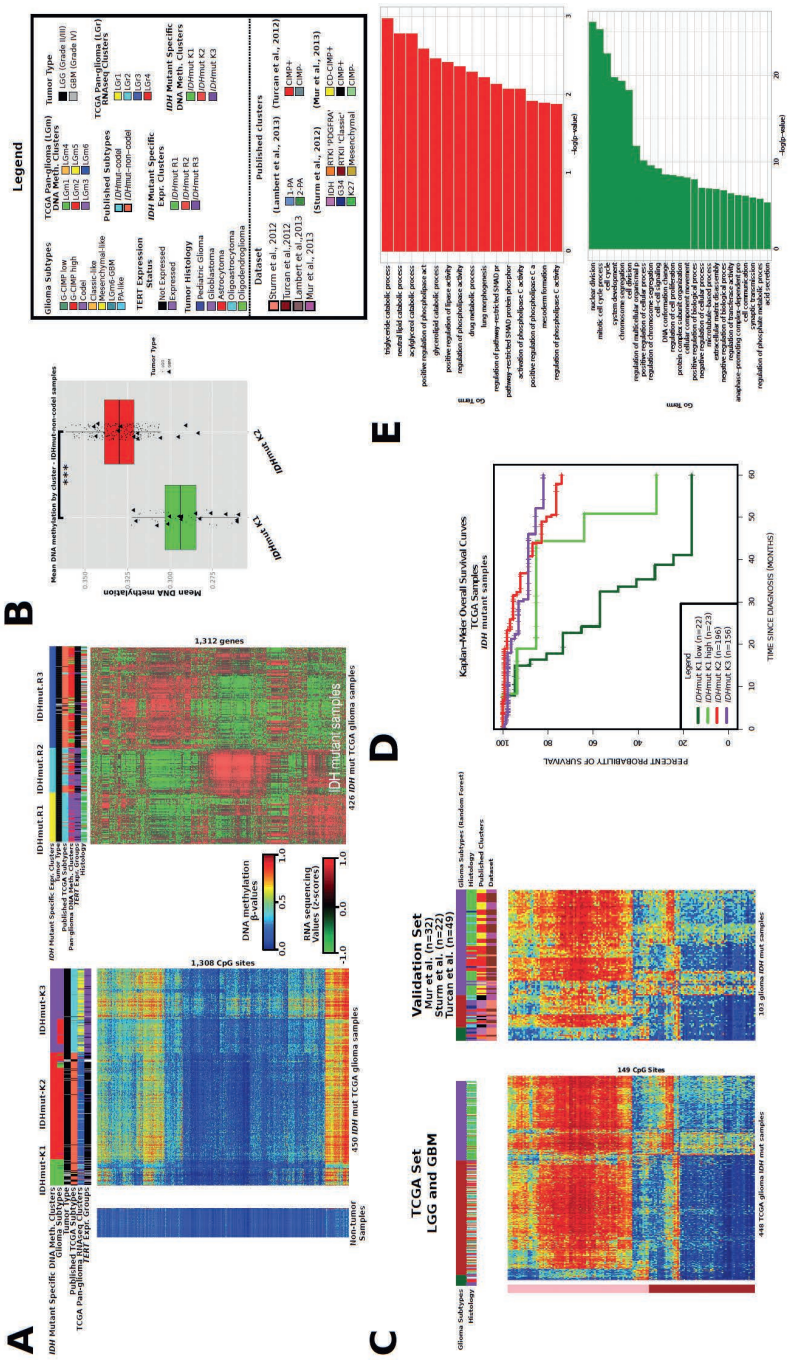
↑ **Figure 2-4. Pan-glioma DNA methylation and transcriptome subtypes.**

A. Boxplot of the mean DNA methylation beta-values genome-wide (20,036 probes) for each sample distributed by the six Pan-glioma DNA methylation clusters (left) and divided by tumor type (right). Significant differences are highlighted with * (p-value < 0.01) and *** (p-value < 1e-04). **B.** Principal component analysis of 932 TCGA glioma samples and 77 non-tumor brain samples (Guintivano et al., 2013) performed on 19,520 CpG probes (genome-wide). **C.** LGG-GBM mRNA merging analysis. Clustered heatmap of merged data with 569 GBM and 463 LGG non-duplicate samples, and 2000 most variable genes. Consensus clustering revealed 9 clusters. The 3 left-most clusters show predominantly LGG samples, 3 clusters show predominantly GBM samples, whereas 3 clusters show a mixture of GBM and LGG samples. The LGG IDH-wildtype samples clustered mostly with the GBM classical samples, whereas many of the LGG IDH-mutant-non-

codel samples cluster with the GBM G-CIMP samples. **D.** Functional Copy Number (fCN) gene signature Heatmap. Genes with Spearman's correlation between CN and Expression above 0.5, abs (FC>1.5) and abs(ΔCN>0.5) define the fCN signature. The Heatmap illustrate the samples unsupervised clustering given the fCN signature. RNA expression levels range from green (low) to red (high). Each row reports the annotation of a different analysis performed in the paper. Last row reports the fCN annotation. **E.** Clustered heatmap of unsupervised hierarchical clustering of 473 samples (columns) and 196 antibodies (rows). The annotation bars (shown on top) were not used for clustering. The legend for the annotation bars is shown on the left. Two clusters can be found that largely correspond to tumor type. The left cluster has largely LGG samples and one GBM sample. However, the right cluster has mostly GBM samples but 26 LGG samples, 17 of which have no mutations in *IDH*1/2. In the heatmap, low, medium, and high expression is represented by blue, white, and red colours, respectively.

We extended our analysis using Tumor Map (Supplementary Methods) to perform integrated co-clustering analysis of, the combined gene expression (*n* = 1,196) and DNA methylation (*n* = 867) profiles. An interactive Tumor Map version is publicly available at http://tumormap.ucsc.edu/?p=ynewton.gliomas-paper. Tumor Map assigns samples to a hexagon in a grid so that nearby samples are likely to have similar genomic profiles and allows visualizing complex relationships between heterogeneous genomic data samples and their clinical or phenotypical associations. Thus, clusters in the map indicate groups of samples with high similarity of integrated gene expression and DNA methylation profiles (**Figure 2-3c**). The map confirms clustering by IDH status and additionally shows islands of samples that share previously reported GBM cluster memberships [22, 23]. To assess clustering sensitivity to pre-processing we tried complementary methods and obtained similar results (**Figure 2-4c**).

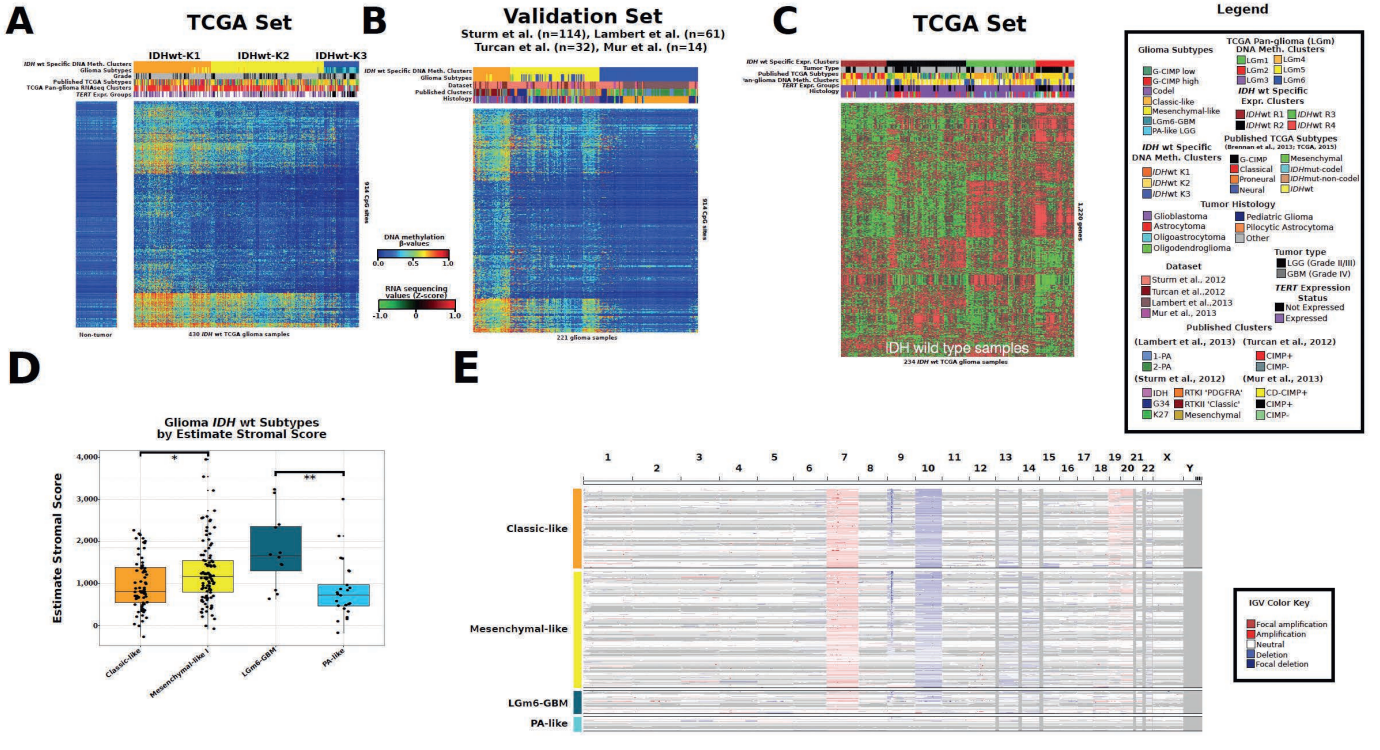
To identify genes whose copy number changes are associated with concordant changes in gene expression, we combined expression and copy number profiles from 659 samples to define a signature of 57 genes with strong functional copy number (fCN) change (***Table S3B****, available online with the full-text of this article at www.cell.com*). The fCN signature clustered gliomas into three macro-clusters, LGfc1-3, strongly associated with IDH and 1p/19q status (**Figure 2-4d**). The fCN analysis revealed the functional activation of a cluster of HOXA genes in the IDH-wildtype LGfc2 cluster, which were previously associated with glioma stem cell maintenance [108].


A
B
C
D
E
Legend
TCGA Set
LGG and GBM
Validation Set
Mur et al. (n=32)
Sturm et al. (n=22)
Turcan et al. (n=49)
1,312 genes
1,308 CpG sites
149 CpG Sites
450 IDH mut TCGA glioma samples
426 IDH mut TCGA glioma samples
448 TCGA glioma IDH mut samples
103 glioma IDH mut samples
Non-tumor Samples
IDH mutant samples
Kaplan-Meier Overall Survival Curves
TCGA Samples
IDH mutant samples
PERCENT PROBABILITY OF SURVIVAL
TIME SINCE DIAGNOSIS (MONTHS)
IDHmut K1 low (n=22)
IDHmut K1 high (n=23)
IDHmut K2 (n=93)
IDHmut K3 (n=156)
Mean DNA methylation by cluster - IDHmut-non-codel samples
IDHmut K1
IDHmut K2
IDHmut K3
Go Term
-log(p-value)

← **Figure 2-5. Identification of a distinct IDH-mutant subtype defined by epigenomics.**

A. Left, Heatmap of DNA methylation data. Unsupervised consensus clustering analysis using 1,308 CpG tumor specific CpG probes defined among the TCGA IDH-mutant gliomas. Column-wise represents 450 IDH-mutant glioma samples, row-wise represents probes. Samples are ordered according to the consensus cluster output, and rows are ordered by hierarchical clustering. DNA methylation beta-values ranges from 0 (low) to 1 (high). Three clusters were defined, each cluster separated and labeled. Non-tumor brain samples are represented on the left of the heatmap (Guintivano et al., 2013). Additional tracks are included at the top of the heatmaps to identify each sample membership within separate cluster analysis (Glioma subtypes, tumor type, previous published subtypes (Brennan *et al.* Cell, 2013, TCGA Research Network, NEJM, 2015), RNA sequencing and *TERT* expression). Legend is provided for the heatmap. Right, Clustering of *IDH*-mutant samples transcriptional profiles. Unsupervised clustering of gene expression separated by *IDH* status 426 samples confirming the presence of three main groups resembling the clusters reported in (TCGA Network, New Eng J Med 2015) where all GBM G-CIMP cluster together with the LGG IDH-mutant-non-codel. **B.** Boxplot of the average DNA methylation beta-value genome-wide (20,000 probes) for each sample grouped by *IDH*mut K1 and *IDH*mut K2. Dots represent LGG tumors and triangles represent GBM tumors. Significant difference is highlighted with *** (p-value < 2.2×10^{-16}) **C.** Left, Heatmap of DNA methylation data. Supervised statistical analysis using 149 CpG tumor specific CpG probes that define each TCGA IDH-mutant glioma subtype. Column-wise represents 448 IDH-mutant (codels and non codels) TCGA glioma samples, row-wise represents probes. DNA methylation beta-values ranges from 0 (low) to 1 (high). Right, Heatmap of DNA methylation data for the validation data set (Sturm et al., 2012; Turcan et al., 2012; Mur et al., 2013), using the 149 CpG tumor specific probes that define each TCGA *IDH-mutant* glioma subtype. Non-TCGA glioma samples were classified into one of the three IDH-mutant type specific clusters using the random forest machine learning method. DNA methylation beta-values ranges from 0 (low) to 1 (high). Additional tracks are included at the top of the heatmap to identify tumor histology, published clusters (Published Clusters) and each sample membership according to its dataset (Study). Legend is provided for the heatmap. **D.** Kaplan-Meier survival curves showing samples separated by IDHmut K1 low, IDHmut K1 high, IDHmut K2 and IDHmut K3. Tick represent censorship. **E.** Pathway analysis of differentially expressed genes between IDHmut K1, IDHmut K2, ranked by p-value. In top red panel there are the categories enriched in IDHmutK2 whereas in bottom green panel the categories enriched in IDHmutK2.

Finally, we clustered reverse phase protein array profiles, consisting of 196 antibodies on 473 samples. Two macro clusters were observed and in contrast to the transcriptome/methylome/fCNV clustering, the primary discriminator was based on glioma grade (LGG vs GBM) rather than IDH status (**Figure 2-4e**). Compared to the LGG-like cluster, the GBM-like cluster had elevated expression of IGFBP2, fibronectin, PAI1, HSP70, EGFR, phosphoEGFR, phosphoAKT, Cyclin B1, Caveolin, Collagen VI, Annexin1 and ASNS, whereas the LGG class showed increased activity of PKC (alpha, beta and delta), PTEN, BRAF, and phosphoP70S6K.


A
TCGA Set
IDHwt-K1
IDHwt-K2
IDHwt-K3
B
Validation Set
Sturm et al. (n=114), Lambert et al. (n=61)
Turcan et al. (n=32), Mur et al. (n=14)
C
TCGA Set
IDH wild type samples
Legend
D
Glioma IDH wt Subtypes
by Estimate Stromal Score
Estimate Stromal Score
Classic-like
Mesenchymal-like
LGm6-GBM
PA-like
E
IGV Color Key
Focal amplification
Amplification
Neutral
Deletion
Focal deletion

← Figure 2-6. Identification of a distinct subgroup of IDH-wildtype gliomas with pilocytic astrocytoma features.

A. Heatmap of DNA methylation data. Unsupervised consensus clustering analysis using 914 CpG tumor specific probes defined among the TCGA IDH-wildtype gliomas. Column-wise represents 430 IDH-wildtype TCGA glioma samples, row-wise represents probes. Samples are ordered according to the consensus cluster output, and rows are ordered by hierarchical clustering. DNA methylation beta-values ranges from 0 (low) to 1 (high). Three clusters were defined, each cluster separated and labeled. Non-tumor brain samples are represented on the left of the heatmap (Guintivano et al., 2013). Additional tracks are included at the top of the heatmaps to identify each sample membership within separate cluster analysis (Glioma subtypes, tumor type, previous published subtypes Brennan et al. Cell, 2013, TCGA Research Network, NEJM, 2015), RNA sequencing and TERT expression). Legend is provided for the heatmap. **B.** Heatmap of DNA methylation data for the validation data set (Sturm et al., 2012; Turcan et al., 2012; Lambert et al., 2013; Mur et al., 2013), using the 914 CpG tumor specific probes defined in panel S4A. Non-TCGA glioma samples were classified into one of the three IDH-wildtype specific clusters using the random forest machine learning method. The second track from top to bottom shows the classification of non-TCGA glioma samples into one of the seven glioma subtypes also using the random forest machine learning method. DNA methylation beta-values ranges from 0 (low) to 1 (high). Additional tracks are included at the top of the heatmap to identify each sample membership according to its dataset (Dataset), to previous published clusters (Published Clusters) and to tumor histology. Legend is provided for the heatmap. **C.** Clustering of IDH-wildtype samples transcriptional profiles. Unsupervised clustering of gene expression separated by IDH status showed that the LGr4 cluster identified in the pan-glioma unsupervised analysis splits into four mixed LGG/GBM clusters (234 samples), where the first two, although separated by a relatively small number of genes, are respectively enriched with Classical subtype (59%) and LGm4 samples and the second with Mesenchymal (75%) subtype and LGm5 samples, the third enriched with Proneural subtype (85%) and a fourth mostly containing LGG IDH-wildtype samples. **D.** Boxplot of the estimate stromal score for each sample distributed by the four glioma IDH wildtype subtypes. Significant differences are highlighted with * (p-value < 0.05) and ** (p-value < 0.005). **E.** IGV screenshot demonstrating differences in copy number landscape across glioma subtypes.

The above results confirm IDH status as the major determinant of the molecular footprints of diffuse glioma. To further elucidate the subtypes of diffuse glioma, we performed unsupervised clustering within each of the two IDH-driven macroclusters. We used 1,308 tumor-specific CpG probes defined among the IDH mutation cohort (*n* = 450) and identified three IDH-mutant specific DNA methylation clusters (**Figure 2-5a**). Using 914 tumor specific CpG probes in the IDH-wildtype cohort (*n* = 430), we uncovered three IDH-wildtype specific clusters (**Figure 2-6a**). The sets of CpG probes used to cluster each of the two IDH-driven datasets overlapped significantly with the 1,300 probes that defined the pan-glioma DNA methylation clustering (1162/1,300, 89% and 853/1,300, 66%, for IDH-mutant and

IDH-wildtype, respectively). The clusters identified by separating IDH-mutant and IDH-wildtype gliomas showed strong overall concordance with pan-glioma DNA methylation subtypes (***Table S3A**, available online with the full-text of this article at www.cell.com*). Similarly, unsupervised clustering of 426 IDH-mutant RNA-seq profiles resulted in three subtypes (**Figure 2-5a**) and analysis of the 234 IDH-wildtype samples led to four mixed LGG/GBM clusters that showed enrichment for previously identified GBM expression subtypes (**Figure 2-6c**) [22].

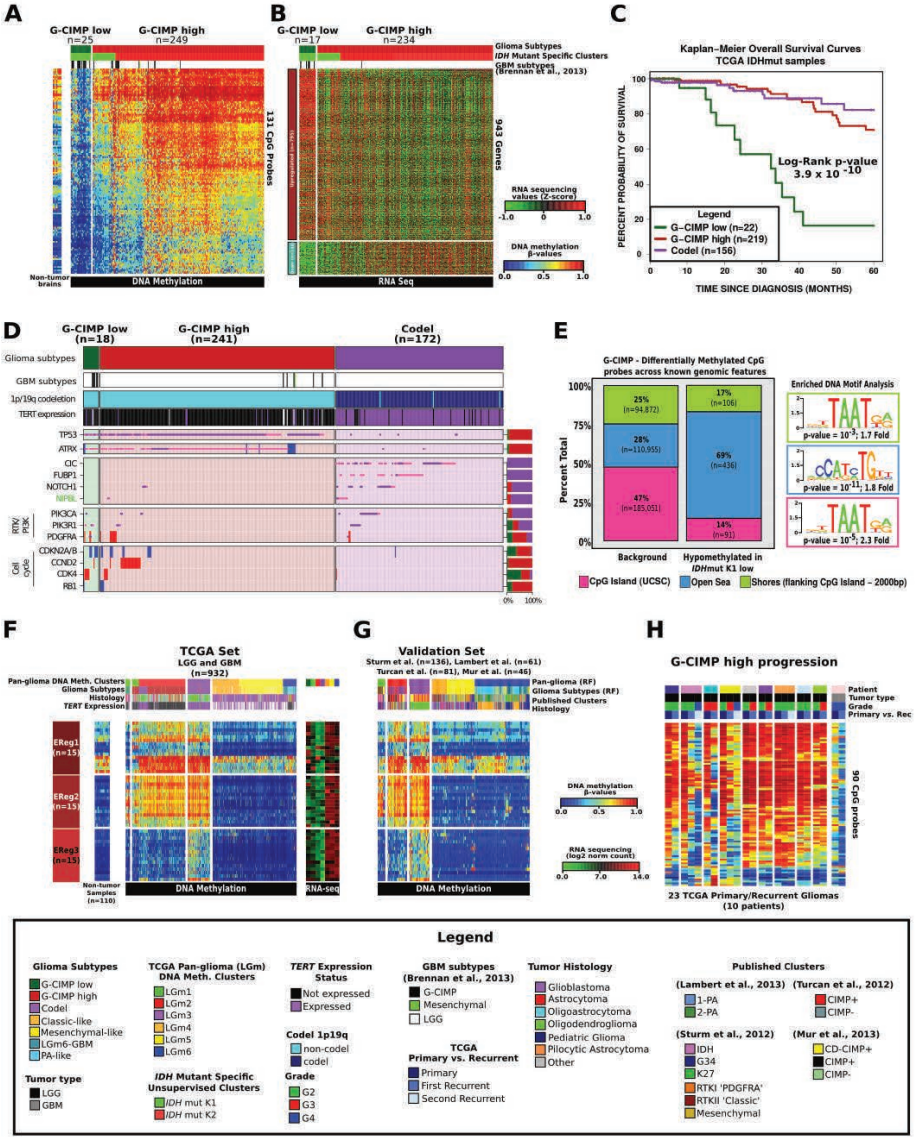
An epigenetic signature associated with activation of cell cycle genes segregates a subgroup of IDH-mutant LGG and GBM with unfavorable clinical outcome

The three epigenetic subtypes defined by clustering IDH-mutant glioma separated samples harboring the 1p/19q co-deletion into a single cluster and non-codel glioma into two clusters (**Figure 2-5a**). Conversely, non-codel glioma grouped nearly exclusively into a single expression cluster and codels were split in two separated expression clusters (**Figure 2-5a**). A distinct subgroup of samples within the IDH-mutant-non-codel DNA methylation clusters manifested relatively reduced DNA methylation (**Figure 2-5b**). The unsupervised clustering of IDH-mutant glioma was unable to segregate the lower methylated non-codel subgroup as the 1,308 probes selected for unsupervised clustering included only 19 of the 131 differentially methylated probes characteristic for this subgroup (FDR < 10^{-15}, difference in mean methylation beta-value > 0.27). The low-methylation subgroup consisted of both G-CIMP GBM (13/25) and LGGs (12/25) and was confirmed using a non-TCGA dataset (**Figure 2-5c**). The tumors with higher methylation in the split cluster were very similar to those grouped in the second non-codel cluster and a supervised comparison identified only 12 probes as differentially DNA methylated (**Figure 2-7a**; **Figure 2-7b**). We concluded that IDH-mutant glioma is composed of three coherent subgroups: 1. The Codel group, consisting of IDH-mutant-codel LGGs; 2. The G-CIMP-low group, including IDH-mutant-non-codel glioma (LGG and GBM) manifesting relatively low genome wide DNA methylation; and 3. The G-CIMP-high group including IDH-mutant-non-codel glioma (LGG and GBM) with higher global levels of DNA methylation. The newly identified G-CIMP-low group of glioma was associated with significantly worse survival as compared to the G-CIMP-high and Codel groups (**Figure 2-5d**). The clinical outcome of the tumors classified as G-CIMP-high was as favorable as that of Codel tumors, the subgroup generally thought to have the best prognosis among glioma patients (**Figure 2-7c**;

Figure 2-5d). We compared the frequencies of glioma driver gene alterations between the three types of IDH-mutant glioma and found that 15 of 18 G-CIMP-low cases carried abnormalities in cell cycle pathway genes such as *CDK4* and *CDKN2A*, relative to 36/241 and 2/172 for G-CIMP-high and Codels, respectively (**Figure 2-7d**). Supervised analysis between gene expression of G-CIMP-low and G-CIMP-high resulted in 943 differentially expressed genes. We mapped the 943 deregulated genes to 767 nearest CpG probes (max distance 1kb) and found the majority of the CpG probes (486/767, 63%) to show a significant methylation difference (FDR<0.05, difference in mean methylation beta-value > 0.01) between G-CIMP-low and G-CIMP-high, suggesting a mechanistic relation between loss of methylation and increased transcript levels.

Recent analysis of epigenetic profiles derived from colon cancers showed that transcription factors may bind to regions of demethylated DNA [109]. Therefore, we asked whether transcription factors may be recruited to the DNA regions differentially methylated between G-CIMP-low samples and G-CIMP-high samples from the same methylation cluster, using 450k methylation profiles (n = 39). Globally, we detected 643 differentially methylated probes between 27 G-CIMP-low and 12 G-CIMP-high samples (absolute diff-mean difference >= 0.25, FDR <= 5%). Most of these probes (69%) were located outside of any known CpG island but positioned within intergenic regions known as open seas (**Figure 2-7e**). This represents a 2.5-fold open sea enrichment compared to the expected genome-wide distribution of 450K CpG probes (Chi-Square p-value < $2.2*10^{-16}$). We also observed a 3.4-fold depletion within CpG islands (Chi-Square p-value < $2.2*10^{-16}$).

Using this set of intergenic CpG probes, we asked whether a DNA motif signature associated with distal regulatory elements. Such a pattern would point to candidate transcription factors involved in tumorigenesis of the G-CIMP-low group. A *de novo* motif scan and known motif scan identified a distinct motif signature –TGTT- (geometric test p-value = 10^{-11}, fold enrichment = 1.8), known to be associated with the OLIG2 and SOX transcription factor families (**Figure 2-7e**) [110]. This observation was corroborated by the higher expression levels of *SOX2* as well as 17 out of 20 other known SOX family members in G-CIMP-low compared to G-CIMP-high (fold difference > 2). The primary function of SOX2 in the nervous system is to promote self-renewal of neural stem cells and, within brain tumors, the glioma stem cell state [111]. Interestingly, SOX2 and OLIG2 have been described as neurodevelopmental transcription factors being essential for GBM propagation [112].



↑ **Figure 2-7. Identification of a distinct G-CIMP subtype defined by epigenomics.**

A. Heatmap of probes differentially methylated between the two IDH-mutant-non-codel DNA methylation clusters allowed the identification of a low-methylation subgroup named G-CIMP-low. Non-tumor brain samples (n=12) are represented on the left of the heatmap. **B**. Heatmap of genes differentially expressed between the two IDH-mutant-non-codel DNA methylation clusters. **C.** Kaplan-Meier survival curves of IDH-mutant methylation subtypes. Ticks represent censored values. **D.** Distribution of genomic alterations in IDH-mutant subtypes for genes frequently altered in IDH-mutant glioma. **E**. Genomic distribution of 633 CpG probes differentially demethylated between co-clustered G-CIMP-low and G-CIMP-high. CpG probes are grouped by UCSC genome

browser defined CpG Islands, shores flanking CpG island +/- 2kb and open seas (regions not in CpG islands or shores). **F.** DNA methylation heatmap of TCGA glioma samples ordered per **Figure 2-3**a, and the epigenetically regulated (EReg) gene signatures defined for G-CIMP-low, G-CIMP-high and Codel subtypes. The mean RNA sequencing counts for each gene matched to the promoter of the identified cgID across each cluster are plotted to the right. **G.** Heatmap of the validation set classified using the random forest method. 1,300 probes defined in **Figure 2-3**a. **H.** Heatmap of probes differentially methylated between G-CIMP-low and G-CIMP-high in longitudinally matched tumor samples.

Supervised gene expression pathway analysis of the genes activated in the G-CIMP-low group as opposed to G-CIMP-high group revealed activation of genes involved in cell cycle and cell division consistent with the role of SOX in promoting cell proliferation (**Figure 2-5e**). The enrichment in cell cycle gene expression provides additional support to the notion that development of the G-CIMP-low subtype is associated with activation of cell cycle progression and may be mediated by a loss of CpG methylation and binding of SOX factors to candidate genomic enhancer elements.

To validate the G-CIMP-low, G-CIMP-high and Codel IDH-mutant subtypes, we compiled a validation cohort from published studies including 324 adult and pediatric gliomas [84, 86, 113, 114]. The CpG probe methylation signatures used to classify the validation set are provided on the publication portal accompanying this publication (https://tcga-data.nci.nih.gov/docs/publications/lgggbm_2015/). Among them, 103 were identified as IDH-mutant on the basis of their genome wide DNA methylation profile. We classified samples in the validation set using the probes that defined the IDH-mutant specific DNA methylation cluster analysis integrated in a supervised random forest method. The analysis recapitulated the clusters generated from the TCGA collection (**Figure 2-5c**). In order to determine epigenetically regulated (EReg) genes that may be characteristic of the biology of the IDH-mutant diffuse glioma subtypes, we compared 450k methylation DNA methylation profiles and gene expression levels between 636 IDH-mutant and IDH-wildtype gliomas and 110 non-tumor samples from eleven different tissue types [115]. From the list of epigenetically regulated genes we extracted 263 genes that were grouped into EReg gene signatures which showed differential signals amongst the three IDH-mutant subtypes (**Figure 2-7f**). These trends were confirmed in the validation set (**Figure 2-7g**).

We investigated the possibility that the G-CIMP-high group is a predecessor to the G-CIMP-low group by comparing the DNA methylation profiles from ten IDH-mutant-non-codel LGG and GBM primary-recurrent cases with the TCGA

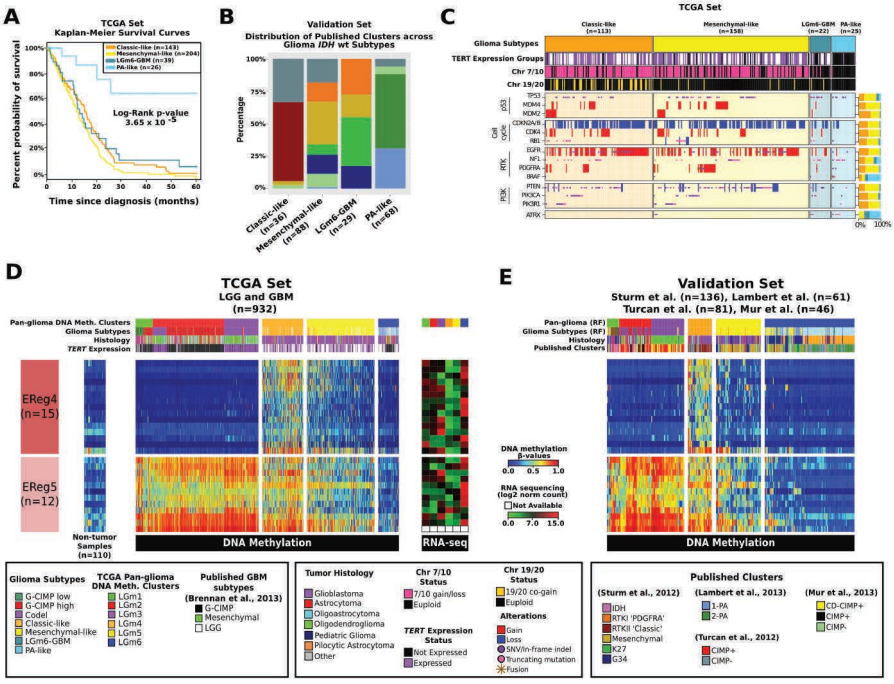
cohort. We evaluated the DNA methylation status of probes identified as differentially methylated (*n* = 90) between G-CIMP-low and G-CIMP-high (FDR < 10^{-13}, difference in mean methylation beta-value > 0.3 and < -0.4). Four out of ten IDHmut-non-codel cases showed a demethylation pattern after disease recurrence while partial demethylation was demonstrated in the remaining six recurrences, supporting the notion of a progression from G-CIMP-high to G-CIMP-low phenotype (**Figure 2-7h**).

An IDH-wildtype subgroup of histologically-defined diffuse glioma is associated with favorable survival and shares epigenomic and genomic features with pilocytic astrocytoma

IDH-wildtype gliomas segregated into three DNA methylation clusters (**Figure 2-6a**). The first is enriched with tumors belonging to the classical gene expression signature and was labeled Classic-like, whereas the second group, enriched with mesenchymal subtype tumors, was labeled Mesenchymal-like (***Table S1**, available online with the full-text of this article at www.cell.com*)[22]. The third cluster contained a larger fraction of LGG in comparison to the other IDH-wild type clusters. We observed that the IDH-wildtype LGGs but not the IDH-wildtype GBM in this cluster displayed markedly longer survival (log-rank *P*-value =3.6 x 10^{-5}; **Figure 2-8a**) and occurred in younger patients (mean 37.6yr vs 50.8yr, *t*-test *P*-value = 0.002). Supervised analysis of differential methylation between LGG and GBM in the third DNA methylation cluster did not reveal any significant probes despite significant differences in stromal content (p-value < 0.005; **Figure 2-6d**), suggesting that this group cannot be further separated using CpG methylation markers.

Next, we sought to validate the methylation-based classification of IDH-wildtype glioma in an independent cohort of 221 predicted IDH-wildtype glioma samples, including 61 grade I pilocytic astrocytomas (PA). Towards this aim, we used a supervised random forest model built with the probes that defined the IDH-wildtype clusters. Samples classified as Mesenchymal-like showed enrichment for the Sturm *et al.* Mesenchymal subtype (29/88) and gliomas predicted as Classic-like were all RTK II 'Classic' (22/22), per the Sturm *et al.* classification (**Figure 2-8b**; **Figure 2-6b**) [86]. We observed that PA tumors were unanimously classified as the third, LGG-enriched group (**Figure 2-6b**). Based on the molecular similarity with PA we labeled the LGGs in the third methylation cluster of IDH-wild type tumors as PA-like. The GBMs in this group were best described as LGm6-GBM, for their original pan-glioma methylation cluster assignment and tumor grade.

Pilocytic astrocytomas are characterized by frequent alterations in the MAPK pathway, such as *FGFR1* mutations, *KIAA1549-BRAF* and *NTRK2* fusions [116]. The frequency of mutations, fusions and amplifications in eight PA-associated genes (*BRAF*, *NF1*, *NTRK1*, *NTRK2*, *FGFR1*, and *FGFR2*) rated from 11% (*n* = 12/113) of Classic-like, 13% (*n* = 21/158) of Mesenchymal-like IDH-wildtype tumors to 32% (*n* = 7/22) of LGm6-GBM and 52% (*n* = 13/25) of PA-like LGG (Fisher Exact Test (FET) *P*-value < 0.0001; **Figure 2-8c**). Conversely, only two of 25 (8%) PA-like LGG tumors showed *TERT* expression, compared to five of 12 LGm6-GBM (43%), 60 of 65 Classic-like (92%) and 82 of 98 Mesenchymal-like (84%, FET *P*-value < 0.0001).



↑ **Figure 2-8. A distinct subgroup of IDH-wildtype diffuse glioma with molecular features of pilocytic astrocytoma.**

A. Kaplan-Meier survival curves for the IDH-wildtype glioma subtypes. Ticks represent censorship. **B.** Distribution of previous published DNA methylation subtypes in the validation set, across the TCGA IDH-wildtype specific DNA methylation clusters. **C.** Distribution of genomic alterations in genes frequently altered in IDH-wildtype glioma. **D.** Heatmap of TCGA glioma samples ordered according to **Figure 2-3a** and two EReg gene signatures defined for the IDH-wildtype DNA methylation clusters. Mean RNA sequencing counts for each gene matched to the promoter of the identified cgID across each cluster are plotted to the right. **E.** Heatmap of the validation set classified using the random forest method using the 1,300 probes defined in **Figure 2-3a**.

The PA-like group was characterized by relatively low frequency of typical GBM alterations, in genes such as *EGFR*, *CDKN2A/B*, and *PTEN* and displayed euploid DNA copy number profiles (**Figure 2-6e**). To ascertain that the histologies of the PA-like subgroup had been appropriately classified, we conducted an independent re-review. This analysis confirmed the presence of the histologic features of diffuse glioma (grade II or grade III) in 23 of the 26 cases in the cluster. The remaining three cases were re-named as PA (grade I). An independent review of the magnetic resonance diagnostic images from thirteen cases showed a similar pattern, with the majority of tumors showing behavior consistent with grade II or grade III glioma. Taken together, the epigenetic analysis of the IDH-wildtype group of adult glioma revealed the existence of a novel subgroup sharing genetic and DNA methylation features with pediatric PA, and favorable clinical outcome compared to diffuse IDH-wildtype glioma. This group may include but extends beyond *BRAF*-mutated grade II oligodendroglioma that were previously recognized as a unique clinical entity [117].

Through comparison of the methylation profiles of 636 glioma and 110 non-neoplastic normal samples from different tissue types, we defined EReg signatures consisting of 27 genes that showed differential signals amongst IDH-wildtype subtypes in the TCGA (**Figure 2-8d**) and the validation set (**Figure 2-8e**). EReg4 comprised a group of 15 genes hypermethylated and downregulated in particularly Classic-like. EReg5 was defined as a group of 12 genes associated with hypomethylation and upregulation in LGm6/PA-like compared to all other LGm clusters. These ERegs aided in characterizing the biological importance of IDH-wildtype subtypes and were subsequently used to evaluate the prognostic importance of the IDH-wildtype clusters.

The epigenetic classification of glioma provides prognostic value independent of age and grade

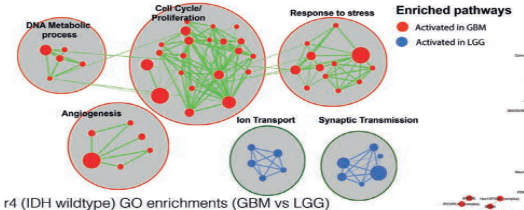
In order to assess whether the DNA methylation-based subtypes we identified carry prognostically relevant information independent of known overall survival predictors, we constructed a series of survival regression models. To find the optimal model for survival prediction we studied covariates individually and in combination with other covariates. Age at diagnosis, histology, IDH/codel subtype, *TERT* expression and epigenetic subtype all contribute to survival in single-predictor analysis (log-rank *P*-value < 0.05, ***Table S4****, available online with the full-text of this article at www.cell.com*).

Table 2-2. Methylation subtypes are prognostically relevant in multivariable analysis and in external validation data.

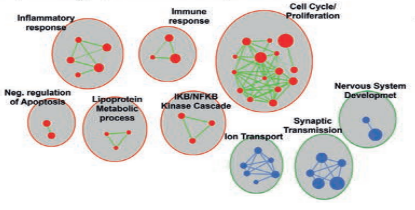
		Discovery (*n* = 809)			Validation (*n* = 183)		
		C-Index: 0.835 ± 0.019			C-Index: 0.745 ± 0.032		
Predictor	**Levels**	**N**	**HR (95% CI)**	**Signif.**	**N**	**HR (95% CI)**	**Signif.**
Age at diagnosis	per year	809	1.05 (1.03-1.06)	***	183	1.02 (1-1.04)	*
WHO Grade	II	214	1.0 (ref)		41	1.0 (ref)	
	III	241	1.96 (1.15-3.33)	*	51	1.24 (0.55-2.76)	
	IV	354	2.38 (1.3-4.34)	*	91	2.6 (1.08-6.3)	*
Subgroup	IDHmut-codel	156	1.0 (ref)		57	1.0 (ref)	
	G-CIMP-low	22	5.6 (2.49-12.62)	***	2	0 (0-Inf)	
	G-CIMP-high	219	1.92 (1.05-3.51)	*	15	1.25 (0.43-3.66)	
	Classic-like	143	5.4 (2.79-10.44)	***	22	4.55 (1.8-11.49)	*
	Mesenchymal-like I	204	8.71 (4.59-16.53)	***	61	5.55 (2.52-12.21)	***
	LGm6-GBM	39	5.79 (2.78-12.1)	***	22	6.8 (2.58-17.91)	**
	PA-like	26	2.02 (0.71-5.71)		4	3.64 (0.79-16.78)	.

Survival regression analysis indicates that an optimal model of prognosis includes age, grade and methylation subtype. These predictors are statistically significant in both our discovery dataset and an external validation dataset. Signif. codes: 0 '***' 0.001 '**' 0.05 '*' 0.1 '.'

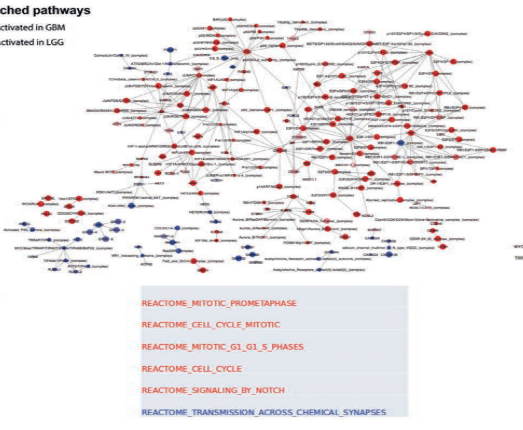
r3 (IDH mutant) GO enrichments (GBM vs LGG)



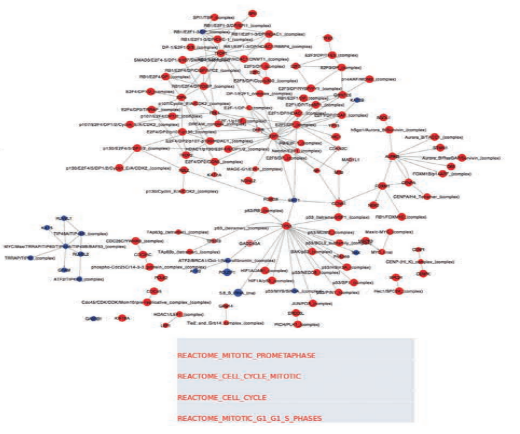
r4 (IDH wildtype) GO enrichments (GBM vs LGG)



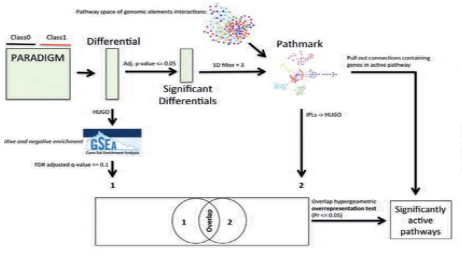
C. LGr3 (IDH mutant) PARADIGM results (GBM vs LGG)



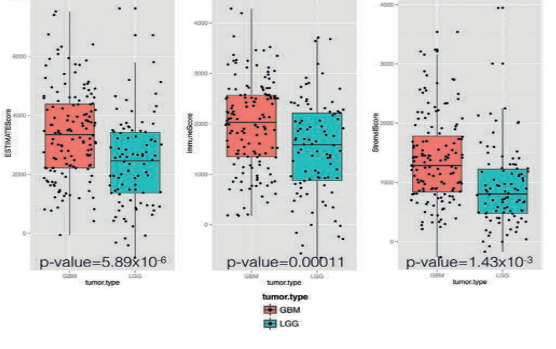
D. LGr4 (IDH wildtype) PARADIGM results (GBM vs LGG)



e process for extracting significantly active pathway from the glioma data



F.EstimateScore by Tumor type (IDH-wildtype)



← Figure 2-9. Progression of LGG to GBM is marked by cell cycle/proliferation or invasion/microenvironmental changes.

The pathways involved with progression from LGG to GBM were identified through supervised analysis of co-clustered LGG and GBM using Gene Set Enrichment Analysis. Gene sets were compiled from the Gene Ontology pathway database. Significantly enriched gene sets (FDR < 0.1, p-value < 0.005) were depicted as an annotation module network using Cytoscape and EnrichmentMap. Nodes represent enriched gene sets, which are grouped and annotated by their similarity. Node size is proportional to the total number of genes within each gene set. Proportion of shared genes between gene sets is represented as the thickness of the line between nodes. **A.** Progression of LGG IDH-mutant-non-codel to GBM G-CIMP in LGr3 was strongly marked by a hyper-proliferation signature and revealed four major gene sets groups related to cell cycle and hyperproliferation, DNA metabolic processes, response to stress and angiogenesis. **B.** Similar analysis of the gene sets activated in the GBM compared to the LGG component of LGr4 (IDH-wildtype) identified an inflammation and immunologic response signature characterized by the activation of several chemokines and interleukins enriching sets involved in inflammatory and immuno response, negative regulation of apoptosis, cell cycle and proliferation, IKB/NFKB kinase cascade. **C.** Differential regulatory networks describing differential molecular activities between GBM and LGG in LGr3. Dichotomies were selected by only choosing those where samples form tight linearly separable clusters in the high dimensional genomic space. The size of the node is inversely proportional to the magnitude of the p-value computed by LIMMA for each differential. Curated canonical MSigDB pathways significantly represented in these networks are listed below each network, following the same color scheme as described above. **D.** Same as in C for LGr4. **E.** Overview of the adopted pipeline for extracting significant pathways. **F.** Distribution of Estimate, Immuno and Stromal score by tumor type in the IDH-wildtype samples.

As expected, age was a highly significant predictor (*P*-value < 0.0001, C-Index 0.78) and was included in all subsequent multi-predictor models. We found that histology and grade are highly correlated. Histology provided only marginal improvement to a model that includes grade (likelihood ratio test (LRT) *P*-value = 0.08) and was therefore not included in further analyses.

Conversely, grade markedly impacted a histology-based predictor model (LRT *P*-value = 0.0005, ***Table S4****, available online with the full-text of this article at www.cell.com*) and was retained in the subsequent models. In contrast to previous reports [12], we failed to observe a statistically significant and independent survival association with *TERT* expression (LRT *P*-value = 0.82, ***Table S4****, available online with the full-text of this article at www.cell.com*) or *TERTp* mutations, after accounting for age and grade (LRT *P*-value = 0.85, not shown). Thus, the optimal survival prediction model includes age, grade and epigenetic subtype (LRT *P*-value < 0.0001, C-Index 0.836; **Table 2-2**).

To confirm that the epigenetic subtypes provide independent prognostic information we tested the survival model on the validation dataset. Epigenetic subtypes in these samples were determined as described above. The distinction

between LGm6-GBM and PA-like gliomas was made on the basis of tumor grade and not by DNA methylation signature. Using a subset of 183 samples in the validation cohort with known survival, age and grade we found that epigenetic subtypes are significant independent predictors of survival in the multivariate analysis (LRT *P*-value < 0.0001, C-Index 0.746, **Table 2-2**). This generalization of our model supports the epigenetic subtypes as a means to improve the prognostication of glioma.

Activation of cell cycle/proliferation and invasion/microenvironmental changes marks progression of LGG to GBM

We observed that in spite of morphological differences between LGG and GBM, such as high cell density and microvascular proliferation, clustering of gene expression profiles frequently grouped LGG and GBM together within the same subtype. Gene Set Enrichment Analysis of the genes activated in G-CIMP GBM as opposed to the IDH-mutant-non-codel within LGr3 (**Figure 2-3b**) revealed four major groups, including cell cycle and hyperproliferation, DNA metabolic processes, response to stress, and angiogenesis (**Figure 2-9a**, ***Table S5****, available online with the full-text of this article at www.cell.com*).

These biological functions are consistent with the criteria based on mitotic index used by pathologists to discriminate lower and high-grade glioma and the significance of activated microglia for tumor aggressiveness [118]. Conversely, compared with the G-CIMP GBM, IDH-mutant-non-codel LGG in LGr3 were characterized by enrichment of genes associated with neuro-glial functions such as ion transport and synaptic transmission, possibly suggesting a more differentiated nature.

The comparison of co-clustered GBM and LGG in LGr3 by the PARADIGM algorithm that integrates DNA copy number and gene expression to infer pathway activity, confirmed that GBMs express genes associated with cell cycle, proliferation and aggressive phenotype, through activation of a number of cell cycle, cell replication and NOTCH signaling pathways whereas LGGs exhibit an enrichment of neuronal differentiation specific categories including synaptic pathways (**Figure 2-9c**, ***Table S5****, available online with the full-text of this article at www.cell.com*).

The analysis of the genes activated in GBM versus the LGG component of LGr4, which grouped IDH-wildtype tumors, identified an inflammation and immunologic response signature characterized by the activation of several chemokines (*CCL18, CXCL13, CXCL2, CXCL3*) and interleukins (*IL8, CXCR2*) enriching sets involved

in inflammatory and immune response, negative regulation of apoptosis, cell cycle and proliferation, and the IKB/NFKB kinase cascade Map (**Figure 2-9b**, ***Table S5****, available online with the full-text of this article at www.cell.com*). These characteristics suggest differences in the relative amount of microglia. We used the ESTIMATE method to estimate the relative presence of stromal cells which revealed significantly lower (p-value 10^{-6}) stromal scores of LGG IDH-wildtype versus GBM IDH-wildtype (**Figure 2-9f**) [119]. Resembling the functional enrichment for LGG within LGr3, functional enrichment of LGG IDH-wildtype in comparison to GBM within LGr4 showed activation in LGG of special glial-neuronal functions involved in ion transport, synaptic transmission and nervous system development. Similar differential pathway activities are observed in GBM versus LGG when we split LGr4 in the three IDH-wildtype methylation clusters (**Figure 2-9e**; Table S5).

Finally, we aimed to identify transcription factors that may exert control over prominent gene expression programs, known as master regulators. Master regulator analysis comparing the IDH-wildtype group to the IDH-mutant group revealed transcription factors that were upregulated in IDH-wildtype gliomas and showed an increase in expression of target genes, including *NKX2-5*, *FOSL1*, *ETV4*, *ETV7*, *RUNX1*, *CEBPD*, *NFE2L3*, *ELF4*, *RUNX3*, *NR2F2*, *PAX8* and *IRF1 (**Table S5**, available online with the full-text of this article at www.cell.com*). No TFs were found to be upregulated in IDH-mutant gliomas relative to IDH-wildtype gliomas (at a log fold change >1).

Discussion

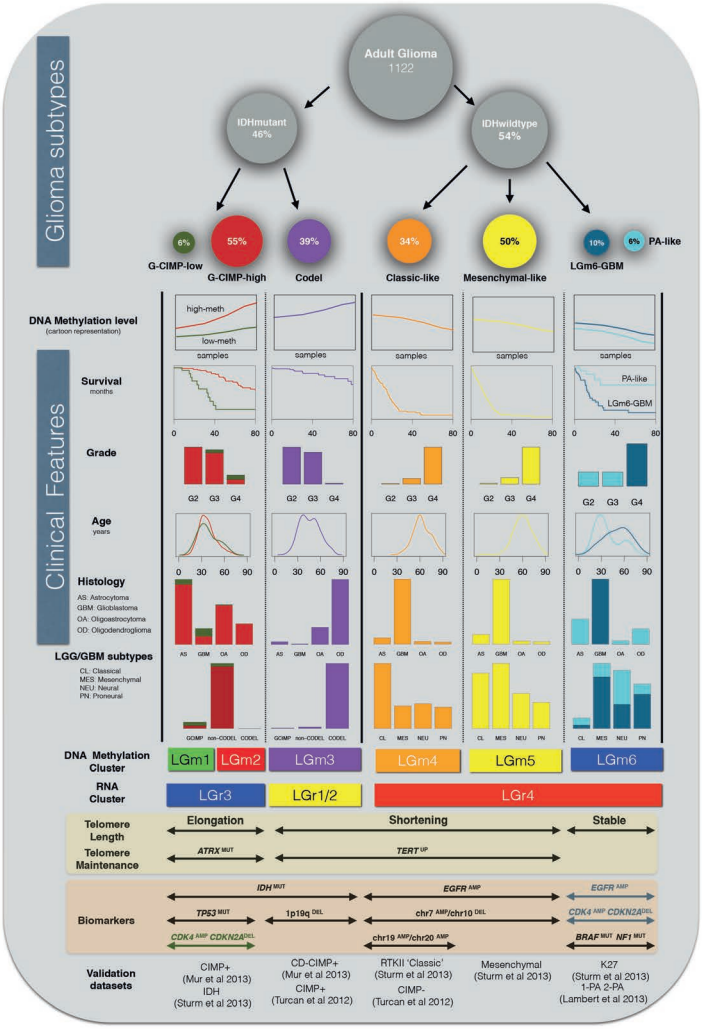
This study represents the largest multi-platform genomic analysis performed to date of adult diffuse glioma (WHO grades II, III and IV). A simplified graphical summary of the identified groups and their main clinical and biological characteristics is reported in **Figure 2-10**. The clustering of all diffuse glioma classes and grades within similarly shaped methylation-based and expression-based groups has allowed us to pinpoint specific molecular signatures with clinical relevance. The DNA methylation classification proposed should be considered as a basis and it is likely that future studies involving significantly larger cohorts and more refined profiling methods will be able to further reduce intra-subtype heterogeneity. The dissection of the IDH-mutant non-codel G-CIMP LGG and GBM into two separate subgroups (G-CIMP-low and G-CIMP-high) based on the extent of genome-wide DNA methylation has crucial biological and clinical relevance. In particular, the identification of the G-CIMP-low subset, characterized by activa-

tion of cell cycle genes mediated by SOX binding at hypomethylated functional genomic elements and unfavorable clinical outcome is an important finding that will guide more accurate segregation and therapeutic assessment in a group of patients in which correlations of conventional grading with outcome are modest [120, 121]. The finding that G-CIMP-high tumors can emerge as G-CIMP-low glioma at recurrence identify variations in DNA methylation as crucial determinants for glioma progression and provides a clue to the mechanisms driving evolution of glioma. Our results unify previous observations that linked the cell cycle pathway to malignant progression of low-grade glioma [122]. Future updates of the TCGA glioma clinical annotation and independent validation of our findings may be able to consider additionally important clinical confounders such as extent of resection and performance status, to further optimize the weights of the currently known prognostic variables and their association to the molecular subtypes we identified.

Analysis of IDH-wildtype glioma revealed the PA-like LGG subset that harbors a silent genomic landscape, confers favorable prognosis relative to other IDH-wildtype diffuse glioma and displays a molecular profile with high similarity to PA. Re-review by neuropathologists and neuroradiologists confirmed that the majority were correctly diagnosed as diffuse glioma, emphasizing the need for integration of molecular signatures intro clinical classification [117] for this subgroup of patients that may be spared potentially unnecessary intensive treatments.

The large number of exomes in our dataset allowed identification of novel glioma-associated somatic alterations, including in the *KRAS* and *NRAS* genes which were frequently used to genetically engineered glioma mouse models [123]. Our analysis further nominates glial tumors to join an increasing number of tumor types characterized by a deactivated cohesin pathway [102, 124]. Cohesin mutant tumors may infer increased sensitivity to DNA damage agents and PARP inhibitors [125] suggesting that gliomas with genetic alterations of key cohesin regulatory factors may represent biomarkers and therapeutic opportunities.

Overexpression of *TERT* mRNA was found to be associated with increased telomere length in urothelial cancer [126]. Our results revealed that in gliomas, increased telomere length is associated with *ATRX* mutations, suggesting an ALT mechanism. ALT has been associated with sensitivity to inhibition of the protein kinase ATR [127].



↑ **Figure 2-10. Overview of major subtypes of adult diffuse glioma.**

Integrative analysis of 1,122 adult glioma resulted in seven different subtypes with distinct biological and clinical characteristics. The groups extend across six DNA methylation subtypes of which the LGm6 cluster was further separated by tumor grade, into PA-like and LGm6-GBM. The size of the circles is proportional to the percentages of samples within each group. DNA methylation plot is a cartoon representation of overall genome-wide epigenetic pattern within glioma subtypes. Survival information is represented as a set of Kaplan-Meier curves, counts of grade, histology and LGG/GBM subtypes within the groups are represented as bar-plots, whereas age is represented as density. Labeling of telomere length and maintenance status is based on the enrichment of samples within each column, similarly for the biomarkers and the validation datasets.

In summary, our pan-glioma analysis has expanded our knowledge of the glioma somatic alteration landscape, emphasized the relevance of DNA methylation profiles as a modality for clinical classification, and quantitatively linked somatic TERT pathway alterations to telomere maintenance. Combined, these findings are an important step forward in our understanding of glioma as discrete disease subsets, and the mechanisms driving gliomagenesis.

Experimental procedures

Patient and Sample Characteristics

Specimens were obtained from patients, with appropriate consent from institutional review boards. Details of sample preparation are described in the Extended Experimental Procedures.

Data Generation

In total, tumors from 1,132 patients were assayed on at least one molecular profiling platform, which platforms included: (1) whole genome sequencing, including high coverage and low pass whole genome sequencing, (2) exome sequencing (3) RNA sequencing (4) DNA copy-number and single-nucleotide polymorphism arrays, including Agilent CGH 244K, Affymetrix SNP6.0 and Illumina 550K Infinium HumanHap550 SNP Chip microarrays (5) gene expression arrays, including, Agilent 244K Custom Gene Expression, Affymetrix HT-HGU133A and Affymetrix Human Exon 1.0 ST arrays (5) DNA methylation arrays, including Illumina GoldenGate Methylation, Illumina Infinium HumanMethylation27, and Illumina Infinium HumanMethylation450 BeadChips (7) reverse phase protein arrays, (8) miRNA sequencing and (9) miRNA Agilent 8 x 15K Human miRNA-specific microarrays. Details of data generation have been previously reported [24, 100]. To ensure cross-platform comparability, features from all array platforms were compared to a reference genome.

Data Analysis

The data and analysis results can be explored through the Broad Institute FireBrowse portal (*http://firebrowse.org/?cohort=GBMLGG*), the cBioPortal for Cancer Genomics (*http://www.cbioportal.org/study.do?cancer_study_ id=lgggbm_tcga_pub*), in a Tumor Map (*http://tumormap.ucsc.edu/?p=ynewton.gliomas-paper*), the TCGA transcript fusion portal (*http://www.tumorfusions.org*), TCGA Batch Effects

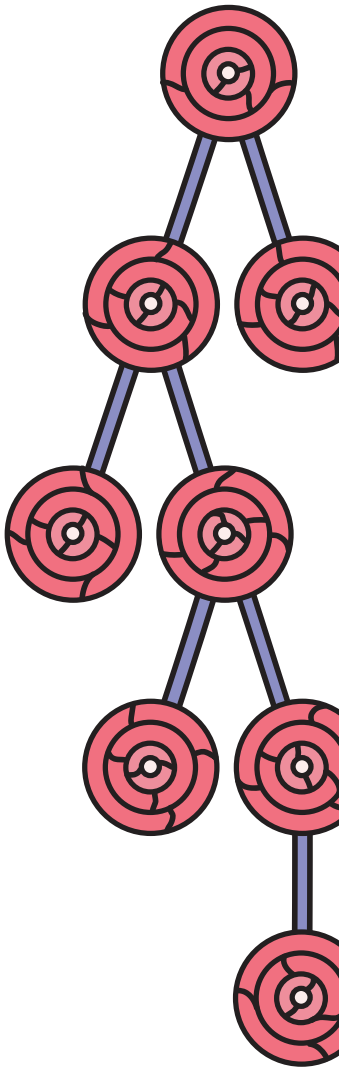
(http://bioinformatics.mdanderson. org/tcgambatch/), Regulome Explorer (*http://explorer.cancerregulome.org/*), Next-Generation Clustered Heat Maps (*http://bioinformatics.mdanderson. org/TCGA/NGCHMPortal/*). See also Supplemental Information and the TCGA publication page (*https://tcga-data.nci.nih.gov/docs/publications/lgggbm_2015/*).

Acknowledgements

This study was supported by NIH grants: U24CA143883, U24CA143858, U24CA143840, U24CA143799, U24CA143835, U24CA143845, U24CA143882, U24CA143867, U24CA143866, U24CA143848, U24CA144025, U54HG003067, U54HG003079, U54HG003273, U24CA126543, U24CA126544, U24CA126546, U24CA126551, U24CA126554, U24CA126561, U24CA126563, U24CA143731, U24CA143843, P30CA016672, P50 CA127001, R01 CA190121, P01 CA085878; Cancer Prevention & Research Institute of Texas (CPRIT) R140606; São Paulo Research Foundation (FAPESP) 2014/02245-3 and 2015/07925-5.

Author contributions

Conceptualization and project administration: R.G.W.V., A.I., and H.N.; Supervision: S.R.S., K.D.A., P.W.L., M.G., D.H., D.J. B., D.H.G., R.R., C.C.L., J.B.S., C.G.C.J., D.P.C.T., W.K.A.Y., J.H., L.C., M.M., and T.M.; Formal Analysis: R.G.W.V., A.I., H.N., M.C., F.P.B., T.M.M., T.S.S., O.M., Y.N., S.M.P., P.Z., L.P., Amie R., G.R., R.A., J.W. G.M., S.L., S.A., Arvind R., B.A.M., A.D.C., and H.Z.; Investigation: D.J.B., L.C., and L.P.; Data Curation: D.J.B., L.P., and F.P.B; Writing - Original Draft: R.G.W.V., A.I., H.N., M.C., F.P.B., T.M.M., and T.S.S.; Manuscript review: DJB, KAD, SRS, MW, NL, DHG.



CHAPTER 3

Longitudinal molecular trajectories of diffuse glioma in adults

Published in print

Barthel FP, Johnson KC, Varn FS, Moskalik AD, Tanner G, The GLASS Consortium, Huse JT, Groot JF de, Stead LF, Verhaak RGW. *Longitudinal Molecular Trajectories of Diffuse Glioma in Adults.* Nature. 2019 Dec;576(7785):112-120.

Abstract

The evolutionary processes that drive universal therapeutic resistance in adult patients with diffuse glioma remain unclear [106, 128]. Here, we analyzed temporally separated DNA sequencing data and matched clinical annotation from 222 patients with glioma. Through mutational and copy number analyses across the three major subtypes of diffuse glioma, we observed that driver genes detected at initial disease were retained at recurrence, while there was little evidence of recurrence-specific gene alterations. Treatment with alkylating-agents resulted in a hypermutator phenotype at different rates across glioma subtypes, and hypermutation was not associated with differences in survival. Acquired aneuploidy was frequently detected in recurrent gliomas characterized by presence of an IDH mutation but without 1p/19q codeletion and further converged with acquired cell cycle alterations and poor outcomes. We show that the clonal architecture of each tumor remains similar over time and that absence of clonal selection was associated with increased survival. Finally, we did not observe differences in immunoediting levels between initial and recurrent glioma. Our results collectively argue that the strongest selective pressures occur early during glioma development and that current therapies shape this evolution in a largely stochastic manner.

Introduction

Diffuse glioma is the most common malignant brain tumor in adults and invariably relapse despite treatment with surgery, radiotherapy, and chemotherapy. The molecular landscape of glioma at diagnosis has been extensively characterized [15, 16, 22, 100, 129-131]. While these efforts have led to the identification of driver genes and clinically relevant subtypes [59, 61], it is unknown how the glioma genetic landscape evolves over time and in response to therapy.

Intratumoral heterogeneity is a well-recognized characteristic of gliomas and results from selective pressures such as a limited availability of nutrients, clonal competition, and treatment [71, 132-134]. Tumors are thought to circumvent these growth bottlenecks via dynamic competition of subclones resulting in the most favorable environment for tumor sustenance [128]. Recent studies have suggested that stochastic changes in clone frequency (i.e. neutral evolution) and immunogenic surveillance may further contribute to the observed intratumoral heterogeneity [135, 136]. An understanding of evolutionary dynamics at multiple time points is needed to develop strategies aimed at delaying or preventing the onset of tumor progression.

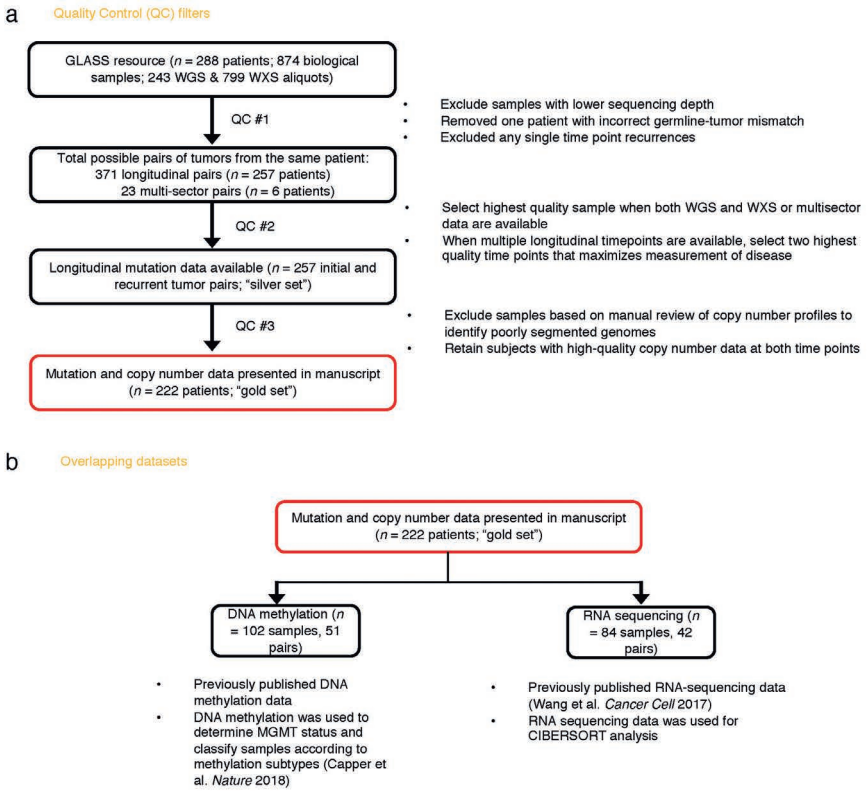
To investigate clonal dynamics over time and in response to therapeutic pressures, we established the Glioma Longitudinal Analysis (GLASS) Consortium. GLASS is a community-driven effort that seeks to overcome the logistical challenges in constructing adequately powered longitudinal genomic glioma datasets by pooling datasets from patients treated at institutions worldwide [137]. We have analyzed longitudinal profiles across the three molecular glioma subtypes to identify the molecular processes active at initial and recurrent time points. These analyses identified few common features of glioma evolution across subtypes, and instead pointed toward highly variable and patient-specific trajectories of genomic alterations.

Results

GLASS cohort

We pooled existing and newly generated longitudinal DNA sequencing datasets from 288 patients treated at 35 hospitals (**Table 3-1**, **Figure 3-1**). After applying quality filters, tumor samples from 222 patients with high-quality data in at least two time points were classified according to molecular markers into

three major glioma subtypes: 1. IDH-mutant and chromosome 1p/19q co-deleted (IDHmutant-codel; *n* = 25) 2. IDH-mutant without chromosome 1p/19q codeletion (IDHmutant-noncodel; *n* = 63) and 3. IDH wild type (IDHwt; *n* = 134), in alignment with the World Health Organization classification of Central Nervous System tumors [59, 61]. For each patient we selected two time-separated tumor samples, henceforth initial and recurrence, for further analysis.



↑ **Figure 3-1. Sample Selection.**

a. Quality control workflow steps identifying all GLASS samples available as a resource and the identification of the highest quality set of patient pairs (*n* = 222) used for the presented mutational and copy number analyses. **b.** Additional available datasets.

Table 3-1. Sequencing Centers and Hospitals with corresponding GLASS barcode designations.

GLASS Barcode	Sequencing Centers	Data Accession Numbers/ Additional Information	Hospitals	Patients (n)	Tumor Samples (n)	Contributors
6	*TCGA*	TCGA Data Portal *(https://tcga-data.nci.nih.gov/docs/publications/gbm_2013/)*	*Henry Ford Hospital*	7	26	TCGA GBM [24]
14			*Emory University*	3	7	
19			*Case Western*	3	10	
19	*Case Western*	Karnofsky Performance Status	*Case Western*	13	27	Barnholtz-Sloan J
AT	*Medical University of Vienna - CeMM*	*MGMT methylation status*	*Medical University of Vienna - CeMM*	8	21	Wohrer A
CU	*Columbia University*	Sequence Read Archive *(SRP074425)*	*Fondazione IRCCS, Istituto Neurologico C. Besta, Milan*	19	38	Iavarone A, Finocchiaro G [139]
DF	*Dana Farber Cancer Institute*	NA*	*Dana Farber Cancer Institute*	2	4	Beroukhim P
			Case Western	7	14	
DK	*German Cancer Research Center*	NA*	*NA*	15	30	Radlwimmer B et al.
DH	*TCGA*	TCGA Data Portal *(https://tcga-data.nci.nih.gov/docs/publications/lgg_2015)* CGHub *(https://cghub.ucsc.edu)*	*University of Florida*	1	2	TCGA LGG [13]
DU			*Henry Ford Hospital*	6	14	
FG			*Case Western*	3	7	
TM			*University of South Wales*	1	2	
TQ			*University of Sao Paulo*	3	7	

Table 3-1. Sequencing Centers and Hospitals with corresponding GLASS barcode designations. *Continued*

GLASS Barcode	Sequencing Centers	Data Accession Numbers/ Additional Information	Hospitals	Patients (n)	Tumor Samples (n)	Contributors
HK	*Jackson Laboratory*	NA*	*Chinese University of Hong Kong*	5	10	Verhaak RG [73]
JX		NA*	*Radboud University in Nijmegen*	1	2	
MD		NA*	*MD Anderson Cancer Center*	39	81	
MD	*MD Anderson Cancer Center*	Interval between surgeries and treatment, Karnofsky Performance Status, location of radiation site, chemotherapy treatment reason. NA*	*MD Anderson Cancer Center*	35	58	
HF		EGA: *EGAS00001001033*	*Henry Ford Hospital*	11	22	
LU	*University of Leeds (UK)*	NA*	*Leeds Teaching Hospital*	1	2	Stead LF [140]
			Royal Preston Hospital	4	8	
			The Walton Centre	4	8	
MG	*Massachusetts General Hospital*	dbGaP (phs1424.v1.p1)	*Massachusetts General Hospital (5 affiliated hospitals)*	12	242	Brastianos PK
NS	*Northern Sydney Cancer Centre*	NA*	*Northern Sydney Cancer Centre*	1	4	Khasraw M

Table 3-1. Sequencing Centers and Hospitals with corresponding GLASS barcode designations. *Continued*

GLASS Barcode	Sequencing Centers	Data Accession Numbers/ Additional Information	Hospitals	Patients (n)	Tumor Samples (n)	Contributors
SF	*UC San Francisco*	*EGA: EGAS00001000579, EGAS00001000685, EGAS00001001255*	*UC San Francisco*	40	95	Costello J [122, 141]
			Erasmus Medical Center			
			Groupe Hospitalier Pitie-Salperiere			
			University of Tokyo			
SM	*Samsung Medical Center*	EGA: *EGAS00001001041, EGAS00001001800*	*Samsung Medical Center*	35	75	Nam D [142]
		GEO: GSE63035	*Seoul National University Hospital*			
SU	*Kyoto University*	*EGA: EGAS00001001044*	*Nagoya University*	9	22	Suzuki H [99]
			Kumamoto University			
			Kyushu University			
			Oita University			
			Tokyo Women's Medical University			

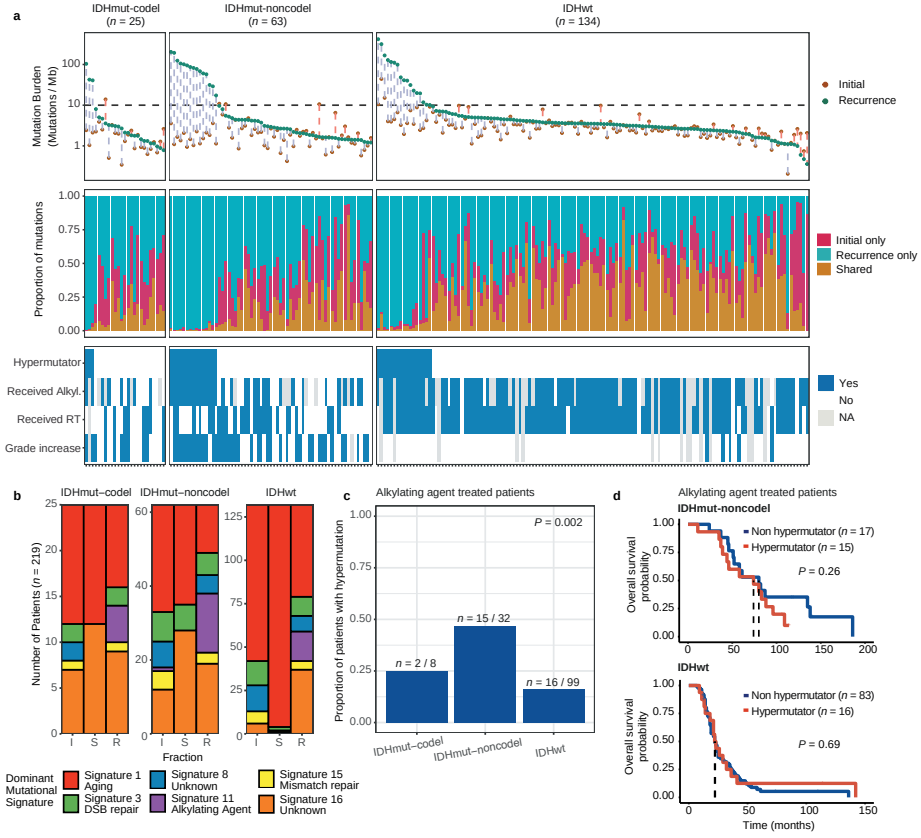


Figure 3-2. Temporal changes in glioma mutational burden and processes.

a. Each column represents a single patient (*n* = 222) at two separate timepoints grouped by glioma subtype and ordered left-to-right by decreasing mutation frequency at recurrence. Top, mutation frequency differences between initial and recurrent tumors. Blue dotted line indicates increased mutation frequency while a red dotted line indicates decreased mutational frequency. Stacked bar plot reflects the proportion of total mutations shared (mustard), private to initial (magenta), or private to recurrence (blue). Clinical information including hypermutation status, therapy, and grade changes. **b**. Stacked bar plot (n=219) indicating the dominant mutational signature among initial, recurrent and shared mutation fractions stratified by glioma subtype. **c**. The proportion of glioma recurrences with alkylating agent-related hypermutation, grouped by glioma subtype. Fisher's exact test was used to compare proportions between subtypes. **d**. Kaplan-Meier curve depicting overall survival in hypermutant (red) versus non-hypermutant (blue) alkylating agent treated patients amongst IDHwt (left, *n* = 99) and IDHmut-noncodel (right, *n* = 32) tumors. Log-rank test P-values are shown.

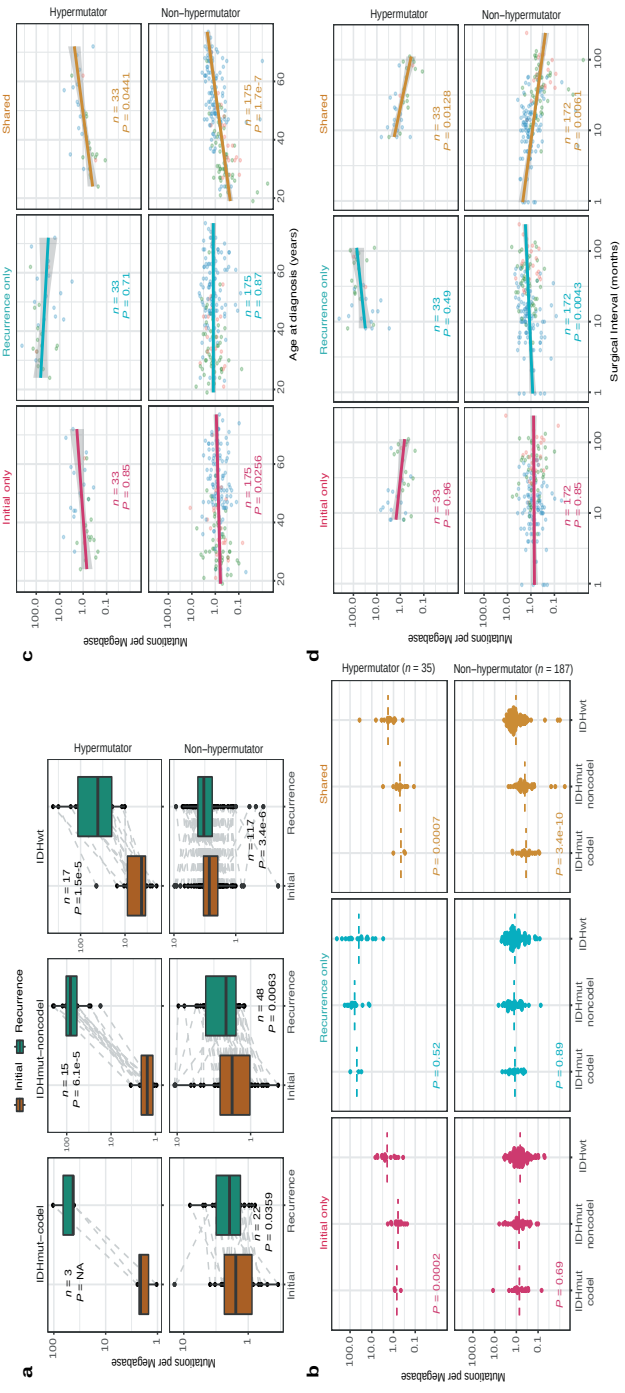
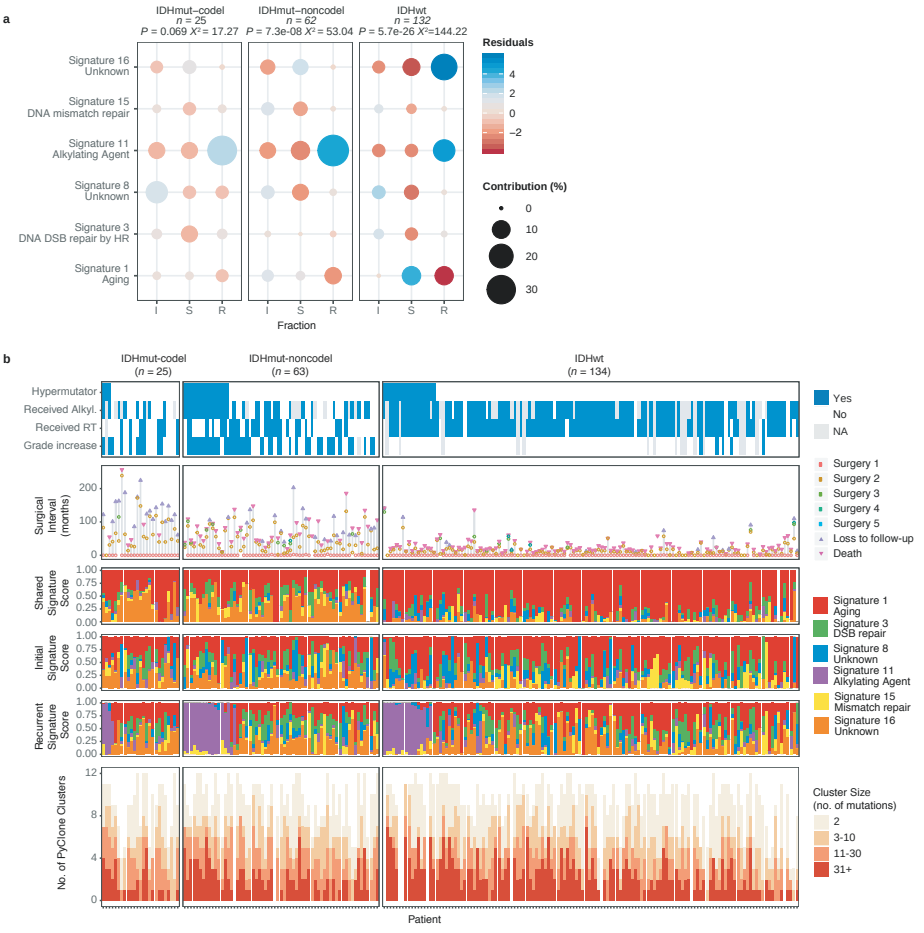


Figure 3-3. Mutation burden by time point and subtype.

a. Boxplots and paired lines depicting coverage adjusted mutation frequencies in initial and matched recurrent samples across three subtypes. Wilcoxon signed-rank test P-values and sample sizes are indicated. **b.** Bee swarm plot depicting coverage adjusted mutation frequencies in fractions by subtype. Dashed line indicates the mean. One-way ANOVA P-values comparing three subtypes are indicated. **c.** Scatter plot showing the relationship between age at diagnosis and coverage adjusted mutation burdens by subtype and fraction. Linear model P-values are indicated and were adjusted by subtype. **d.** Similar to the analysis presented in **c**, but showing the relationship between time to recurrence and coverage adjusted mutation burdens.



↑ **Figure 3-4. Mutational signatures by fraction and subtype.**

a. Correlation plot showing the Pearson's chi-squared (X^2) residuals for each signature by fraction and subtype. A X^2 was performed for each subtype and P-values are indicated. Positive residuals (blue) indicate a positive correlation, whereas negative residuals (red) indicate an anticorrelation. The point size reflects the contribution to X^2 estimate. **b.** The same ordered of patients as **Figure 3-2**a along with relevant clinical information is provided alongside the fraction-specific mutational signatures. PyClone mutational clusters are also presented.

Mutational burdens and processes over time

We first evaluated temporal changes in mutational burden and processes to understand general patterns of glioma evolution. Mutation burdens in initial tumors were comparable with previously reported rates [15, 100, 138]. 2.20 mutations (single-nucleotide variants and small insertions/deletions) per Megabase (Muta-

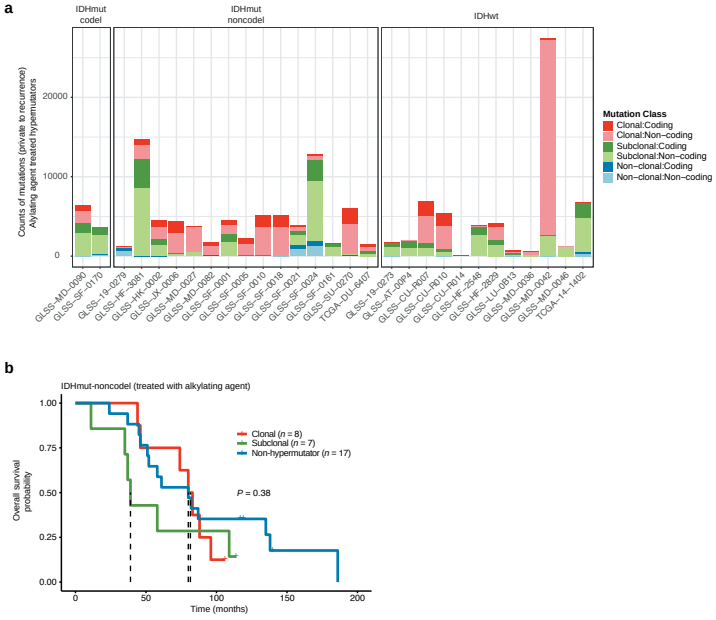
tions/Mb) for IDHmutant-codels; 2.52 Mutations/Mb for IDHmutant-noncodels; and 2.85 Mutations/Mb for IDHwt glioma (**Figure 3-2a**; **Figure 3-3a**). Excluding DNA hypermutation cases (> 10 Mutations/Mb, n = 35), the mutation burden increased after recurrence in 70% of the cohort (**Figure 3-3a**). To study changes during tumor progression, we separated mutations into three fractions: initial only, recurrence only, or shared. Interestingly, private fraction but not shared fraction mutation burdens were comparable between subtypes (**Figure 3-3b**). Patient age at diagnosis was significantly associated with the shared mutational burden and to a lesser extent the mutation burden private to the initial tumor (**Figure 3-3c**). On average, tumors with longer time to recurrence had slightly higher mutation burdens (**Figure 3-3d**).

These fraction-specific differences in mutation burden suggested that the activity of distinct mutational processes may also be time-dependent. We therefore classified mutations in each fraction according to the Catalogue of Somatic Mutations in Cancer (COSMIC) signature database [143]. As expected, signature activity was closely related to subtype and fraction (**Figure 3-2b**, **Figure 3-4a**). Signature 1 (aging) was nearly always the dominant signature amongst shared mutations in IDHwt tumors, whereas the shared fraction in IDHmut-noncodel and IDHmut-codel tumors - tumor subtypes associated with a younger age of diagnosis - additionally showed a strong presence of signature 16 (unknown etiology). Signatures 3 (double strand break repair) and 15 (mismatch repair) along with signature 8 (unknown etiology) were mostly confined to the private fractions, suggesting that these processes were of lesser importance to tumor maintenance than those associated with aging.

Treatment of glioma includes alkylating agents that can induce post-treatment hypermutation [73, 139, 141]. We observed enrichment of the associated signature 11 in recurrent tumors with a mutational load exceeding 10 Mutations/Mb and treated with alkylating agents (**Figure 3-2a**, **Figure 3-4b**). Treatment-associated hypermutation occurred most frequently among IDHmutant-noncodels (47%), followed by IDHmutant-codels (25%), and IDHwt gliomas (16%) (**Figure 3-2c**). The difference in the proportion of hypermutation events was significantly different between the three glioma subtypes (Fisher's exact-test P = 2.0e-03), suggesting that IDHmutant noncodels are most sensitive to developing a hypermutator phenotype [144].

Treatment-induced hypermutation has been associated with disease progression [141]. We did not find overall survival differences between alkylating agent-treated hypermutators and alkylating agent-treated non-hypermuta-

tors independent of age, subtype, and *MGMT* methylation status (**Figure 3-2d, Table 3-2, Table 3-3**). In order to further assess the pathogenicity of acquired mutations, we studied their clonality [145]. Newly acquired clonal mutations have penetrated most of the tumor (i.e., a selective sweep) between initial and recurrence and mark clonal expansion [146]. Conversely, acquired subclonal mutations are less prevalent, and therefore less likely to drive disease progression. Previous reports have suggested that alkylating agent-associated mutations hypermutation are frequently clonal [147]. We found that in 48% of hypermutated tumors a majority of the recurrence-only mutations were clonal, potentially reflecting cases where a selective sweep occurred (**Figure 3-5a**).



↑ **Figure 3-5. Hypermutator clonality.**

a. Bar plots represent counts of recurrence-only mutations per hypermutator tumor that were known to receive alkylating agent therapy and were successfully run through the PyClone algorithm. Colors indicate mutation clonality and color intensity indicates whether the mutations resulted in coding changes. **b.** Kaplan-Meier curve comparing alkylating agent-treated patients with IDHmut-noncodel hypermutator tumors that were predominantly clonal (*n* = 8), predominantly subclonal (*n* = 7), versus IDHmut-noncodel non-hypermutators known to be treated with alkylating agents and had available PyClone data (*n* = 17). Log-rank P-value is shown.

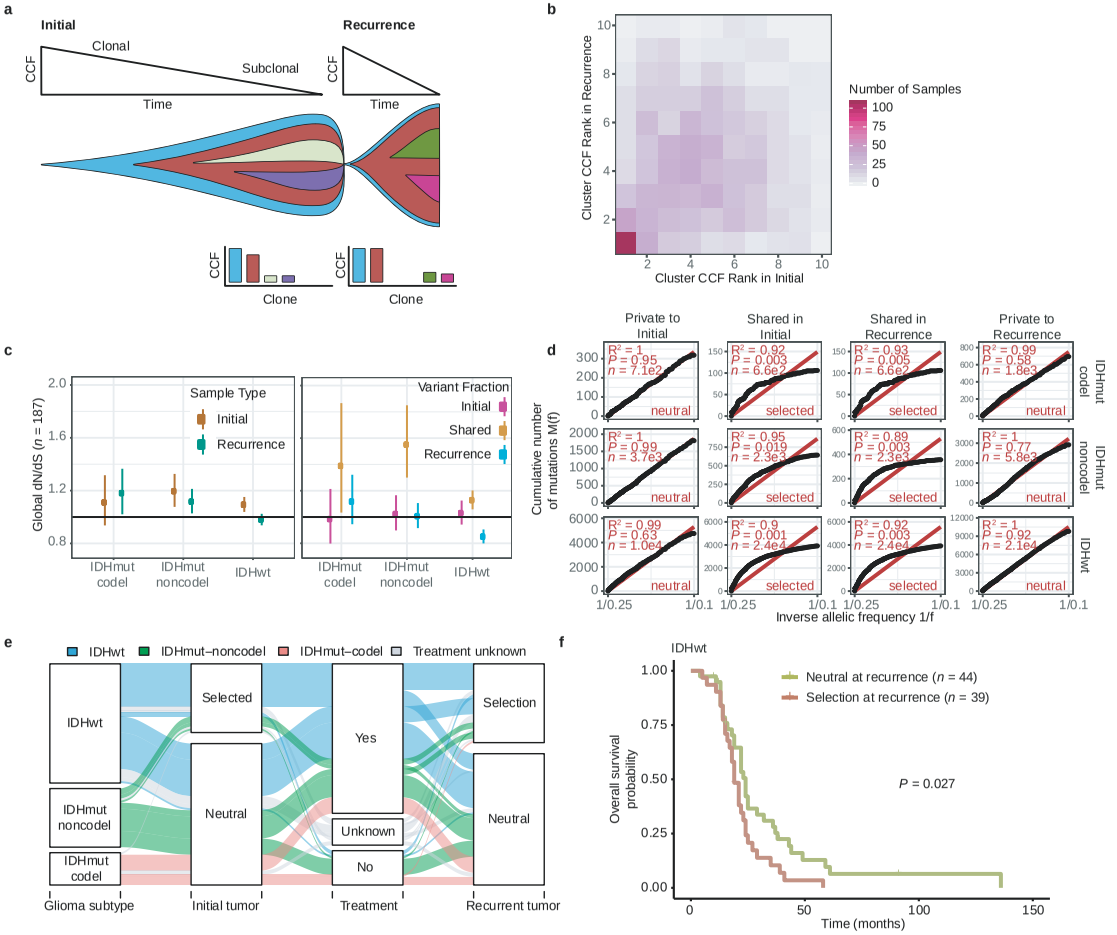
However, IDHmut-noncodel hypermutators with predominantly clonal mutations did not show differences in survival compared with those harboring predominantly subclonal mutations (log-rank test *P* = 0.38, **Figure 3-5b**). Alkylating agents such as temozolomide prolong survival of adult patients with glioma [148, 149]. Our results show that treatment-induced hypermutation is common across subtypes and does not associate with a reduced overall survival supporting the noted benefit of alkylating agent therapy.

Table 3-2. Cox proportional hazards analysis for alkylating agent associated hypermutation (*n* = 158 patients)

Covariate	Levels (*n* = patients)	HR (95% CI)	*P*-value
Age at first Diagnosis			
	Years	1.04 (1.02-1.06)	**4.61E-06**
IDH-codel subtype			
	IDHmut-codel (*n* = 8)	(ref)	
	IDHmut-noncodel (*n* = 32)	7.3 (1.71-31.36)	**7.31E-03**
	IDHwt (*n* = 98)	15.3 (3.65-64.15)	**1.91E-04**
Hypermutation			
	No hypermutation (*n* = 105)	(ref)	
	Hypermutation (*n* = 33)	1.03 (0.66-1.61)	6.58E-01

Table 3-3. Cox proportional hazards analysis for alkylating agent associated hypermutation and MGMT methylation (*n* = 65 patients)

Covariate	Levels (*n* = patients)	HR (95% CI)	*P*-value
Age at first Diagnosis			
	Years	1.05 (1.03-1.08)	**2.75E-05**
IDH-codel subtype			
	IDHmut-codel (*n* = 0)		
	IDHmut-noncodel (*n* = 13)	(ref)	
	IDHwt (*n* = 52)	1.95 (0.82-4.62)	1.31E-01
Hypermutation			
	No hypermutation (*n* = 50)	(ref)	
	Hypermutation (*n* = 15)	0.92 (0.43-2.00)	6.58E-01
MGMT			
	Methylated (*n* = 31)	(ref)	
	Unmethylated (*n* = 34)	3.10 (1.66-5.78)	**3.94E-04**


a
Initial
Recurrence
Clonal
Subclonal
CCF
Time
Clone
b
Cluster CCF Rank in Recurrence
Cluster CCF Rank in Initial
Number of Samples
c
Global dN/dS (n = 187)
Sample Type
Initial
Recurrence
Variant Fraction
Shared
IDHmut codel
IDHmut noncodel
IDHwt
d
Private to Initial
Shared in Initial
Shared in Recurrence
Private to Recurrence
Cumulative number of mutations M(f)
neutral
selected
Inverse allelic frequency 1/f
e
IDHwt
IDHmut−noncodel
IDHmut−codel
Treatment unknown
Selected
Neutral
Yes
Unknown
No
Selection
Glioma subtype
Initial tumor
Treatment
Recurrent tumor
f
Neutral at recurrence (n = 44)
Selection at recurrence (n = 39)
P = 0.027
Overall survival probability
Time (months)

← **Figure 3-6. Quantifying selective pressures during glioma evolution.**

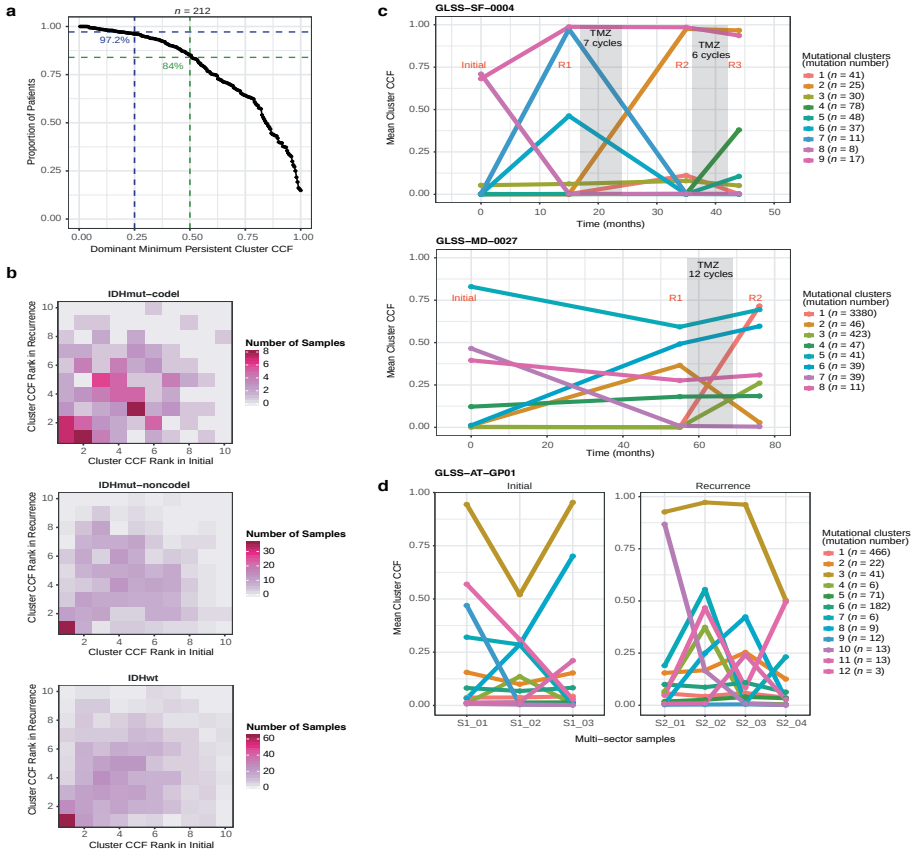
a. Schematic depiction of cancer cell fraction (CCF) values during tumor evolution indicating clonality and associated relative timing. **b.** Comparison of PyClone clusters ranked by CCF in matched initial and recurrent tumors. **c**. *Left*: dN/dS ratio for all variants (i.e. global) in initial and recurrent tumors for each subtype. Hypermutators were not included (*n* = 187). Dots represent the global dN/dS ratio with associated Wald confidence intervals. *Right*: global dN/dS ratios for variant fractions per subtype. **d.** Cumulative distribution of subclonal mutations by their inverse variant allele frequency. Mutations were separated by timepoint, variant fraction, and glioma subtype. Deviation from a linear relationship, significant Kolmogorov-Smirnov P-values and R^2 below 0.98 indicate selection. **e.** Sankey plot indicating the breakdown of SubClonalSelection evolutionary modes by subtype and therapy (*n* = 104). The sizes of the bands reflect sample sizes and band colors highlight the glioma subtype. Gray coloring reflects instances when treatment information was not available. **f.** Kaplan-Meier curve showing survival differences between IDHwt recurrent tumors demonstrating selection (*n* = 39) compared with neutrally evolving tumors (*n* = 44). Log-rank P-value is indicated.

Selective pressures during glioma evolution

Environmental and treatment-induced pressures may drive changes in clonal architecture at recurrence. To evaluate selection over time we clustered copy number changes and mutations based on their cancer cell fraction (CCF). CCF values represent the fraction of cancer cells harboring a given alteration and reflect the relative timing of events, since alterations that are present in a subset of cancer cells likely occurred later than events present in all cancer cells (**Figure 3-6a**). Most tumors (84%) demonstrated a mutational cluster with CCF > 50% that persisted from the initial tumor into recurrence, likely reflecting the tumor trunk and harboring the tumor-initiating driver mutations (**Figure 3-6b**, **Figure 3-7a**) [150]. To determine changes in clonal dominance over time we ranked clusters within each sample by their CCF and found similarities in clonal architecture throughout the course of disease (Kendall rank correlation, tau = 0.20, *P* = 3.76E-24, **Figure 3-6b**, **Figure 3-7b-d**). These results suggested that the clonal structure at initial disease mostly persisted into recurrence.

To deepen our assessment of selective pressures, we evaluated selection in initial and recurrent tumors by determining the normalized ratio between non-synonymous and synonymous mutations (dNdScv). Higher ratios (> 1) suggest positive selection, and ratios less than one suggest negative selection. We found evidence for positive selection at both time points despite differences between subtypes (**Figure 3-6c**). Separating mutations into mutational fractions demonstrated that shared but not private mutations showed positive dN/dS ratios in all three glioma subtypes indicating that only shared mutations (including truncal

mutations) are likely subject to positive selection (**Figure 3-6c**). The dN/dS ratio of initial-only mutations showed that these are neither positively nor negatively selected for, while recurrence-only mutations were subject to negative selection in IDHwt.



↑ **Figure 3-7. Clonal structure evolution over time.**

a. The minimum cancer cell fraction of the most persistent (shared between initial and recurrence) PyClone cluster. **b.** Comparison of PyClone clusters ranked by CCF in matched initial and recurrent tumors, as **Figure 3-6b** but separated by subtype. **c-d.** Examples of cluster CCF dynamics over time in three separate samples, including **(c)** two multi-timepoint samples **(d)** and one multi-sector sample. These additional data are available in the GLASS resource, but only two time-separated samples were used throughput the manuscript to ensure clarity.

To verify the reduced selective pressure in the private mutations we used an orthogonal method to test for evidence of selection (neutralitytestr) [151]. The

method uses variant allele frequency distributions and estimated mutation rates to detect whether profiles significantly deviate from a model of neutral evolution (i.e. as depicted by a linear relationship in **Figure 3-6d**). In accordance with dNdScv results, private mutations demonstrated dynamics consistent with neutral evolution (**Figure 3-6d**). Shared subclonal mutations deviated from linearity and were consistent with selection both in non-hypermutators and hypermutators (**Figure 3-6d**, **Figure 3-8a-b**), providing additional evidence that the strongest selective forces occur early in gliomagenesis.

Table 3-4. Breakdown of evolution mode (evolution at primary - evolution at recurrence) separated by subtype (*n* = 104 patients). N = neutral, S = selection.

	Glioma subtype		
Evolution mode	**IDHmut-codel (*n* = 16)**	**IDHmut-noncodel (*n* = 29)**	**IDHwt (*n* = 59)**
N-N	14	19	21
N-S	1	3	12
S-N	0	2	9
S-S	1	5	17

Table 3-5. Breakdown of selection status at recurrence by subtype (*n* = 145 patients).

	Glioma subtype		
Selection at recurrence	**IDHmut-codel (*n* = 19)**	**IDHmut-noncodel (*n* = 43)**	**IDHwt (*n* = 83)**
Neutral	16	32	44
Selection	3	11	39

Cohort-level analysis of selection masks the heterogeneity that exists in individual evolutionary trajectories. To determine the selective effects at each tumor time point we used a Bayesian framework (SubClonalSelection) which simultaneously provides sample-specific probabilities for both selection and neutrality while modeling sources of noise in sequencing data. The classification of a sample as "selection" or "neutral" is determined by whichever model has the greater probability. Classification as "neutral" reflects the accumulation of random mutations that are not subject to selection. Given the stringent algorithm requirements, 183 patients were included in this analysis with at least one time point, and 104 patients with both time points (16 IDHmutant-codels, 29 IDHmutant-noncodels, 59 IDHwt, **Table 3-4**). Neutral to neutral was the most common evolutionary trajectory across all three subtypes (52%), and IDHwt tumors displayed the highest

observed selection at any time point with selection detected in 64% of tumors (Fisher's exact test *P* = 0.01, **Figure 3-6e, Table 3-5**). IDHwt gliomas with evidence for selection at recurrence had a shorter overall survival than IDHwt gliomas classified as neutral at recurrence (*P* = 2.7E-02; log-rank statistic, **Figure 3-6f**), suggesting that subclonal competition associates with more aggressive tumor behavior. To address the limitations of smaller sample sizes in the IDH-mutant subtypes, we performed a Cox proportional hazards model including age at first diagnosis, all three glioma subtypes, and mode of selection at recurrence. This analysis revealed that selection at recurrence was significantly associated with shorter survival across subtypes (HR = 1.53 95% CI 1.00-2.41, *P* = 4.8E-02, **Table 3-6**). We next investigated whether radiation and chemotherapy imposed a selective effect, by comparing the evolutionary status at recurrence with treatment and other clinical variables. We did not observe significant associations between subclonal selection and radiation therapy or chemotherapy (Fisher's exact-test *P* > 0.05, **Table 3-7**), suggesting that standard therapeutic approaches for glioma have limited impact on the subclonal tumor architecture. While high-depth sequencing datasets may be required to detect subtle selective effects [146], our analyses raise the possibility that the survival benefit derived from standard chemoradiation results from tumor cell elimination where treatment sensitivity of individual cells is not determined by genetic factors.

Table 3-6. Multivariate Cox proportional hazards model testing the association between evolution status at recurrence and overall survival (*n* = 131 patients).

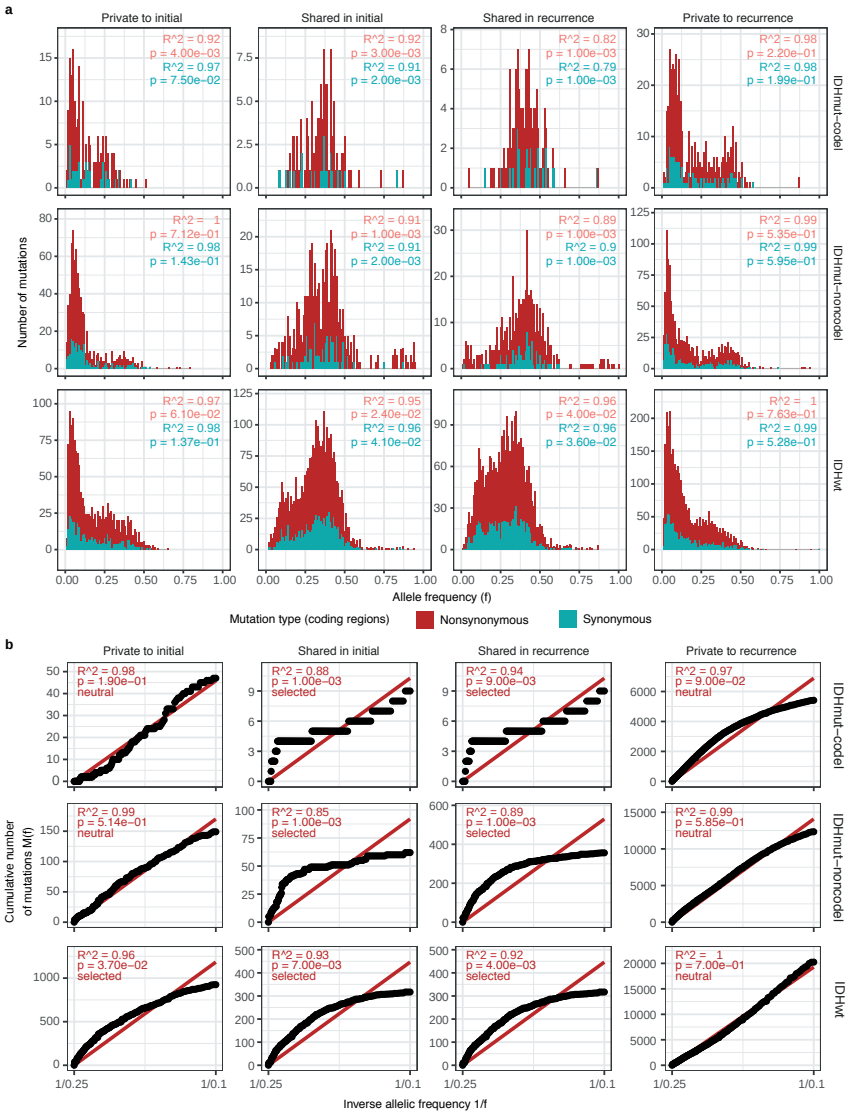
Covariate	Levels	HR (95% CI)	*P*-value
Age at first Diagnosis			
	Years	1.04 (1.02-1.06)	5.58E-05
IDH-codel subtype			
	IDHmut-codel (*n* = 18)	(ref)	
	IDHmut-noncodel (*n* = 43)	6.2 (2.1-18.2)	9.83E-04
	IDHwt (*n* = 70)	25.7 (8.5-77.9)	9.26E-09
Evolution status at recurrence			
	Neutral (*n* = 87)	(ref)	
	Selection (*n* = 44)	1.55 (1.0-2.4)	4.90E-02

Table 3-7. Relationship between selection status at recurrence and therapy. P-values were calculated using a two-sided Fisher's exact test. Patients with available SubClonalSelection and known therapy information and sample size is indicated.

IDHmut-codel							
Alkylating-agent (*n* = 15 patients)				**Radiation therapy (*n* = 18 patients)**			
	No	**Yes**	**P-value**		**No**	**Yes**	**P-value**
Neutral	8	5		**Neutral**	11	5	
Selection	2	0	0.52	**Selection**	1	1	1.0
IDHmut-noncodel							
Alkylating-agent (*n* = 39 patients)				**Radiation therapy (*n* = 43 patients)**			
	No	**Yes**	**P-value**		**No**	**Yes**	**P-value**
Neutral	14	15		**Neutral**	20	12	
Selection	3	7	0.46	**Selection**	8	3	0.72
IDHwt							
Alkylating-agent (*n* = 65 patients)				**Radiation therapy (*n* = 69 patients)**			
	No	**Yes**	**P-value**		**No**	**Yes**	**P-value**
Neutral	6	29		**Neutral**	4	33	
Selection	4	26	0.74	**Selection**	2	30	0.68

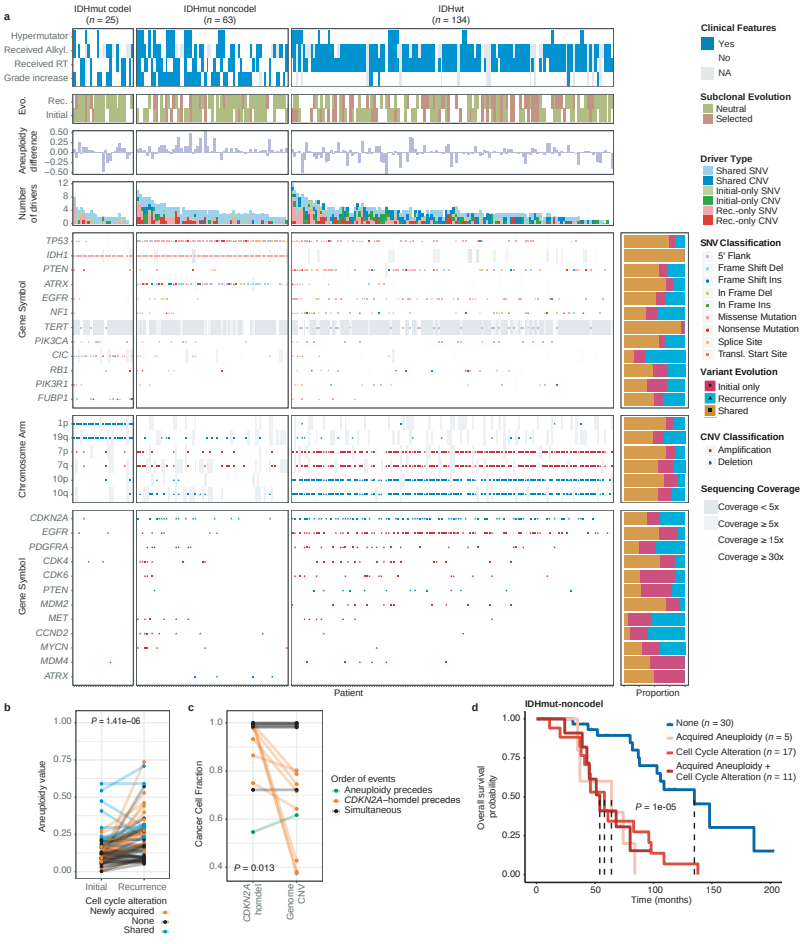
Driver alteration frequencies across time

We evaluated how stability, acquisition, and loss of mutation and copy number drivers [15] over time impact glioma evolution. We used dNdScv to nominate 12 candidate mutation driver genes at both time points ($Q < 0.05$, **Figure 3-9**, **Figure 3-10a**) and determined significant copy number alterations that recapitulated previously identified drivers (**Figure 3-10b**). Mutations in *IDH1* and co-occurring 1p/19q chromosome-arm loss have been suggested as glioma-initiating events [128], which was corroborated by the observation that these events were never lost or acquired during the surgical interval (**Figure 3-9a**, **Figure 3-11a**). Similarly, we observed that *TERT* promoter mutations were almost always shared in the IDHmutant-codel and IDHwt, though many samples lacked sufficient coverage in this GC-rich region. Chromosome 7 gains and chromosome 10 losses were present in a large majority of IDHwt initial tumors and persisted into recurrence.



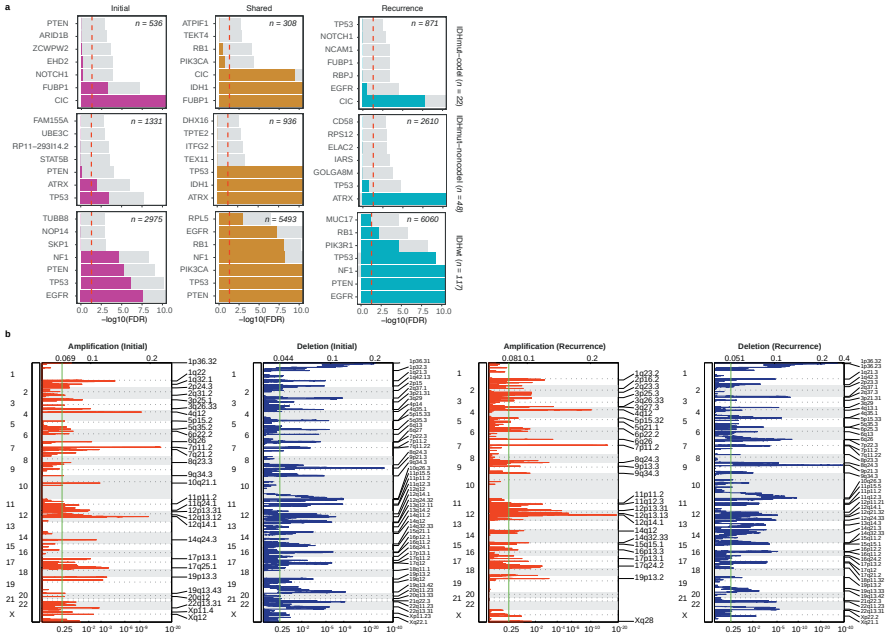
↑ **Figure 3-8. Variant allele fraction distribution.**

a. Non-hypermutator variant allele fraction distributions for copy neutral variants in coding regions (*n* = 181 patients). Variants are separated by subtype, fraction, and also whether the variant was non-synonymous or synonymous mutation in a coding region. R^2 goodness-of-fit measure and associated P-values are shown for both mutation types. Note that this data considers only the coding portion of genome while **Figure 3-6d** presents both coding and non-coding. **b.** The cumulative distribution of the subclonal mutations in copy-neutral regions for hypermutators (*n* = 31 patients). For each variant fraction and subtype, the R^2 goodness-of-fit measure and P-values are shown.



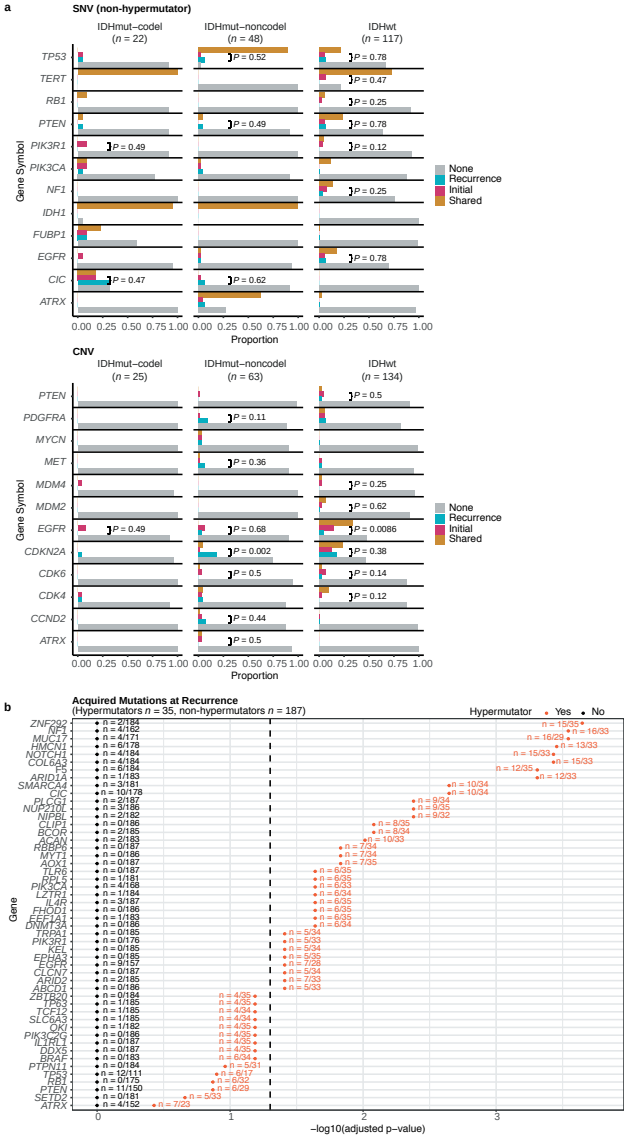
↑ **Figure 3-9. Patterns of glioma driver frequencies over time.**

a. Driver dynamics for SNVs nominated by the dNdScv and CNVs nominated by GISTIC (*n* = 222). Each column represents a single patient at two separate time points stratified by subtype and ordered left-to-right by the number of driver alterations. The degree of aneuploidy difference (recurrence – initial) offers a summary metric for increases (> 0) or decreases (< 0) in aneuploidy at recurrence. Variants are marked and different shapes indicate whether a variant was shared or private. The variant type is depicted by its color. Stacked bar plots accompanying each gene/arm provide cohort-level proportions for whether the alteration was shared, lost, or acquired. **b.** Aneuploidy comparison in matching initial and recurrent IDHmut-noncodel tumors. **c.** Within-sample CCF comparison of *CDKN2A* homozygous deletion (homdel) to genome-wide CCF as a proxy for aneuploidy. A relative higher CCF indicates temporal precedence. Wilcoxon signed-rank test P-value is indicated. **d.** Kaplan-Meier curve comparing survival in IDHmut-noncodel tumors with an alteration in the cell cycle, acquired aneuploidy, or both (shades of red) versus unaltered IDHmut-noncodel tumors (blue). Log-rank P-value is shown.



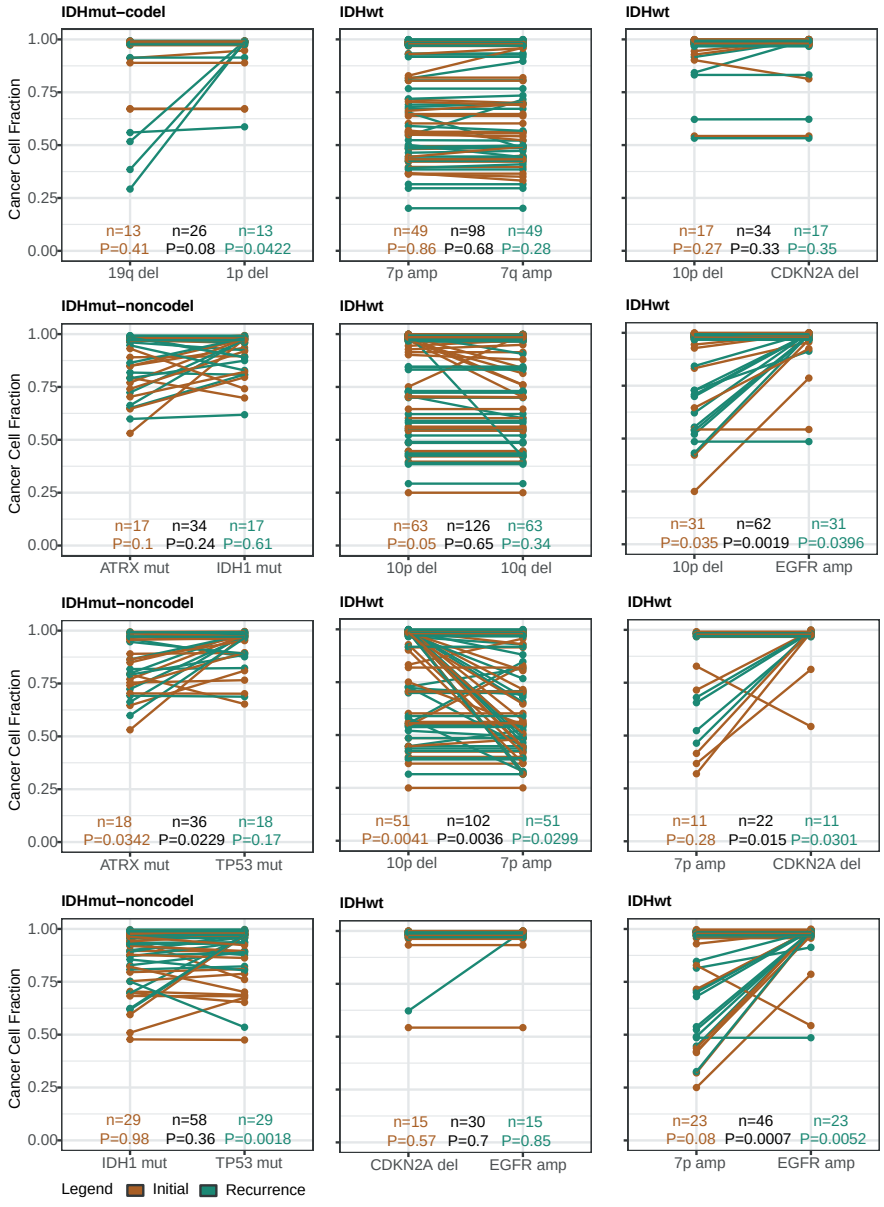
↑ **Figure 3-10. Driver gene nomination.**

a. Local (gene-wise) dNdScv estimates by subtype (rows) and fraction (columns). Genes are sorted by Q-value and P-value. The Q-value is shown in color, whereas the P-value is indicated in light gray. The Q-value threshold of 0.05 is indicated by a horizontal red line. **b.** GISTIC significant amplification (red) and deletion (blue) plots in initial (left) and recurrent tumors (right). Chromosomal locations are ordered on the y-axis, Q-values are shown on the x-axis, and selected drivers are indicated by their chromosomal location on the right.



↑ **Figure 3-11. Driver acquisition over time.**

a. Tabulated numbers of SNV (top) and CNV (bottom) driver events that were shared, initial-only, or recurrence-only. P-values were obtained using a two-sided Fisher test comparing the initial-only fraction to the recurrence-only fraction testing for acquisition. **b.** One-sided Fisher test comparing the initial-only fraction to the recurrence-only fraction amongst previously implicated glioma drivers testing for driver acquisition. P-values were adjusted for multiple testing using the FDR (x-axis). Hypermutators (red) and non-hypermutators (black) were separately analyzed.

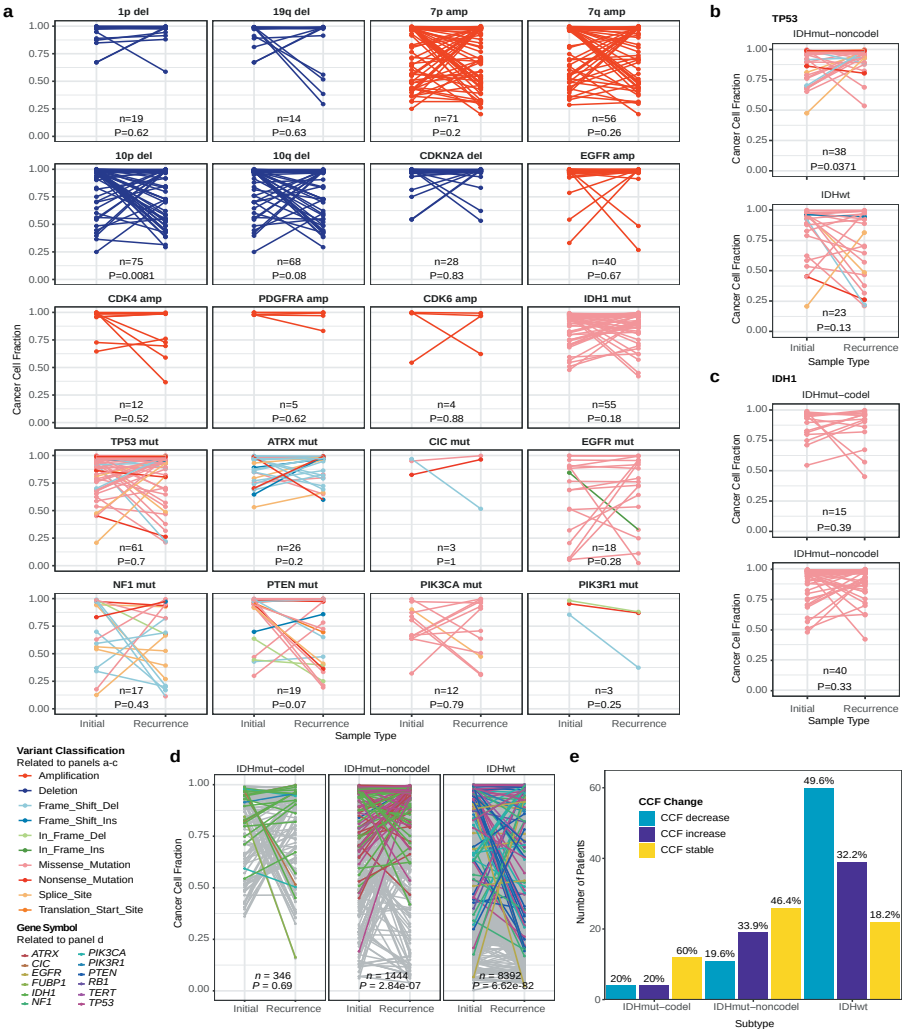


↑ **Figure 3-12. Intra-tumor CCF comparison.**

Ladder plots comparing the CCF of co-occurring drivers in single tumor samples. The color of the lines and points indicates whether the sample shown is an initial (brown) or recurrent (green) tumor. Two-sided Wilcoxon rank-sum test P-values are shown for all initial samples, all recurrent samples, as well as all samples (black).

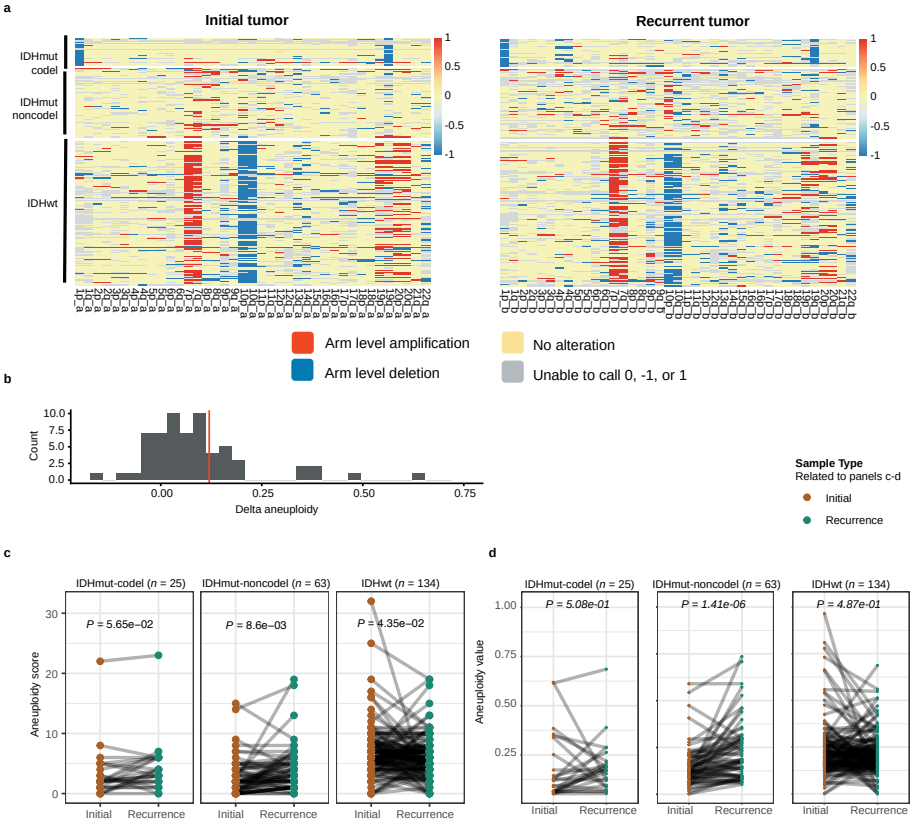
Shifts in the fraction of cancer cells harboring an event may also indicate a time dependency of drivers. We determined changes in cellular prevalence of shared driver events by ordering events in each sample by their CCF (**Figure 3-12**). *ATRX* mutations in IDHmutant-noncodel initial tumors demonstrated lower CCFs than *TP53* ($P = 0.03$) and *IDH1* ($P = 0.10$) mutations, suggesting *IDH1* and *TP53* mutations precede *ATRX* inactivation [128]. There was no difference in CCF between *IDH1* and *TP53* amongst initial gliomas ($P = 0.98$), however, *IDH1* mutations demonstrated significantly lower CCFs compared with *TP53* ($P = 0.0018$) in recurrent gliomas. We did not observe any CCF differences among driver mutations detected in IDHwt tumors at either time point. Chromosome 10 deletion CCFs were higher compared to chromosome 7 amplifications ($P = 0.0036$) implying that chromosome 10 deletions arise earlier [152]. Similarly, there was no difference in CCF between *CDKN2A* deletion and *EGFR* amplification ($P = 0.70$). *EGFR* and chromosomal arm events significantly differed (i.e. 10p del vs *EGFR* amp, P = 0.0019) but not *CDKN2A* deletion and chromosomal events (i.e. 10p del vs CDKN2A del, P = 0.33). The consistently high CCF for *EGFR* amplifications could indicate that these events precede even some larger chromosomal aberrations, while not excluding the possibility that high levels of extrachromosomal *EGFR* [153] artificially inflate CCF.

Longitudinal changes in CCF values provide additional insights into evolutionary dynamics. For instance, the CCF value may increase when a driver event is linked to clonal expansion, or conversely, decrease when a clone is outcompeted. Most individual drivers did not demonstrate significant consistent CCF changes between the initial tumor and recurrence (**Figure 3-13a**). A notable exception was the *TP53* mutation CCF that increased over time ($P = 0.037$) in IDHmut-noncodels, but not IDHwt gliomas ($P = 0.13$, **Figure 3-13b**). We did not observe any differences in *IDH1* CCF over time among IDHmut-noncodel tumors, possibly because the general trend of these tumors to increase in CCF is counteracted by the biological loss of relevance of mutant *IDH1* over time (**Figure 3-13c**). Indeed, a gross comparison of all shared mutation CCFs revealed an increase in recurrent IDHmut-noncodel tumors ($P < 0.0001$), which may reflect increased clonality and a reduction in intratumoral heterogeneity (**Figure 3-13d**). In contrast, shared CCFs decreased in IDHwt tumors, potentially indicating a general increase in intratumoral heterogeneity at recurrence ($P < 0.0001$, **Figure 3-13d**). We confirmed that IDHmutant-noncodel CCF increases and IDHwt decreases were not biased by patients with high mutation burden through the classification of patient-specific shared mutation CCF change (**Figure 3-13e**).



↑ **Figure 3-13. Between time point intra-patient CCF comparison.**

a. Driver-gene CCF comparison between initial and matched recurrences. Lines are colored by variant classification. Two-sided Wilcoxon rank-sum test P-values are shown. **b.** *TP53* CCF by subtype, otherwise as in **(a)**. **c.** *IDH1* CCF by subtype, otherwise as in **(a)**. **d.** Ladder plot visualizing CCF change across all SNVs between initial and recurrent tumors, separated by subtype. Wilcoxon rank-sum test was used to test for differences between time points. **e.** Initial and recurrent mutations in each patient were compared using a Wilcoxon rank-sum test. Bar plot with counts of patients in each subtype are shown. Patients lacking significant change are shown in yellow, those with a significant increase or decrease are shown in dark and light blue, respectively.



↑ **Figure 3-14. Aneuploidy calculation.**

a. Heatmap displaying the chromosomal arm-level events (x-axis) with patients represented in each row. Patients are placed in the same order for both the initial (left) and recurrence (right). White space was inserted as a break between the three subtypes. **b.** Distribution of total aneuploidy difference. Acquired aneuploidy determination (upper-quartile) indicated with a red line. **c.** Comparison of aneuploidy score between initial and recurrent tumors separated by subtype **d.** As **(c)**, comparing aneuploidy value.

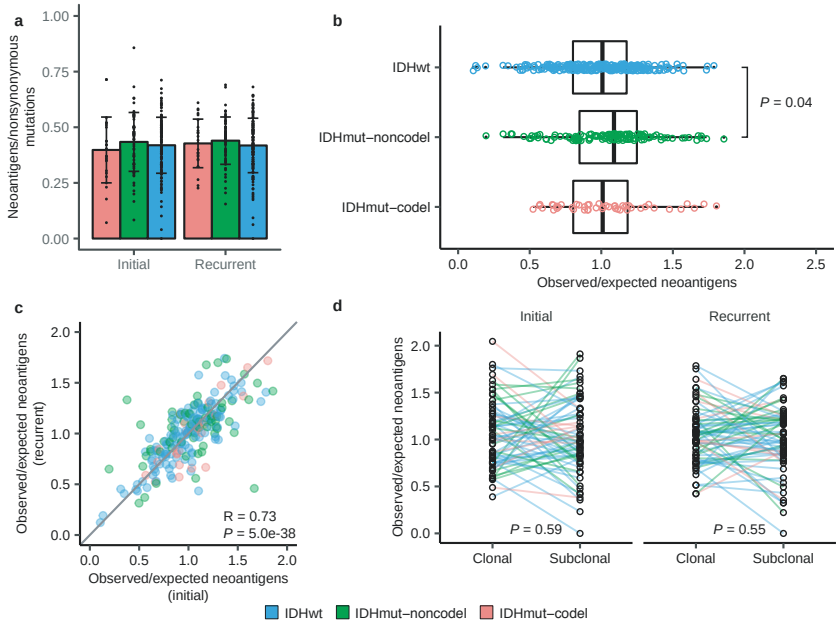
We next investigated whether specific somatic alterations were acquired or lost over time. Gene-specific enrichment of many recurrence-only mutations was found in hypermutated tumors, but there was no enrichment for somatic gene alterations in non-hypermutators suggesting that glioma recurrence is not directed by particular sets of mutations (**Figure 3-11b**). Within subtypes we detected an enrichment in *CDKN2A* homozygous deletions (**Figure 3-9a**, **Figure 3-11a**) in recurrent IDHmutant-noncodels, which was corroborated by additional cell

cycle gene alterations (focal gain of *CCND2, CDK4*, *CDK6*, and mutation or homozygous loss of *RB1*). Mutations in cell cycle checkpoint control genes are associated with genomic instability [154]. Therefore, we analyzed aneuploidy levels by determining the proportion of the genome that had undergone aneuploidy events (**Figure 3-14a-b).** We observed that IDHmutant-noncodel tumors had a higher level of aneuploidy at recurrence (Wilcoxon rank sum test P = 1.4E-06 total aneuploidy, p = 8.6E-03 arm-level aneuploidy, **Figure 3-14c-d**) with tumors carrying acquired cell cycle gene alterations displaying the largest increases in aneuploidy (P = 7.6E-06; Wilcoxon rank sum test, **Figure 3-9b**). We reasoned that *CDKN2A* deletions may precede aneuploidy. Homozygous *CDKN2A* deletions had significantly higher CCFs compared to average CNV CCF across the genome (as a surrogate for aneuploidy related copy number changes), suggesting that *CDKN2A* loss occurred prior to aneuploidy (**Figure 3-9c**). These alterations may hasten disease progression as patients with either cell cycle alterations or the largest increases in aneuploidy at recurrence demonstrated significantly shorter survival than patients without these alterations (log-rank test $P < 0.0001$, **Figure 3-9d**). Taken together, the persistence of drivers over time and the paucity of consistent change imply that therapy does not result in selection of specific sets of molecular changes.

Immunoediting activity in glioma

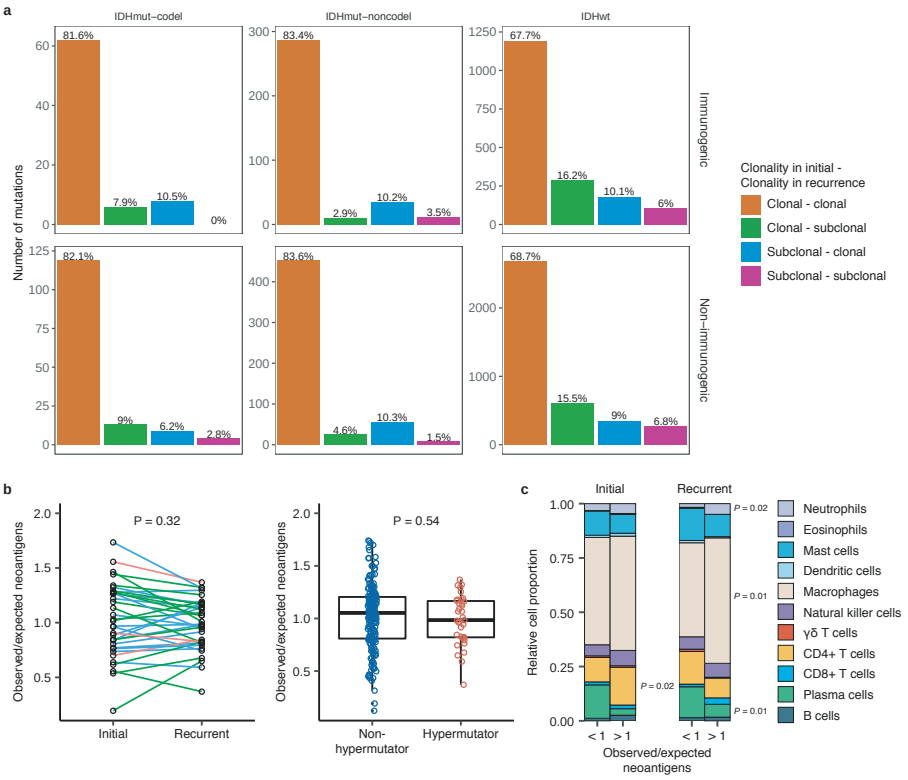
We next investigated how the immune microenvironment affects evolutionary trajectories. The immune system may prune tumor cells carrying immunogenic (neo-)antigens, resulting in the selection of subclones capable of evading the immune response. Evidence of this immunoediting process has been shown in cancer types, including glioma [74, 155-157], and suggests active immunosurveillance that may be therapeutically exploited [158]. We computationally predicted neoantigen-causing mutations [159]. As expected, the neoantigen load across the GLASS cohort was strongly correlated with exonic mutation burden (Spearman's Rho = 0.89), with 42% of nonsynonymous exonic mutations giving rise to neoantigens on average. This fraction did not significantly differ by glioma subtype or between initial and recurrent tumors ($P > 0.05$, Wilcoxon rank-sum test; **Figure 3-15a**). The most common neoantigen arose from the clonal R132H mutation in *IDH1* and was present in of 22 out of 88 IDH-mutant initial and recurrent tumors. Beyond mutations in *IDH1*, no mutations gave rise to a neoantigen found in more than three tumors at a given timepoint (***Supplementary Table 6**, available online with the full-text of this article at www.nature.com*). Across the

dataset, neoantigens and non-immunogenic mutations exhibited similar changes in cancer cell fractions between initial and recurrent tumors indicating a lack of neoantigen-specific selection processes over time (**Figure 3-16a**).



↑ **Figure 3-15. Neoantigen selection during tumor progression.**

a. Mean proportion of coding mutations giving rise to neoantigens (neoantigens/nonsynonymous) stratified by glioma subtype and timepoint (*n* = 222). Error bars represent standard deviation. **b.** Boxplot depicting the distribution of observed to expected neoantigen ratios in the GLASS cohort stratified by glioma subtype. P-value was calculated using the Wilcoxon rank-sum test. Each box spans quartiles, with the lines representing the median ratio for each group. Whiskers represent absolute range, excluding outliers. **c.** Scatterplot depicting the association between the observed-to-expected neoantigen ratio in a patient's initial versus recurrent tumor. Each point represents a single patient. R represents Pearson correlation coefficient. Panels b and c only include samples with at least 3 neoantigens in the initial and recurrent tumors (*n* = 131, 63, and 24 for IDHwt, IDHmut-noncodel, and IDHmut-codel, respectively). **d.** Ladder plot depicting the difference in observed-to-expected neoantigen ratio between a tumor's clonal and subclonal neoantigens. Each set of points connected by a line represents one tumor. Tumors are stratified by whether they were a patient's initial or recurrent tumor. Lines are colored by each patient's glioma subtype. Panel d only includes samples with at least 3 clonal neoantigens and at least 3 subclonal neoantigens in both the initial and recurrent tumors (*n* = 35, 20 and 9 for IDHwt, IDHmut-noncodel, and IDHmut-codel, respectively). P-value was calculated using a paired two-sided t-test. Colors in each panel represent the glioma subtype and are denoted at the bottom of the figure.



↑ **Figure 3-16. Neoantigen evolution and cellular analysis.**

a. Bar plots representing the number of shared mutations that give rise to neoantigens (top row, "immunogenic") and those that do not give rise to neoantigens (bottom row, "non-immunogenic") stratified by longitudinal clonality ("(clonality in initial)-(clonality in recurrence)") and further separated by subtype. Percentage of longitudinal clonality per subtype and mutation immunogenicity are presented above the respective bars. **b.** *Left:* Ladder plot depicting the difference in observed-to-expected neoantigen ratio between the initial and recurrent tumors of patients with hypermutated tumors at recurrence. Each set of points connected by a line represents one tumor ($n = 70$). *Right:* Boxplot depicting the distribution of observed to expected neoantigen ratios in recurrent tumors stratified by hypermutator status ($n = 35$ and 183 for hypermutators and non-hypermutators, respectively). Each box spans quartiles, with the lines representing the median ratio for each group. Whiskers represent absolute range, excluding outliers. P-values for panel b were calculated using a paired and unpaired two-sided t-test, respectively. **c.** Stacked bar plots depicting the average relative fraction of 11 CIBERSORT cell types in the neoantigen depleted (< 1) and non-depleted (> 1) initial and recurrent tumor subgroups. Asterisks to the right of each plot indicate a significant difference ($P < 0.05$, Wilcoxon rank-sum test) between the depleted and non-depleted groups for the noted cell type at that time.

We then examined the extent to which immunoediting occurred by comparing each sample's observed neoantigen rate to an expected rate that was empirically derived from our dataset. The output of this approach is a normally distributed set of ratios centered at 1. Samples with an observed-to-expected neoantigen ratio < 1 exhibit evidence of neoantigen depletion relative to the rest of the dataset, and thus are more likely to have been immunoedited. We found that none of the three glioma subtypes harbored observed-to-expected ratios that significantly differed from 1 ($P > 0.05$, one sample t-test), though IDHwt tumors exhibited significantly lower scores compared to IDHmut-noncodels (t-test, $P = 0.04$; **Figure 3-15b**). We additionally did not observe an association between the observed-to-expected ratio and survival when adjusting for subtype and age (Wald test, $P > 0.05$), nor was there a difference between samples with neutral evolution dynamics compared to those exhibiting evidence of subclonal selection. When comparing samples longitudinally, we found that the observed-to-expected neoantigen ratio was strongly correlated between initial and recurrent tumors of each patient (Pearson's R = 0.73, $P = 5E-38$), suggesting that the neoantigen depletion level in the recurrence reflects that of the initial tumor (**Figure 3-15c**).

Immunoediting is most likely to take place in the tumors with high cytolytic activity and low levels of immunosuppressive activity [157]. Hypermutators, which have high neoantigen loads, have previously been associated with highly cytolytic microenvironments [74]. However, we did not observe any differences in the observed-to-expected neoantigen ratio between hypermutated recurrent tumors and their initial counterparts, nor did we observe differences between hypermutated and non-hypermutated recurrent tumors, indicating that immunoediting activity is not related to the total number of mutations in a sample (Wilcoxon rank-sum test $P > 0.05$; **Figure 3-16b**). To more directly determine whether there were immunologic factors associated with neoantigen depletion, we analyzed CIBERSORT immune cell fractions from a subset of samples that had undergone expression profiling in a previous study ($n = 84$ from 42 tumor pairs) [74, 160]. Initial tumors with an observed-to-expected neoantigen ratio >1 exhibited significantly higher levels of CD4+ T cells than those with a ratio < 1, while recurrent tumors with a ratio > 1 exhibited significantly higher levels of macrophages, neutrophils, and significantly lower levels of plasma cells relative to those with ratio < 1 ($P < 0.05$, Wilcoxon rank-sum test; **Figure 3-16c**).

While we did not detect many factors associated with the observed-to-expected neoantigen ratio, we did observe that the ratio was significantly associated with the total number of unique HLA loci in a patient (Spearman's Rho = 0.28,

P = 2E-9), reflecting similar findings in lung cancer [161]. This may bias analyses comparing the ratio across patients. To determine whether immunoediting varies over time in a patient-agnostic manner, we compared the observed-to-expected neoantigen ratio derived from a sample's clonal mutations, which likely arose earlier in tumor evolution, to that derived from their subclonal mutations, which likely arose later. We did not observe a significant difference in the observed-to-expected neoantigen ratio of each patient's clonal and subclonal neoantigens, regardless of glioma subtype or whether the sample was an initial tumor or recurrence (*P* > 0.05, paired t-test; **Figure 3-15d**). Together, these analyses suggest that neoantigens in glioma are not exposed to differing levels of selective pressure throughout their development.

Discussion

We reconstructed the evolutionary trajectories of 222 patients with glioma to better understand treatment failures and tumor progression. The longitudinal molecular profiles revealed common features such as acquired hypermutation and aneuploidy, but highlighted the individualistic paths of post-treatment glioma evolution. Our results provide evidence that current standard of care therapies do not frequently coerce glioma down predictable paths. Instead, an unexpected number of gliomas appeared to stochastically evolve following early driver events. We expect that continuing to profile patient tumors over time using comprehensive sequencing approaches will identify additional common evolutionary paths. Our results here highlight the exciting prospects of several ongoing efforts that may inform new glioma therapies.

The observation that treatment-induced hypermutation occurred across subtypes, but did not confer a detrimental effect on patient survival leaves the clinical significance of glioma hypermutation uncertain [73, 139, 141, 144, 147]. Future analyses that consider the number of therapy cycles and *MGMT* DNA methylation status will help to elucidate factors that predispose tumors to hypermutation and identify therapies that effectively exploit this phenotype's vulnerabilities (e.g., high mutation burden). Acquired cell cycle alterations and aneuploidy in recurrent IDHmut-noncodel gliomas also provide a rationale to target these more aggressive phenotypes with CDK inhibitors [162] or with compounds that disrupt microtubule dynamics [163]. Finally, our analyses revealed that immunoediting activity does not vary in glioma over time, though we did observe variation between individual patients. Additional molecular and immunological data are needed to fully understand the impact this variability has on glioma evolution

and to devise therapies directed at a glioma's immunogenicity [135]. To this end, we found that clonal neoantigens arising from the *IDH1* R132H mutation persisted from the initial tumor into the recurrence, justifying neoantigen vaccine approaches as treatments for initial and recurrent glioma [164, 165].

Collectively, these findings help shape our perspective on what constitutes an optimal treatment, and what approaches would result in the greatest removal or killing of glioma cells possible. Genomic characterization efforts such as TCGA have greatly increased our understanding of glioma biology, but were limited to a single snapshot in evolutionary time. The GLASS resource provides a framework to study the patterns of glioma evolution and treatment response.

Acknowledgements

This work is dedicated to the memory of Simone Bischoff-Lardenoije and is made possible by the patients and their families whom generously contributed to this study. This work is supported by the National Brain Tumor Society, Oligo Research Fund; Cancer Center Support grants P30CA16672 and P30CA034196; Cancer Prevention & Research Institute of Texas (CPRIT) grant number R140606; Agilent Technologies (R.G.W.V.); the National Institutes of Health- National Cancer institute for the following grants: NCI CA170278 (L.M.P., M.M.T., N.H.), NCI R01CA222146 (L.M.P, N.H.), NCI R01CA230031 (J.H.C., J.N.), NCI R01CA188288 (J.S.B., R.B., P.B., K.L.L., A.C., A.E.S.), R01CA179044 (Antonio Iavarone), U54CA193313 (Antonio Iavarone). The National Brain Tumor Society (W.K.A.Y.; J.D.G). Brain Tumour Northwest tissue bank (including the Walton research tissue bank) is supported by the Sidney Driscol Neuroscience Foundation and part of the Walton Centre and Lancashire Teaching Hospitals NHS Foundation Trusts (A.B., M.D.J.). This work was supported by a generous gift from the Dabbiere family (J.F.C.). Support is also provided by a Leeds Charitable Foundation grant (9R11/14-11 to LFS), University of Leeds Academic Fellowship (11001061) (L.F.S.) and Studentship (11061191) (G.T.) as well as Leeds Teaching Hospitals NHS Trust (Aruna Chakravarti, Azzam Ismail). The Leeds Multidisciplinary Research Tissue Bank staff was funded by the PPR Foundation and The University of Leeds (S.C.S.). Funds were received from The Brain Tumour Charity (C.W., Grants 10/136 & GN-000580, B.A.W., 200450). G.T. is funded by EKFS 2015_Kolleg_14. R01CA218144 (P.S.L, E.J.C, J.C. A.K.L.) and Strain for the Brain, Milwaukee, WI (P.S.L, E.J.C, J.C. A.K.L.). E.K is recipient of an MD-Fellowship by the Boehringer Ingelheim Fonds and is supported by the German National Academic Foundation. The Leeds Multidisciplinary Research Tissue Bank staff was funded by the PPR Foundation and part of the University of Leeds (S.C.S.).

GLASS-Austria was funded by the Austrian Science Fund project KLI394 (A.W.). GLASS-Germany was funded by the German Ministry of Education and Research (BMBF) 031A425 (G.R., P.L.) and German Cancer Aid (DKH) 70-3163-Wi 3 (M.W.). GLASS-NL receives support from KWF/Dutch Cancer Society project11026 (MCMK, PW, RGWV, PJF, JMN, MS, BAW). We thank the University of Colorado Denver Central Nervous System Biorepository (D.R.O.) for providing tissue samples. Sponsoring was also received from the National Institute of Neurological Disorders and Stroke (NINDS R01NS094615, R.G.), National Health and Medical Research Council project grant (A.M.D.). F.S.V. is supported by a postdoctoral fellowship from The Jane Coffin Childs Memorial Fund for Medical Research. F.P.B. is supported by the JAX Scholar program and the National Cancer Institute (K99 CA226387); K.C.J. is the recipient of an American Cancer Society Fellowship (130984-PF-17-141-01-DMC). We thank the Jackson Laboratory Clinical and Translation Support team for coordinating all data transfer agreements. We thank Matt Wimsatt for assistance in graphic design.

Conflicts of interest

R.G.W.V. declares equity in Boundless Bio, Inc. M.K. receives research grants from BMS and ABBVie. P.K.B. is a consultant for Lilly, Genentech-Roche, Angiochem and Tesaro. P.K.B. receives institutional funding from Merck and Pfizer and honoraria from Merch and Genentech-Roche. W.K.A.Y serves in a consulting or advisory role at DNAtrix Therapeutics. M.W. receives funding from Acceleron, Actelion, Bayer, Isarna, Merck, Sharp & Dohme, Merck (EMD, Darmstadt), Novocure, OGD2, Pigur and Roche as well as honoraria from BMS, Celldex, Immunocellular Therapeutics, Isarna, Magforce, Merck, Sharp & Dohme, Merck (EMD, Darmstadt), Northwest Biotherapeutics, Novocure, Pfizer, Roche, Teva and Tocagen. G.R. receives funding from Roche and Merck (EMD, Darmstadt) as well as honoraria from AbbVie. M.S. is a central reviewer for Parexel Ltd and honoraria are paid to the institution. G.T. reports personal fees from Bristol-Myers-Squibb, personal fees from AbbVie, personal fees from Novocure, personal fees from Medac, travel grants from Bristol-Myers-Squibb, education grants from Novocure, research grants from Roche Diagnostics, research grants from Medac, membership in the National Steering board of the TIGER NIS (Novocure) and the International Steering board of the ON-TRK NIS (Bayer).

Author contributions

D.M.A., D.A., P.B., J.S.B., R.B., C.B., P.K.B., D.J.B., A.B., A.C., E.J.C., J.C., G.F., M.N.F., Antonio I., M.D.J., M.K., P.S.L., M.L., P.L., K.L.L., T.M.M., A.M.M., D.N., N.N., H.N., C.Y.N., S.P.N., Houtan N., D.R.O., C.P., L.M.P., G.R., B.R., J.K.S., S.C.S., A.E.S., M.S., L.F.S., H.S., E.G.V.M., C.W., M.W., G.W., A.W., contributed to sample acquisition and processing, sequencing data coordination was performed by H.K, F.P.B and K.C.J., and clinical data coordination by A.D.M., and O.A.. Data analysis was led by F.P.B. and K.C.J. in collaboration with S.B.A., P.B., B.C., J.H.C., H.K., E.K, T.M.M., H.N., J.N., M.S., L.F.S., G.T., F.S.V. and R.G.W.V.. Clinical analysis was performed by A.D.M., L.M.P., and C.W.. Pathology review was completed, in part, by Aruna Chakrabarty, J.T.H., Azzam Ismail., and A.W.. F.P.B., K.C.J., A.D.M., F.S.V., and R.G.W.V. wrote the manuscript. K.D.A. and J.F.D. took charge in coordinating GLASS-MDACC; L.F.S. was the lead coordinator of the GLASS-Leeds cohort and B.A.W. of GLASS-Netherlands. R.G.W.V was the project lead and coordinator. Funding for the project was received by K.D.A., E.B.C., H.G., J.T.H., S.C.S., L.F.S.. All co-authors discussed the results and commented on the manuscript and Supplementary Information.

Methods

Data reporting

No statistical methods were used to predetermine sample size.

DNA sequencing and data collection

The GLASS dataset consists of both unpublished and published sequencing data as outlined in Supplementary Table 1. Among the cohort were exomes from 436 glioma samples (200 patients), whole-genome from 165 glioma samples (78 patients), with overlapping exome/whole-genome data on 78 glioma samples (38 patients). A matching germline sequence was available for all patients. The dataset includes 257 sets of at least two time-separated tumor samples, seventeen standalone recurrences, and 19 patients with at least two geographically distinct tumor portions. More specifically, the dataset includes exome or whole-genome sequencing data on 211 primary gliomas, 234 first recurrences, 32 second recurrences, 11 third recurrences and one fourth recurrence (***Supplementary Table 7****, available online with the full-text of this article at www.nature.com*).

Newly generated whole genome sequencing data for the Chinese University of Hong Kong (HK), Northern Sydney Cancer Centre (NS) and MD Anderson Cancer

Center (MD) cohorts were subjected to 150 base paired-end sequencing. The HK samples were sequenced using a HiSeqX while the NS and MD cohorts were sequenced using a NovaSeq according to Illumina's protocols. Whole exome capture was performed using the following platforms as reported in previous publications. Agilent SureSelect Human All Exon 50Mb capture kit was used for patients SF-0001- SF-0021, Agilent SureSelect Human All Exon V4 capture kit was used for patients SF-0024 – SF-0029 in the UC San Francisco cohort. Agilent SureSelect Human All Exon v4 or v5 was used to capture samples in the Kyoto University cohort. Samsung Medical Center cohort reported using Agilent SureSelect kit for patients SM-R056 – SM-R071, SM-R075, SM-R076, SM-R095- SM-R114 while Illumina TruSeq Exome-capture kit was used for patient SM-R072. Exome capture was performed using Agilent SureSelect Human All Exon 50 Mb in The Cancer Genome Atlas (TCGA)-GBM cohort and Agilent SureSelect Human All Exon v2.0, 44Mb kit in the TCGA-LGG cohort. Columbia University cases were captured using Agilent V3 50M kit, sequencing 90bp PE for samples R009-TP, R009R1, R011TP, R011R1, R014TP, R014R1, R017-R1, R018-R1, R019-R1. Mapping files of initial tumor and normal samples of patients R017 – R019 were obtained from TCGA through CG-hub. All other samples were captured using Agilent SureSelect XT Human All Exon v4 Kit, PE, 80M reads, 150X on target coverage. Samples in the Henry Ford Hospital cohort were multiplexed and sequenced using Illumina HiSeq 2000 by the Sequencing and Microarray Facility at an average target exome coverage of 100× using 76-bp paired-end reads. Samples in the HK cohort were subjected to 75 base paired-end sequencing for HK-0001 – HK-0004 as performed NextSeq in high output mode. In the Leeds Cohort (LU) SureSelectXT V5 kit (PE100) was used to construct exome libraries. Illumina TruSeq Exome capture kit was used for samples at the Medical University of Vienna – CeMM.

GLASS identifiers

A GLASS barcode system was created, based on TCGA barcode design, in an effort to de-identify patient information and provide an organized framework for the different pieces of the dataset.

GLASS barcodes are composed of 24 characters. The first four characters specify the project (either GLSS or TCGA). All datasets submitted to the GLASS consortium, published and unpublished, were given the GLSS project ID. Samples that were part of the TCGA cohorts (TCGA GBM and TCGA LGG) were given a TCGA designation. The next two characters designate the center where the samples

were either acquired or sequenced (***Supplementary Table 7**, available online with the full-text of this article at www.nature.com*). This is followed by the four-character center specific patient identification that was kept as close as possible to the patient identification provided by the collaborators to allow a simplified trace back process. Patient data is divided by a relative sample type, such as initial tumor (TP), recurrent tumor (R1), normal tissue (NB, NM, etc), or metastatic tumor sample (M1). If there was more than one recurrence the relative number was specified following "R". Some patients had surgeries for which a biospecimen was unavailable. Thus, a surgical number was also provided to indicate temporal ordering (***Supplementary Table 8**, available online with the full-text of this article at www.nature.com*). To include spatially separated samples the portion designation was added, which is followed by one character specifying the type of analyte, either DNA (D) or RNA (R). As there is variation in the sequencing analysis, a three-character designation represents either whole genome (WGS) or whole exome sequencing (WXS). The last part of the GLASS barcode is a six-character designation unique to each barcode that was randomly generated.

Computational pipelines

All pipelines were developed using snakemake 5.2.2 [166]. Unless otherwise stated, all tools mentioned are part of the GATK 4 suite [167]. All data was collected at a central location (The Jackson Laboratory) and was analyzed using homogenous pipelines capable of processing both raw fastq files as well as re-process previously analyzed bam files.

Alignment and pre-processing

Data pre-processing was conducted in accordance to the GATK Best Practices using GATK 4.0.10.1. Briefly, aligned BAM files were separated by read group, sanitized and stripped of alignments and attributes using 'RevertSam', giving one unaligned BAM (uBAM) file per readgroup. Uniform readgroups were assigned to uBAM files using 'AddOrReplaceReadgroups'. Similarly, unaligned fastq files were assigned uniformly designated readgroup attributes and converted to uBAM format using 'FastqToSam'. uBAM files underwent quality control using 'FastQC 0.11.7'. Sequencing adapters were marked using 'MarkIlluminaAdapters'. uBAM files were finally reverted to interleaved fastq format using 'SamToFastq', aligned to the b37 genome ('human_g1k_v37_decoy') using 'BWA MEM 0.7.17', attributes were restored using 'MergeBamAlignment'. 'MarkDuplicates' was then used to merge aligned BAM files from multiple readgroups and to mark PCR and opti-

cal duplicates across identical sequencing libraries. Lastly, base recalibration was performed using 'BaseRecalibrator' followed by 'ApplyBQSR'. Coverage statistics were gathered using 'CollectWgsMetrics'. Alignment QC was performed running 'ValidateSamFile' on the final BAM file and QC results were inspected using 'MultiQC 1.6a0' [168]. A haplotype database for fingerprinting was generated using a modified version of the code on https://github.com/naumanjaved/fingerprint_maps. The tool 'CrosscheckFingerprints' was used to confirm that all readgroups within a sample belong to the same individual, and that all samples from one individual match. Any mismatches were marked and excluded from further analysis.

Variant detection

Variant detection was performed in accordance to the GATK Best practices using GATK 4.1.0.0. Germline variants were called from control samples using Mutect2 in artifact detection mode and pooled into a cohort-wide panel of normals. Somatic variants were subsequently called in individual tumor samples (single-sample mode) and in entire patients using GATK 4.1 Mutect2 in multi-sample mode. Mutect2 was given matched control samples, the aforementioned panel of normals and the gnomAD germline resource as additional controls. Cross-sample contamination was evaluated using 'GetPileupSummaries' and 'CalculateContamination' run for both tumor and matching control samples. Read orientation artifacts were evaluated using 'CollectF1R2Counts' and 'LearnReadOrientationModel'. Somatic likelihood, read orientation, sequence context, germline and contamination filters were applied using 'FilterMutectCalls'.

Variant post-processing

BCFTools 1.9 was used to normalize, sort and index variants [169]. A consensus VCF was generated from all variants in the cohort, removing any duplicate variants. The consensus VCF file was annotated using GATK 4.1 Funcotator and the v1.6.20190124s annotation data source. Allele frequencies (AFs) from multi-sample Mutect2 were used to compare AFs between related samples. Multi-sample Mutect2 calls and filters mutations across a patient as a whole and does not determine mutation calls in a single sample. Single-sample mutation calls were overlaid on the multi-sample calls to infer whether variants were called in individual samples. Single-sample called variants that were not present in the multi-sample callset were discarded.

Mutational burden

Mutational burden was calculated as the number of mutations per megabase (Mb) sequenced. A minimum coverage threshold of 15x was required for each base. DNA hypermutation was defined for recurrent tumors with greater than 10 mutations per Mb sequenced as these values were considered outliers (1.5 times the interquartile range above the upper quartile). Notably, there were a few initial gliomas that demonstrated a mutational frequency above 10 mutations per Mb. However, the "hypermutation" classification was restricted to only patients with this level at recurrence since these likely reflect different evolutionary paths.

Mutational signatures

The relative contributions of the COSMIC mutational signatures were determined from a patient's initial-only, recurrence-only, and shared mutations by solving the non-negative-least squares (NNLS) problem for each set of mutations using the 30 signatures from version 2 (March 2015). Six signatures were dominantly enriched in at least 3% of the fractions and we resolved the NNLS using the reduced six-signature model to increase accuracy and reduce noise.

Copy number segmentation

Copy number identification was performed according to the GATK Best Practices and is outlined briefly here. The pipeline differs slightly for whole genomes and whole exomes. For genomes, the genome was segmented into 10kb bins using 'PreprocessIntervals'. For exomes, overlapping regions between several commonly used capture kits (Broad Human Exome b37, Nextera Rapid Capture, TruSeq Exome, SeqCap EZ Exome V3, Agilent SureSelect V4, Agilent SureSelect V7) were identified using 'bedtools multiIntersectBed'. The tool 'PreprocessIntervals' was used to apply 1kb padding and to merge overlapping intervals. In parallel, 'SelectVariants' was used to subset the gnomAD resource of germline variants to variants with a population AF greater than 5%. Next, 'CollectReadcounts' was used to count reads in the bins generated by 'PreprocessIntervals' separately for autosomes and allosomes. In parallel, 'CollectAllelicCounts' was used to count reference and alternate reads at gnomAD variant sites with a population AF greater than 5%. The cohort was subsequently split into batches determined by sequencing center and 'CreateReadCountPanelOfNormals' was used to create a panel of normal (PON) for each batch. PONs were created separately for allosomes and autosomes, and allosomes were separated further by sex. To further improve the panel of normals, GC content annotation of each interval as determined by

'AnnotateIntervals' were given. Next, 'DenoiseReadCounts' was used to denoise the binned readcounts output by 'CollectReadCounts', given a PON determined by batch, chromosomes (allosomes or autosomes) and sex. Denoised copy ratios were plotted and inspected for quality concerns using 'PlotDenoisedCopyRatios'. The tool 'ModelSegments' is an implementation of a gaussian-kernel binary-segmentation algorithm and was used to merge contiguous segments and assign copy and allelic ratios. The results of this segmentation were plotted using 'PlotModeledSegments' and inspected for quality concerns.

Copy number calling

A copy number caller loosely based on GATK 'CallCopyRatioSegments' (which in turn is based off of ReCapSeg) and GISTIC was implemented to call both arm-level and high-level copy number changes, respectively [170, 171].

Segments (from 'ModelSegments') with a non-log2 copy ratio between 0.9 and 1.1 were determined to be neutral. These segments were then weighted by length and a weighted mean and standard deviation (sd) non-log2 copy ratio (once-filtered) were determined again. Outlier segments are removed and once again a weighted mean and sd non-log2 copy ratio (twice-filtered) were determined. Segments with a non-log2 copy ratio between 0.9 and 1.1 and segments within two standard deviations of the twice-filtered mean were determined to be neutral, and segments outside of these boundaries were determined to have a low-level amplification or deletion, depending on the direction.

The weighted mean and sd of the non-log2 copy ratio (once-filtered) was then determined individually for each chromosome arm. Outlier segments were removed and the weighted mean and sd of the non-log2 copy ratio (twice-filtered) was determined again. In order to determine a high-level amplification and deletion threshold, the most highly amplified and deleted chromosome arms were selected, respectively. The twice-filtered mean plus (high level amplification) or minus (high level deletion) two times the sd of the selected arms were used as high-level thresholds.

Gene level copy number were called by intersecting the gene boundaries with the segment intervals and by calculating the weighted non-log2 copy ratio for that gene. The copy number call for that gene was then determined by comparing the gene-level non-log2 copy ratio to the previously determined thresholds.

dNdScv

The R package dNdScv [172] (https://github.com/im3sanger/dndscv) was run using the default and recommended parameters for all mutations in initial tumor samples, recurrent tumor samples, and for each mutational fraction (unique to initial, unique to recurrent and shared). All analyses were conducted separately within the three main tumor subtypes.

Aneuploidy calculation

The most reductive metric of aneuploidy was computed by taking the size of all non-neutral segments divided by the size of all segments. The resulting aneuploidy value indicates the proportion of the segmented genome that is non-diploid. In parallel, an arm-level aneuploidy score modeled after a previously described method was computed [173]. Briefly, adjacent segments with identical arm-level calls (-1, 0 or 1) were merged into a single segment with a single call. For each merged/reduced segment, the proportion of the chromosome arm it spans was calculated. Segments spanning greater than 80% of the arm length resulted in a call of either -1 (loss), 0 (neutral) or +1 (gain) to the entire arm, or NA if no contiguous segment spanned at least 80% of the arm's length. For each sample the number of arms with a non-neutral event was finally counted. The resulting aneuploidy score is a positive integer with a minimum value of 0 (no chromosomal arm-level events detected) and a maximum value of 39 (total number of autosomal chromosome arms excluding the short arms for chromosomes 13, 14, 15, 21, and 22).

Estimates of evolutionary pressures

Evolutionary pressures were evaluated both by variant status and glioma subtype using the neutralitytestr algorithm as previously described (R-package: neutralitytestr version: 0.0.2, https://github.com/marcjwilliams1/neutralitytestr) [151]. Individual variant allele frequency vectors were merged at the level of glioma subtype by variant status. Only mutations found in copy-neutral regions should were included in these analyses. For all else, default parameters were used. Merged VAF distributions were deemed to be selected when the neutral null hypothesis was rejected using several metrics. Tests for neutrality required that both R^2 values < 0.98 and the area between the two curves of 1) merged VAF data and 2) a normalized distribution expected under neutrality to be significantly different. The SubclonalSelection algorithm was applied to GLASS mutation data to measure the selection strength in individual tumor samples (Julia package: Subclon-

alSelection, https://github.com/marcjwilliams1/SubClonalSelection.jl) [136]. Patients that had samples at both timepoints with a TITAN-defined purity estimate >= 0.5 and >= 25 subclonal mutations in non-diploid regions were included. Mean coverage across all mutations was used as the "read_depth" input parameter and the model was run with the recommended 10^6 iterations and 1000 particles. Samples were classified as neutral or selected based on the model that had the highest probability, in line with the prior applications to TCGA data [136]. Classification based on the highest model probability yielded stable results there was not a significant change in proportions when setting a higher classification probability threshold ($P > 0.05$, Pearson's Chi-square test, for both probability thresholds of 0.6 and 0.7). At all three probability thresholds (0.5, 0.6, and 0.7), Kaplan-Meier survival analyses between selection at recurrence and overall survival continued to indicate that patients with IDHwt tumors that were selected had a worse overall survival ($P = 0.03$ (n=81), $P = 0.01$ (n=66), $P = 0.01$ (n=56) respectively).

Mutation clonality

Each patient's clonal architecture was inferred using PyClone (version 0.13.1) by grouping SNVs into clonal clusters (https://github.com/aroth85/pyclone) [174]. The patient-level input mutation matrix was reduced by limiting to sites with at least 30x coverage across all samples. PyClone was subsequently ran using a binomial density model, connected initiation, and 10000 iterations. Sample purities were provided for each patient and parental copy number (minor and major allele counts) from TITAN were given. PyClone results were post-processed using a burn-in of 1000, thin of 1, minimum cluster size of 2 and a maximum number of clusters per patient of 12. Individual mutations were determined to be clonal if the PyClone cancer cell fraction (CCF) values were >= 0.5, subclonal for mutations with CCF >= 0.1 and CCF <0.5, mutations were considered non-clonal when CCF < 0.1 as previously described [175].

CNV clonality

Allele specific copy number, tumor purity and ploidy estimates were derived using a probabilistic model (TITAN, version 1.19.1) for both whole genome and whole exome sequencing samples [176]. TITAN was supplied with the tumor denoised readcounts output by GATK DenoiseReadCounts and the tumor allelic counts at loci found to be heterozygous in control samples output by ModelSegments. An 'alphaK' (and 'alphaKHigh') parameter of 2500 and 10000 was used for exomes

and genomes, respectively. The patient sex was provided in order to improve fitting allosomes. For each tumor-control pair TITAN was ran assuming an initial ploidy of two or three, and assuming 1 to 3 clusters, resulting in a total of six possible solutions per tumor/control pair. To select the optimal solution, TITAN's internal selectSolution function was used with a threshold of 0.15 giving additional weight to diploid solutions.

Timing analysis

The CCF values output by TITAN or PyClone were used for separately timing copy number changes or mutations. To time specific copy number changes in genes, the average CCF for that gene was calculated. When timing mutations in genes, the highest CCF amongst the non-synonymous mutations was taken.

Neoantigen analyses

Neoantigens in this analysis were defined as all 8-11-mer peptides that arose from an exonic nonsynonymous SNV or indel and bound their respective patient's HLA class I molecules at a binding affinity score (IC50) that was ≤ 500 nM and better than or equal to the wild-type form of the peptide. Each patient's 4-digit HLA class I types were inferred using OptiType (version 1.3.1, https://github.com/FRED-2/OptiType) run on each patient's matched normal sample [177]. VCF files for each tumor sample were annotated using Variant Effect Predictor (ensembl) with the Downstream and Wildtype plugins. Neoantigens from these VCFs were then called using pVACseq (version 4.0.10, https://github.com/griffithlab/pVACSeq) [159] run using netMHCpan (version 2.8, http://www.cbs.dtu.dk/services/NetMHCpan-2.8/) [178]. For each pVACseq run, epitope length was set to 8, 9, 10, or 11, minimum binding affinity fold-change was set to 1, and downstream sequence length was set to full, with default parameters used for all other settings. Downstream neoantigen analyses were performed using the pVACseq output linked to its respective mutation information. Neoantigen-causing mutations were defined as all mutations that gave rise to at least one neoantigen. The observed-to-expected neoantigen ratio was calculated using a previously developed approach that compares each tumor's observed neoantigen rate to an empirically derived expected rate that assumes no selection against neoantigen-causing mutations [157]: From the gold set samples in the GLASS cohort (n = 222), define $\overline{N}_s$ to be the expected number of nonsynonymous missense SNVs per synonymous SNV with trinucleotide context s. $\overline{B}_s$ is then defined as the ex-

pected number of neoantigen-generating missense SNVs per nonsynonymous missense SNV with trinucleotide context s. For a given sample i, define Y_i as the sample's set of synonymous SNVs and $s(m)$ to be a synonymous SNV with trinucleotide context m. The expected number of nonsynonymous missense SNVs, N_{pred}, and neoantigen-causing mutations, B_{pred}, can then be calculated as follows:

$$N_{pred,i} = \sum_{m \in Y_i} \bar{N}_{s(m)}$$

$$B_{pred,i} = \sum_{m \in Y_i} \bar{N}_{s(m)} \bar{B}_{s(m)}$$

To obtain sample i's final neoantigen depletion ratio R_i, the observed number of neoantigen-causing mutations in the sample, $B_{obs,i}$ is divided by the sample's observed number of nonsynonymous missense SNVs, $N_{obs,i}$, and then this ratio is divided by the ratio of $B_{pred,i}$ and $n_{pred,i}$. Thus:

$$R_i = \frac{B_{obs,i}/N_{obs,i}}{B_{pred,i}/N_{pred,i}}$$

For analyses examining clonal/subclonal neoantigen ratios, the observed and expected numbers were calculated by subsetting a sample's SNVs by the respective criteria and then recalculating the ratio as described above. To mitigate overfitting, all analyses presented here utilized samples from patients with at least 3 neoantigen-causing mutations in their primary and recurrent tumors.

Immune cell analyses

CIBERSORT relative immune cell fraction data used in downstream neoantigen analyses were downloaded from a previous publication [74].

Statistical methods

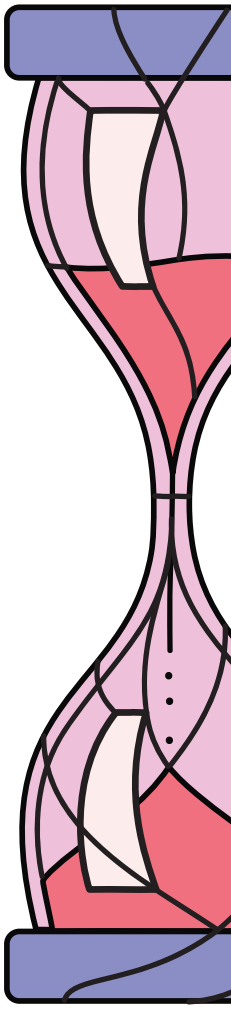
All data analyses were conducted in R 3.4.2, Python 2.7.15, PostgreSQL 10.5, and Julia 0.7. All survival analyses including Kaplan-Meier plots and Cox proportional hazards models were conducted using the R packages survival and survminer.

Data availability

All deidentified, non-protected access somatic variant profiles and clinical data are accessible via Synapse (http://synapse.org/glass). Raw data of the various sequencing datasets can be obtained per the overview provided in the Supplement.

Code availability

All custom scripts and pipelines are available on the project's github page (https://github.com/TheJacksonLaboratory/GLASS).



CHAPTER 4

Spatial concordance of DNA methylation classification in diffuse glioma

Published in print

Verburg N, **Barthel FP** (co-first author), Anderson KJ, Johnson KC, et al., Wesseling P, De Witt Hamer PC, Verhaak RGW. *Spatial concordance of DNA methylation classification in diffuse glioma.* Neuro Oncol. 2021 Dec 1;23(12):2054-2065.

Abstract

Background

Intratumoral heterogeneity is a hallmark of diffuse gliomas. DNA methylation profiling is an emerging approach in the clinical classification of brain tumors. The goal of this study is to investigate the effects of intratumoral heterogeneity on classification confidence.

Methods

We used neuronavigation to acquire 133 image-guided and spatially-separated stereotactic biopsy samples from 16 adult patients with a diffuse glioma (7 IDH-wildtype and 2 IDH-mutant glioblastoma, 6 diffuse astrocytoma, IDH-mutant and 1 oligodendroglioma, IDH-mutant and 1p19q codeleted), which we characterized using DNA methylation arrays. Samples were obtained from regions with and without abnormalities on contrast enhanced T1 weighted and fluid-attenuated inversion recovery MRI. Methylation profiles were analyzed to devise a three-dimensional reconstruction of (epi)genetic heterogeneity. Tumor purity was assessed from clonal methylation sites.

Results

Molecular aberrations indicated that tumor was found outside imaging abnormalities, underlining the infiltrative nature of this tumor and the limitations of current routine imaging modalities. We demonstrate that tumor purity is highly variable between samples and explains a substantial part of apparent epigenetic spatial heterogeneity. We observed that DNA methylation subtypes are often conserved in space in most patients after taking tumor purity and prediction accuracy into account.

Importance of study

Genetic and transcriptional intratumoral heterogeneity is a key feature of diffuse gliomas and thought to drive treatment failure. Epigenetic profiling is an emerging approach used for clinical classification of brain tumors that results in improved alignment of patients and treatments. However, how epigenetic intratumoral heterogeneity may impact the robustness of classification remains unclear. Through DNA methylation profiling of multiple spatially mapped biopsies from diffuse gliomas in 16 adult patients, and eleven patients in a validation cohort, our study shows that epigenetic spatial heterogeneity is confounded by

the presence of non-tumor cells and the prediction accuracy of the classifier. Taking these confounders into account, DNA methylation-based classification is conserved in space in most patients, including tumor presence outside standard imaging abnormalities. These findings emphasize the importance of tumor purity assessment in DNA methylation studies. The uniformity of DNA methylation-based classification within tumors demonstrates the robustness of this method for classification, corroborating its value for clinical studies and practice.

Introduction

Diffuse gliomas are the most common malignant brain tumors in adults [179]. Patients with a diffuse glioma have a poor prognosis and eventually succumb to treatment failure [180]. The diagnosis, treatment and follow-up of diffuse gliomas rely heavily on imaging [180], with magnetic resonance imaging (MRI) as the current standard. Using contrast enhanced T1 weighted (T1c) MRI, diffuse gliomas can be divided into enhancing tumors, predominantly glioblastoma, or non- enhancing tumors, predominantly low-grade gliomas (LGG). T1c MRI is used for enhancing and T2/Fluid-attenuated inversion recovery (FLAIR) MRI for non-enhancing gliomas [181]. However, diffuse glioma infiltration extends beyond the abnormalities detected on standard MRI [182, 183]. Also, the majority of diffuse gliomas recur directly adjacent to the standard MRI-guided surgical cavity [184]. Heterogeneity of tumor cells is a salient feature of diffuse gliomas and thought to be a driver of treatment failure. Treatment exposure may drive the clonal evolution of heterogeneous tumor cell populations, leading to the selection and survival of resistant subpopulations in some gliomas, whereas refractory disease in others may be driven by other factors [185].

Numerous studies have looked at genetic and transcriptomic heterogeneity in diffuse glioma. Recent single-cell transcriptome studies have elucidated transcriptional heterogeneity in regulatory programs that converge on the cell cycle or distinct cellular states [132, 186] while bulk tissue analysis has demonstrated extensive heterogeneity in somatic drivers such as *EGFR* and *PDGFRA [134, 187]* as well as in general somatic alteration burden [71, 73, 99].

DNA methylation is an epigenetic modification where a methyl-group is added to a cytosine, most commonly measured in the CpG dinucleotide context. These modifications are of interest to the neuro-oncology field as genome-wide patterns in DNA methylation profiles provide a robust method for disease classifica-

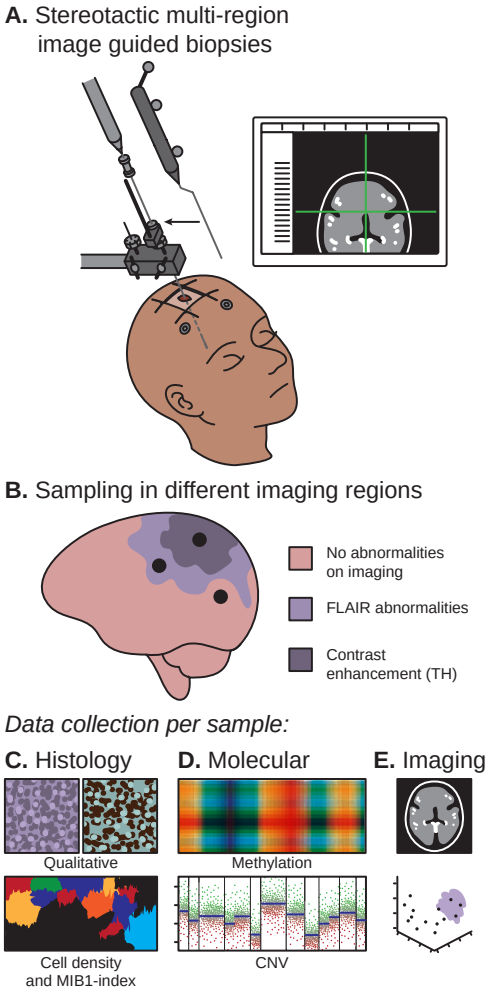
tion and a viable supplement to traditional histopathology [15, 94]. Nevertheless, the extent of intratumoral heterogeneity in DNA methylation remains unclear. In order to improve our understanding of the epigenetic heterogeneity of diffuse gliomas, we present a comprehensive analysis of DNA methylation of a large number of spatially-separated samples taken from regions with and without imaging abnormalities. We devised a three-dimensional reconstruction of the DNA methylation landscape for each tumor, with particular consideration to the variable ratios of tumor and non-malignant cells in each sample. These ratios are quantified as tumor purity using a methylation-based metric that, although not being the perfect test due to non-zero levels of tumor purity in non-malignant samples, has proven to be a major confounder in genomic analyses [77, 188]. Our analysis underlines the infiltrative nature of gliomas beyond visible tumor boundaries and demonstrates a rather homogeneous DNA methylation landscape across space.

Methods

Sample acquisition and study design

A schematic overview of the study design is given in **Figure 4-1**. The exploration cohort consisted of 16 patients with an untreated initial diffuse glioma, treated at the Amsterdam UMC, location VU medical center (VUmc), Amsterdam, The Netherlands. Patient characteristics are given in **Table 4-1**. All patients were participants of the FRONTIER study of which the protocol has been published [189]. This protocol was approved by the Medical Ethics Committee of the VUmc and registered in the Dutch National Trial Register (*www.trialregister.nl*, unique identifier NTR5354). Sampling was performed, using a stereotactic biopsy procedure preceding the craniotomy, to obtain two samples of each biopsy location for, respectively, FFPE and *Molfix*© (patient 1-8) or snap-frozen (patient 9-16) fixation. Samples were obtained from regions with and without abnormalities on T1c and FLAIR MRI in accordance with the study protocol. All procedures were carried out in accordance with the Helsinki Declaration [190]. Written informed consent was obtained from all patients.

The validation cohort comprised 11 patients with 61 FFPE samples from multi-sector sampling of an untreated diffuse glioma treated at the Toronto Western Hospital, Toronto, Canada or USCF Brain Tumor Center, San Francisco, USA. In addition, 64 FFPE samples from 64 patients without a glioma from the German Cancer Network served as controls.



↑ **Figure 4-1. Graphical overview of the methods.**

A. Multiple pre-operatively planned stereotactic biopsies were taken from each patient tumor. **B.** Biopsies were acquired in regions in and outside imaging abnormalities. **C.** Acquired tissue was subject to comprehensive histological and molecular analysis.

DNA isolation

DNA isolation was performed by adding proteinase K and incubation at 56°C using QIAamp DNA Mini Kit (Qiagen). DNA was quantified using a Qubit Fluorometer (ThermoFisher). Genomic DNA was bisulfite converted using Qiaamp DNA FFPE tissue Kit (Qiagen)

DNA methylation profiling by microarray

Data was processed using the minfi packages in R (R Foundation for Statistical Computing, Vienna). Data from the 450k (IlluminaHumanMethylation450k.ilmn12.hg19) and EPIC platforms (IlluminaHumanMethylationEPICanno.ilm10b2.hg19) were processed separately. Detection P-values were calculated for each probe and sample, and samples with an average detection P-value > 0.01 were removed from follow-up analysis. Data was normalized using BMIQ from the wateRmelon package in R. Probes on sex chromosomes and known cross-reactive probes were removed, as were probes mapping to known SNPs and probes with a detection P-value > 0.01. Finally, data from different platforms was merged.

DNA-methylation based classification

Glioma methylation subtype classification was performed using L2-regularized logistic regression using the R package LiblineaR. Classifiers were trained and evaluated on a set of common probes from TCGA glioma samples with known methylation subtypes. The classes LGm6-GBM and PA-like were merged into a single class LGm6-PA as the separation between these classes was based on phenotype. To improve classification accuracy of samples with low tumor purity, DKFZ controls were added to the classifier as separate classes. DKFZ DNA methylation classification was performed using the molecular neuropathology classification tool (version 11b4) [94]. Briefly, the software preprocesses, normalizes, and performs batch adjustment on raw DNA methylation data to generate probabilistic estimates for each defined tumor subtype. The tumor subtype with the highest probability was selected as the classification. Samples with a sub-classification accuracy lower than 0.5 were assigned to the main DKFZ class in accordance with the Molecular Neuropathology guideline.

Table 4-1. Overview of patients and biopsies in exploration and validation cohort.

Cohort	Patient	Histology	WHO grade	IDH status	1p/19 codeletion	Biopsies
Exploration	VUmc-01	Glioblastoma	IV	Mutant	No	8
Exploration	VUmc-02	Glioblastoma	IV	Wildtype	No	8
Exploration	VUmc-03	Astrocytoma	II	Mutant	No	7
Exploration	VUmc-04	Glioblastoma	IV	Mutant	No	11
Exploration	VUmc-05	Oligodendroglioma	II	Mutant	Yes	8
Exploration	VUmc-06	Astrocytoma	II	Mutant	No	8
Exploration	VUmc-07	Glioblastoma	IV	Wildtype	No	9

Table 4-1. Overview of patients and biopsies in exploration and validation cohort. *Continued*

Cohort	Patient	Histology	WHO grade	IDH status	1p/19 codeletion	Biopsies
Exploration	VUmc-08	Glioblastoma	IV	Wildtype	No	10
Exploration	VUmc-09	Astrocytoma	II	Mutant	No	6
Exploration	VUmc-10	Astrocytoma	II	Mutant	No	7
Exploration	VUmc-11	Glioblastoma	IV	Wildtype	No	9
Exploration	VUmc-12	Astrocytoma	II	Mutant	No	8
Exploration	VUmc-13	Glioblastoma	IV	Wildtype	No	9
Exploration	VUmc-14	Glioblastoma	IV	Wildtype	No	7
Exploration	VUmc-15	Astrocytoma	II	Mutant	No	6
Exploration	VUmc-17	Glioblastoma	IV	Wildtype	No	12
Validation	Toronto-01	Astrocytoma	III	Wildtype	No	8
Validation	Toronto-02	Glioblastoma	IV	Wildtype	No	5
Validation	Toronto-03	Glioblastoma	IV	Wildtype	No	7
Validation	Toronto-04	Glioblastoma	IV	Wildtype	No	4
Validation	Toronto-05	Glioblastoma	IV	Wildtype	No	8
Validation	UCSF-01	Astrocytoma	II	Mutant	No	4
Validation	UCSF-04	Astrocytoma	II	Mutant	No	6
Validation	UCSF-17	Oligodendroglioma	II	Mutant	No	3
Validation	UCSF-18	Astrocytoma	II	Mutant	No	4
Validation	UCSF-49	Astrocytoma	III	Mutant	Yes	6
Validation	UCSF-90	Astrocytoma	II	Mutant	No	6

Methylation purity estimation and simplicity score

DNA methylation measurements of tumor purity included the PAMES algorithm and simplicity score [74, 188]. For the PAMES non-malignant central nervous system samples from the German Cancer Research Center (DKFZ) were used as a control. PAMES operates in three steps. First, AUCs are calculated for each probe discriminating between tumor and normal. Second, a selection of the most informative probes is made. Third, tumor purity is calculated on input samples using these probes

DNA copy number aberrations inferred from EPIC microarray

Using the R/Conumee package, copy number aberrations were inferred from the 450k and EPIC array data. Merged data from the control samples was used as baseline control for all analyses. Genomic data was used to calculate aneuploidy.

Immunohistochemistry and qualitative assessment
FFPE samples from the exploration cohort were stained using hematoxylin and eosin (HE) and MIB-1. Two expert neuropathologists independently, blinded for imaging results, assessed the presence or absence of tumor in each sample. Consensus was obtained in case of disagreement. The patient's histopathological diagnosis was made based on resection material using routine procedures and according to the WHO 2016 criteria [61].

Histopathological analysis of whole-slide scans
Using a Hamamatsu Nanozoomer XR, FFPE slides stained with HE and MIB-1 of each sample were digitalized. The 40x magnification images were converted to multiple mosaic images using NDPITools software. Cellularity, defined as number of cells per mm2, was calculated with Cellprofiler. Proliferation index, defined as percentage of Ki-67 positive nuclei of all nuclei, was calculated using local developed software.

Radiologic evaluation of sample locations
Standard imaging sequences from the patients in the exploration cohort included T1-, T2-, T2/FLAIR and T1c MRI. For each sample location presence of an abnormal signal for each imaging sequence was independently assessed by a neurosurgeon and neurosurgical resident with ample experience in glioma imaging. Consensus was obtained in case of disagreement.

Sample-to-tumor surface distance
Tumors were segmented on FLAIR and, in case of contrast enhancement, also on T1c MRI, using Brainlab Software, by a neurosurgical resident with ample experience in glioma imaging. The segmentations and sample coordinates were exported in 3D T1c MRI space. Sample to tumor-surface distances were obtained, using Matlab, by calculating the distance between the sample coordinate and the nearest surface coordinate of the tumor segmentation, using FLAIR for non-enhancing and both FLAIR and T1c MRI for enhancing tumors.

Statistical analysis
Median values with interquartile range were used to describe non-normally distribute data. Mann-Whitney-U test was used to compare distributions between subgroups. Correlations were calculated with the Spearman or Pearson's correlation and compared using Fisher's z transformation. Comparison of percentages

between subgroups was performed using Fisher's test Normalization and scaling of purity measurement modalities was performed by subtracting the mean and dividing by the standard deviation. To compare absolute purity estimates, the normalized and scaled purity measurements were rescaled using the PAMES mean and SD. P values less than 0.05 was considered statistically significant. R (version 3.5.3) was used for all statistical analyses.

Heterogeneity analysis

Each probe per patient was classified as methylated (B>=0.3) or unmethylated (B<0.3). A table of all possible pairwise combinations of samples was generated. Each pair of samples was evaluated for heterogeneity by counting the number of identical (homogeneous) probes, the number of differing (heterogeneous) probes and percentages were subsequently calculated. Each pair was annotated according to the metadata for each sample in the comparison.

For each patient and sample type we tabulated all possible combinations of any number samples, iteratively including between 1 and the total number of possible samples. The proportion of heterogeneous and homogeneous probes was calculated when considering each sample in a given set. For each patient/sample type and sample number we then calculated the mean and standard deviation of the proportion heterogeneous across all sets.

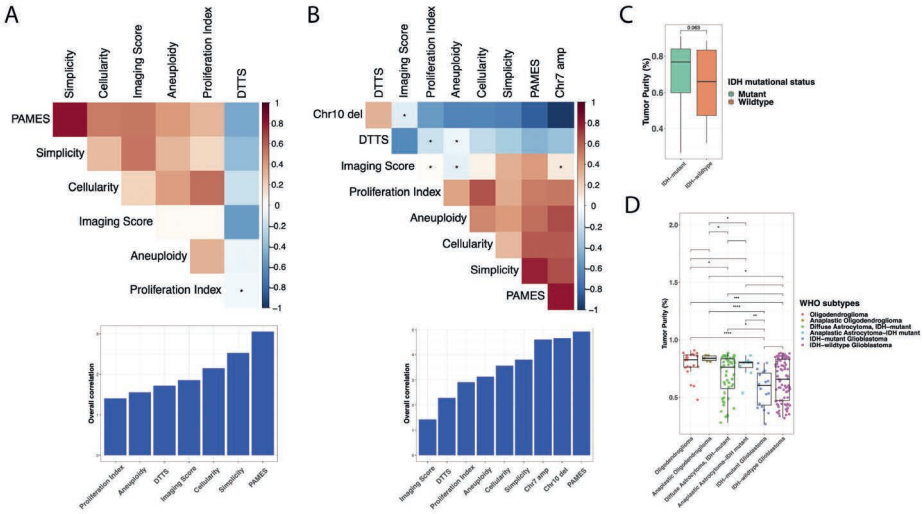
Results

We obtained 133 multi-region image-guided samples from eight patients (six glioblastoma, IDH-wildtype, 2 glioblastoma, IDH-mutant) with an enhancing tumor (76 samples) and eight patients (six diffuse astrocytoma, IDH-mutant, one glioblastoma, IDH-wildtype and one oligodendroglioma, IDH-mutant and 1p19q codeleted) with a non-enhancing tumor (47 samples) (**Figure 4-1**, **Figure 4-2**). In enhancing gliomas, 12 samples were taken outside both T1c and FLAIR abnormalities (T1c-/FLAIR-), 44 samples outside (T1c-/ FLAIR+) and 20 inside T1c abnormalities (T1c+/ FLAIR+). In non-enhancing gliomas, 16 samples were taken outside (FLAIR-) and 41 inside FLAIR abnormalities (FLAIR+). The maximum sample-to-tumor surface distance in enhancing tumors was 36.5 mm and 25.6 mm, respectively assessed with T1c and FLAIR MRI. In non-enhancing tumors the maximum sample-to-tumor surface distance was 22.5 mm assessed with FLAIR MRI.

VUmc 05
VUmc 03
VUmc 06
VUmc 09
VUmc 10
VUmc 12
VUmc 15
VUmc 01
VUmc 04
VUmc 02
VUmc 07
VUmc 08
VUmc 11
VUmc 13
VUmc 14
VUmc 17
IDH status
1p19q-codel status
Contrast Enhancement
Percent tumor purity (PAMES)
Tumor Reconstruction
Contrast-enhancing
Non-enhancing
Distance to tumor surface (mm)
MRI abnormalities
FLR
T1G
Normal
Abnormal
Cellularity (x10³ cells/mm³)
Percent MIB1 positive cells
Pathology assessment
Normal
Tumor

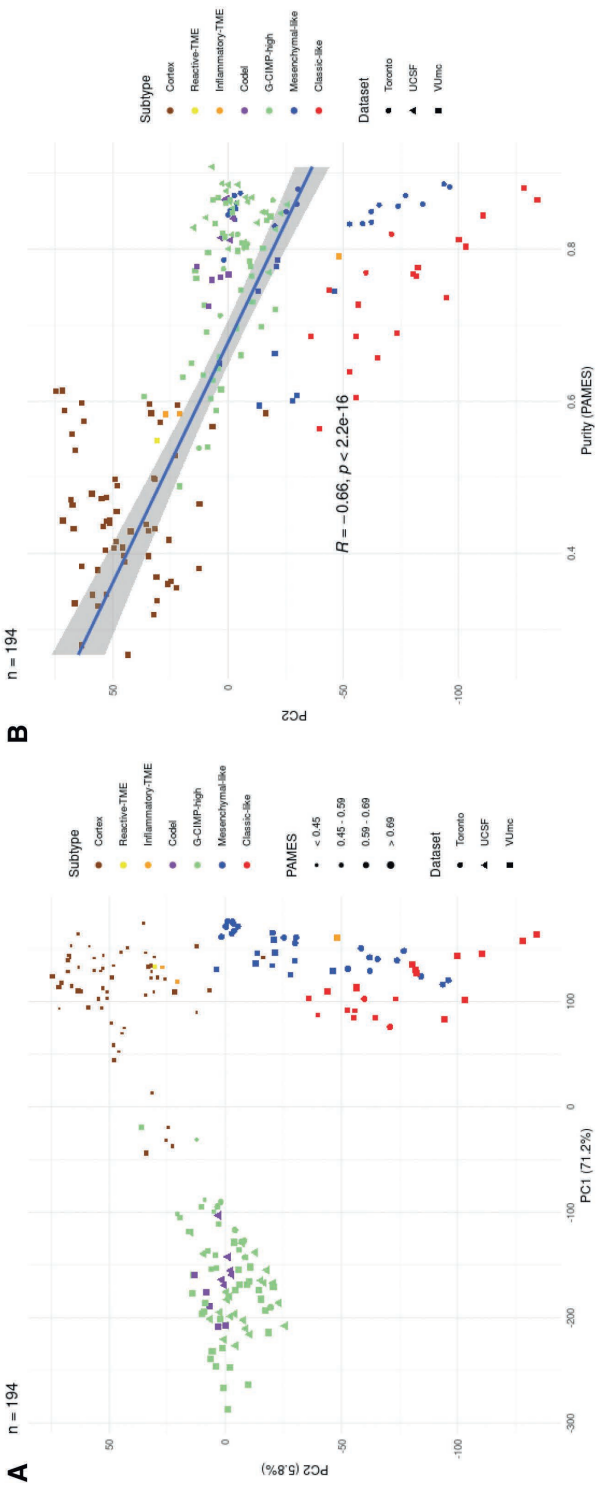
← Figure 4-2. Overview of 133 samples in 16 patients with initial diffuse glioma.

Samples are numbered in order of tumor purity for each patient. First row: IDH status with + representing a mutation, 1p/19q status with + representing a codeletion, and contrast enhancement with + representing presence of contrast on T1c MRI. Second row = Tumor purity assessed with PAMES. Third row = 3D reconstruction of FLAIR (yellow), T1c MRI (red) abnormalities and sample locations. Due to the 2d representation of a 3D object samples behind and outside the abnormalities might appear to be within the abnormalities. Fourth row = Euclidean distance (mm) between sample coordinate and tumor surface assessed with FLAIR (yellow) and T1c MRI (red). Negative values indicate samples obtained within the tumor volume. Fifth row = presence of abnormalities on FLAIR (top) and T1c MRI (bottom) at sample location. Sixth row = Median cellularity and percentage of MIB1 positive cells of the sample and final row = Consensus assessment of tumor presence by two neuropathologists. IDH = isocitrate dehydroxygenase, T1c = contrast enhanced T1 weighted, MRI = magnetic resonance imaging, PAMES= Purity Assessment from clonal MEthylation Sites, FLAIR = fluid-attenuated inversion recovery



↑ Figure 4-3. Tumor purity measurements.

A. and **B.** Spearman correlation of different tumor purity estimates in exploration cohort (**A.**) and subset of IDH-wildtype glioblastomas of exploration cohort (**B.**). Correlation matrix with colors representing Spearmans rho with non-significant correlation as marked with an asterix (*) for all patients (upper left) and patients with an IDH-wildtype glioblastoma (upper right). Barplot of sum of absolute correlations for each tumour purity measurement, using only significant correlations, for all patients (lower left) and patients with an IDH-wildtype glioblastoma (lower right). **C.** Boxplot of tumor purity of IDH-mutant and IDH-wildtype tumors in the combined exploration and validation cohort with p-value of Mann Whitney U test. **D.** Boxplot of the tumor purity of histological subtypes in the combined exploration and validation cohort with p-values of the Mann Whitney U tests.


A
n = 194
PC2 (5.8%)
PC1 (71.2%)
Subtype
Cortex
Reactive-TME
Inflammatory-TME
Codel
G-CIMP-high
Mesenchymal-like
Classic-like
PAMES
< 0.45
0.45 - 0.59
0.59 - 0.69
> 0.69
Dataset
Toronto
UCSF
VUmc
B
n = 194
PC2
Purity (PAMES)
R = −0.66, p < 2.2e-16

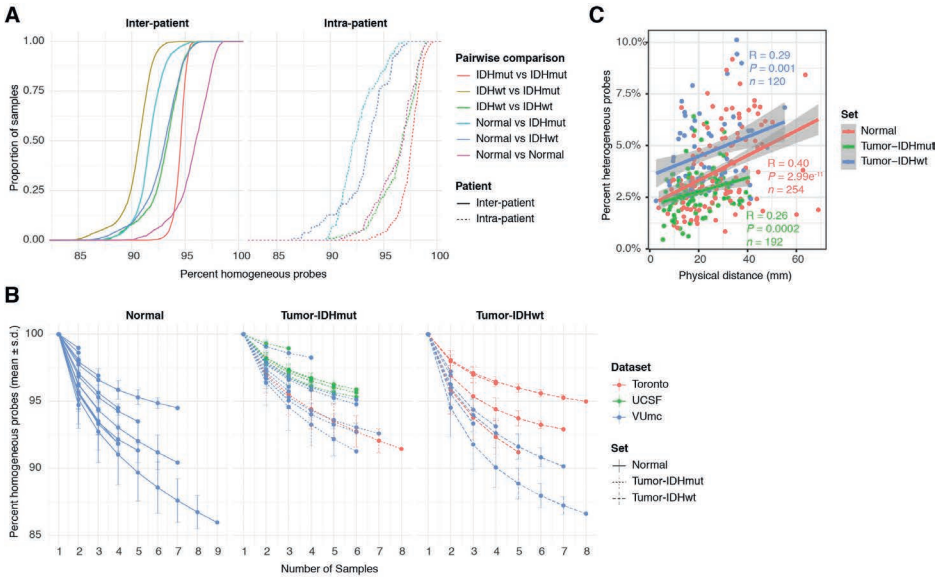
← **Figure 4-4. Exploration of spatial distribution of DNA methylation-based subtypes.**

A. Principal component analysis of exploration and validation cohort (samples=194). **B.** Correlation between the second principal component and tumor purity, with the correlation calculated with Pearson's R.

Tumor purity accounts for a considerable amount of variation in DNA methylation profiles

Since non-malignant cells in a sample influence molecular tumor assessment [74], we sought to quantify tumor purity, defined as the ratio of tumor to non-malignant cells. We evaluated and compared several methods of tumor purity estimation based on histology, MRI, DNA methylation and DNA copy number (**Figure 4-3a-b**). DNA methylation-based purity estimates, Purity Assessment from clonal MEthylation Sites (PAMES) provided the strongest correlations with all other features [188]. There was no difference in tumor purity between samples from IDH-mutant and IDH-wildtype tumors (**Figure 4-3c**). WHO subtype was associated with tumor purity (Kruskal-Wallis $p<0.001$), with the highest tumor purities in oligodendroglioma, IDH-mutant and 1p/19q co-deleted and the lowest for glioblastoma, IDH-mutant (**Figure 4-3d**), most likely due to the known admixture of non-malignant cells in grade IV diffuse glioma [191]. This association between grade and tumor purity is in line with a recent comprehensive analysis of TCGA samples [77].

We performed a principal component analysis of the DNA methylation data to elucidate drivers of differences in methylation (194 samples, **Figure 4-4a**). Included in the analysis were samples from a second cohort consisting of 61 multi-sector tumor samples from 11 gliomas (five diffuse astrocytoma, IDH-mutant, four glioblastoma, IDH-wildtype, one anaplastic astrocytoma, IDH-mutant and one anaplastic oligodendroglioma, IDH-mutant and 1p19q codeleted) [94, 192, 193]. The first principal component (percentage of variance 71.2%) separated samples based on IDH-status (**Figure 4-4a**). The second principal component (percentage of variance 5.8%) was associated with tumor purity, as evidenced by the linear increase in tumor purity and the samples from the control cohort (**Figure 4-4b**). These findings indicate that tumor purity accounts for a considerable amount of variation in DNA methylation profiles.



↑ **Figure 4-5. Spatial heterogeneity of genome-wide methylomes.**

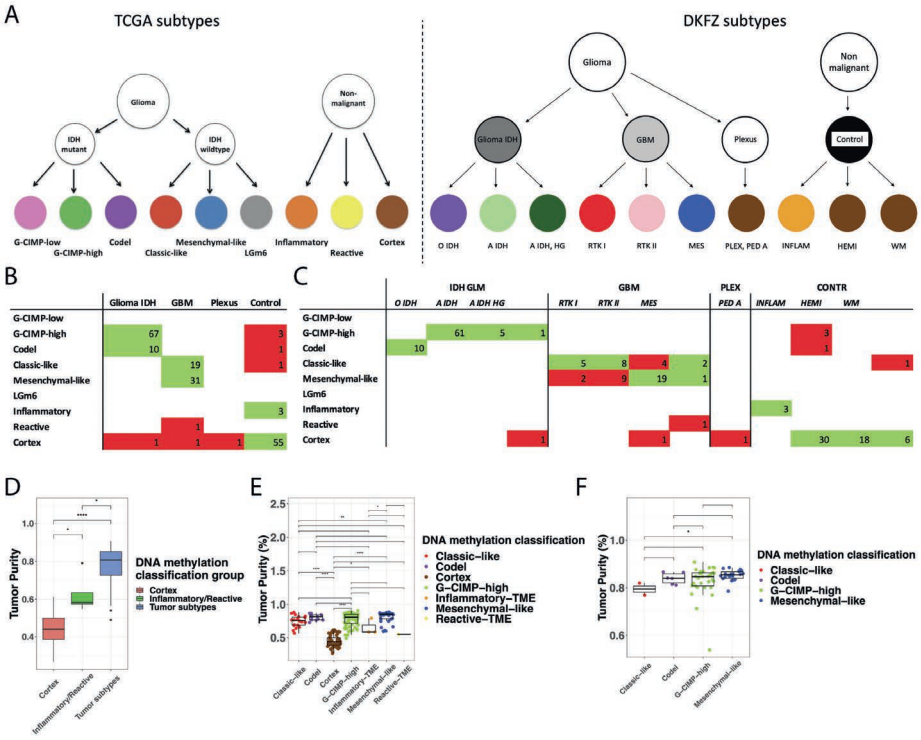
A. Empirical cumulative density function (ECDF) curves reflecting similarity (homogeneity) across all pairwise combinations of samples. Comparisons were separated based on whether they involved two samples from the same patient (intra-patient) or between two patients (interpatient) and based on whether the two samples spanned one or multiple sample types. **B.** Line plot showing the cumulative homogeneity associated with additional samples taken from the same tumor. Lines were colored by dataset, tumor and normal samples were separated, and tumor samples were further separated into IDHmut and IDHwt. **C.** Scatter plot of the relation between distance and methylation heterogeneity. Distance is the Cartesian distance in mm between two samples and methylation heterogeneity as described above. Correlation is calculated with Pearson's R.

DNA methylation heterogeneity is a function of space

To precisely quantify DNA methylation heterogeneity, we performed pairwise comparisons of binarized methylation values between samples. The vast majority of probes were homogeneously methylated (mean 0.93, range 0.83 – 1.0) between samples, suggesting that only a small fraction of probes is responsible for all intratumor heterogeneity. Similar trends have been observed in comparisons of non-malignant samples from the same lineage [194]. Unsurprisingly, any two samples from different unrelated tumors showed less probes with identical methylation (mean 0.93 ± 0.02) compared to any two samples from the same tumor (mean 0.96 ± 0.03). However, this difference was subgroup dependent. For example, any two samples from two unrelated IDH-mutant tumors show more

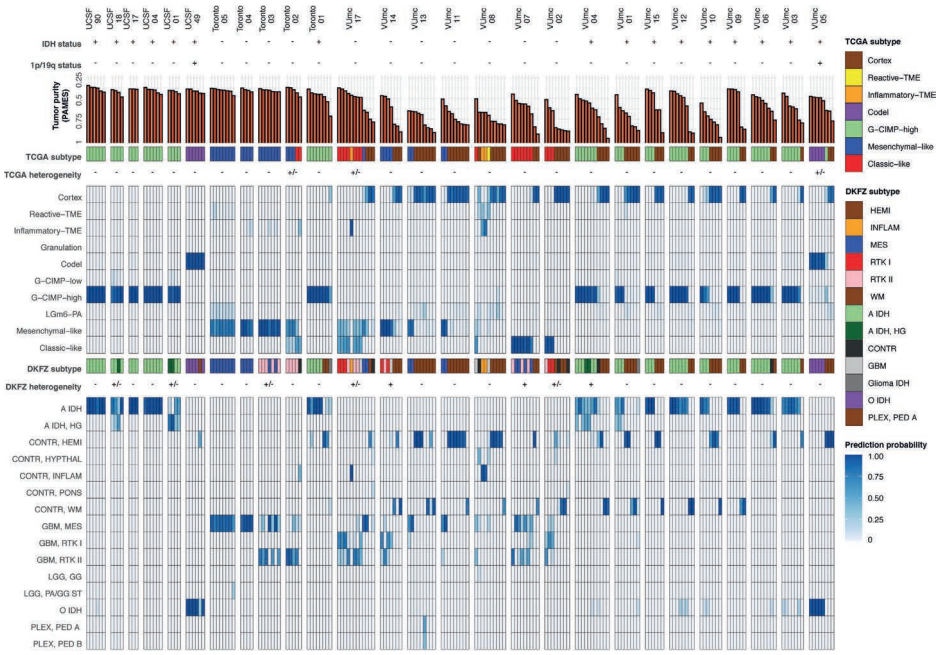
similarity on average than any two samples from two unrelated IDH-wildtype tumors (**Figure 4-5a**), likely related to the propensity of (G-CIMP positive) IDH-mutant tumors to uniformly methylate. As expected, a higher degree of heterogeneity can be observed when comparing samples classified as non-malignant to samples classified as tumor, based on DNA methylation classification, within the same patient. Any two IDH-wildtype tumor samples from the same patient show a comparable degree of heterogeneity when compared to two non-malignant samples from the same patient (Kolmogorov-Smirnov $P = 1.0$, pink and green dashed lines). In comparison, any two IDH-mutant tumor samples from the same patient demonstrate less heterogeneity compared to any two non-malignant samples from the same patient (Kolmogorov-Smirnov $P < 0.001$). These findings may reflect the clonal nature and shared ancestry of IDH-mutant tumor cells, whereas specimens classified as non-malignant harbor cells from a mixture of lineages.

To assess the impact of additional samples on tumor heterogeneity we calculated the percentage of identical probes pooling any number of samples per patient, separating samples classified as tumor and non-malignant (**Figure 4-5b**). The majority of heterogeneity was captured by the first two samples per patient. Although additional samples further contributed to overall heterogeneity, the change in heterogeneity decreased with each additional sample. Next, we investigated the relation between sample-to-sample distance and heterogeneity. These were positively correlated for both non-malignant and tumor samples (**Figure 4-5c**). These results suggest that heterogeneity is a function of space and that increased physical separation between cells increases cell-to-cell heterogeneity.



↑ **Figure 4-6. Comparison of DNA methylation subtypes and tumor purity.**

A. Overview of TCGA (left) and DKFZ subtypes (right) with representative colors. In the DKFZ subtypes, the main classes of Glioma IDH, GBM and Control also have a representative color since some samples could not be subclassified and were therefore assigned a main class. **B.** Cross table of TCGA and DKFZ classification. The classification of a sample as DKFZ subtype Plexus appears to be an incorrect classification, despite the more than 0.5 prediction accuracy, since none of the patients were diagnosed with a plexus tumor. **C.** Cross table of TCGA and DKFZ subclassification **D.** Boxplot showing tumor purity of the TCGA subtype Cortex, the combined TCGA subtypes Inflammatory and Reactive (Inflammatory/Reactive) and the combined TCGA subtypes G-CIMP-high, Mesenchymal-like, Classic-like and Codel (Tumor subtypes) of the exploration cohort. Statistically significant differences using the Mann Whitney U tests are marked with an asterix (*). **E.** Boxplot with the tumor purity of the TCGA subtypes of the exploration cohort. Statistically significant differences using the Mann Whitney U tests are marked with an asterix (*). **F.** Boxplot with the tumor purity of the TCGA subtypes of the validation cohort. Statistically significant differences using the Mann Whitney U tests are marked with an asterix (*). IDH GLM = IDH mutant glioma, GBM = glioblastoma, MNG = meningioma, SCHW = schwannoma, PLEX = plexus tumor, CONTR = control tissue, A IDH = astrocytoma, A IDH HG = astrocytoma high grade, O IDH = oligodendroglioma, MES = mesenchymal, MEL = melanotic, PED A = paediatric A, PED B = paediatric B, WM = white matter, HEMI = hemispheric cortex, REACT = reactive tumor microenvironment, HYPTHAL = hypothalamus, INFLAM = inflammatory tumor microenvironment.

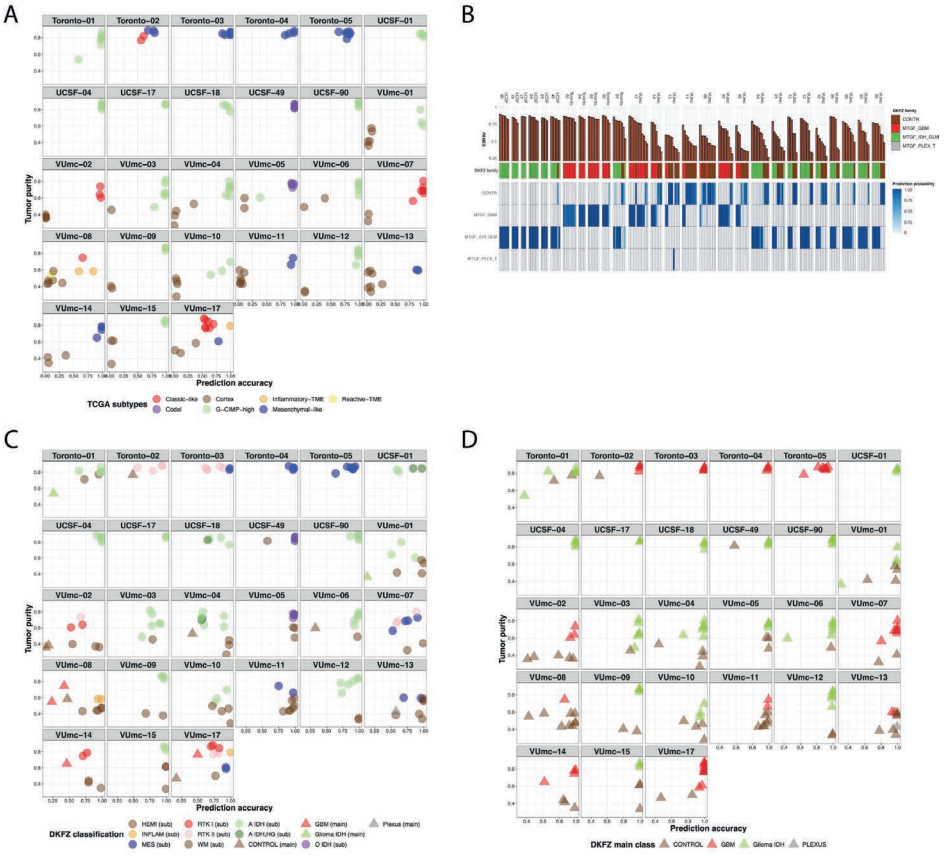


↑ **Figure 4-7. Overview of DNA methylation subtypes.**

Patients of both the exploration and validation cohort are ordered according to their cohort and DNA methylation subtype. Both TCGA and DKFZ subtypes and classification probabilities are given. Intratumoral heterogeneity of DNA methylation subtypes is represented as not present (-), possibly present (+/-) or present (+).

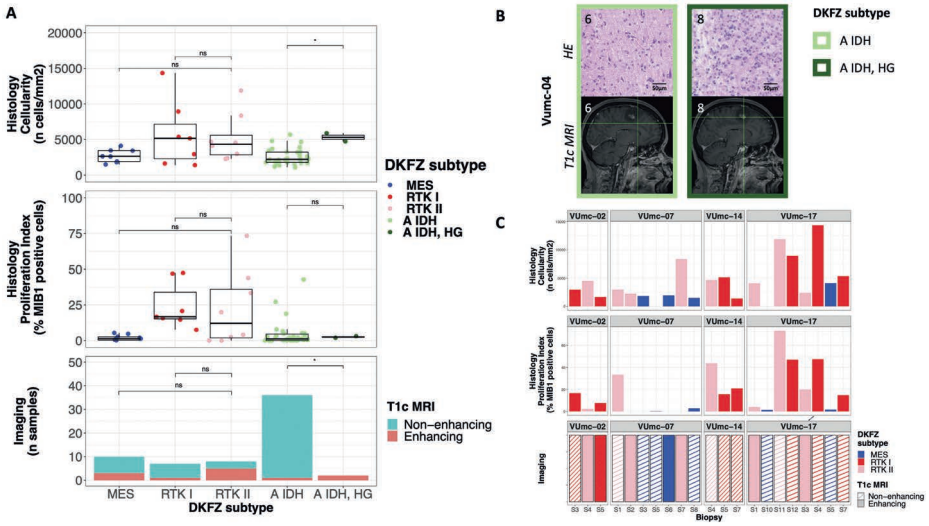
DNA methylation-based classification is highly conserved in space after adjusting for tumor purity and classification ambiguity

To establish the relationship between DNA methylation-based classification and tumor purity, we inferred sample subtypes based on two previous published classifiers (**Figure 4-6a**) [15, 94], which showed a large conformity for classification families (4.6% discordance, **Figure 4-6b**) and slightly lower conformity for family subtypes (12.4% discordance, **Figure 4-6c**). As expected, samples with a low tumor purity were assigned a control subtype whereas high tumor purity samples were assigned a tumor subtype when assessed by the TCGA classifier (**Figure 4-6d**). The differences in subtype assignment and its relation to tumor purity was clearly captured by the principal component analysis (**Figure 4-4a**).



↑ **Figure 4-8. Overview of DNA methylation subtypes.**

A. Plot of tumor purity (y-axis) and prediction accuracy (x-axis) of samples with colors representing the TCGA subtype of all patients. The different position on the x-asis of the different subtypes shows the influence of tumor purity on the heterogeneity that was found. **B.** Patients of both the exploration and validation cohort are ordered according to their cohort and DNA methylation subtype. Tumor purity, DKFZ main class and classification probabilities are given. **C.** Plot of tumor purity (y-axis) and prediction accuracy (x-axis) of samples with colors representing the DKFZ main classes of all patients. **D.** Plot of tumor purity (y-axis) and prediction accuracy (x-axis) of samples with colors representing the DKFZ subtype of all patients. The position of the different subtypes on the y-asis in patient VUmc-02 and VUmc-17, and on the x-axis in patient Toronto-03, UCSF-01 and UCSF-18, shows the influence of, respectively, tumor purity and prediction accuracy on the heterogeneity that was found.



↑ **Figure 4-9. Histology and imaging of patients with DNA methylation subtype heterogeneity.**

A. Boxplots of cellularity (upper) and proliferation index (middle) and barplot of T1c MRI contrast enhancement (lower) for the DKFZ subtypes MES (blue), RTK I (red), RTK II (pink), A IDH (lightgreen) and A IDH, HG (darkgreen) of all patients from the exploration cohort. Statistical significanse is given using Mann Whitney U test for the continuous data (Histology) and Chi-squared test for the frequency data (Imaging), with * representing p<0.05. **B.** Representive examples of the histological and radiological images of the patient VUmc-04: HE stained histology slide of a sample classified as A IDH (S3) and A IDH, HG (S9) showing the higher cellularity in the A IDH, HG sample. Sagittal T1c MRI images of the same samples showing contrast enhancement in the A IDH, HG, but not in the A, IDH sample. **C.** Barplot of cellularity (upper), proliferation index (middle) and contrast enhancement on T1c MRI (lower) of the tumor samples of glioblastoma patients of the exploration cohort with subtype heterogeneity. T1c = contrast enhanced T1 weighted, MRI = magnetic resonance imaging, HE = hematoxylin and eosin.

There were no significant differences between the tumor purity of the different tumor subtypes in the exploration dataset (**Figure 4-6e**), while in the validation dataset (**Figure 4-6f**) the tumor purity was lower in the Classic-like compared to Mesenchymal-like subtype, which are methylation subtypes based on unsupervised hierarchical clustering with a majority of, respectively, Classical and Mesenchymal expression subtypes as described by Verhaak et al [22]. To explore the heterogeneity of tumor subtypes within a tumor, we analyzed which tumor subtypes were recognized within each patient across the core and validation dataset. The majority of patients (24 of 27) did not show heterogeneity in tumor subtype as assessed by the TCGA classifier (**Figure 4-7**). In the three patients

(Toronto-02, VUmc-05 and VUmc-17) with TCGA tumor subtype heterogeneity, the one (VUmc-05, VUmc-17) or two (Toronto-02) discordant samples were the lowest purity tumor sample identifying tumor purity as the confounding factor for the found heterogeneity (**Figure 4-8a**). Also, prediction accuracy of tumor samples was lower in patients with subtype heterogeneity (mean prediction accuracy 71.0%) than without (mean 94.9%, $P < 0.001$).

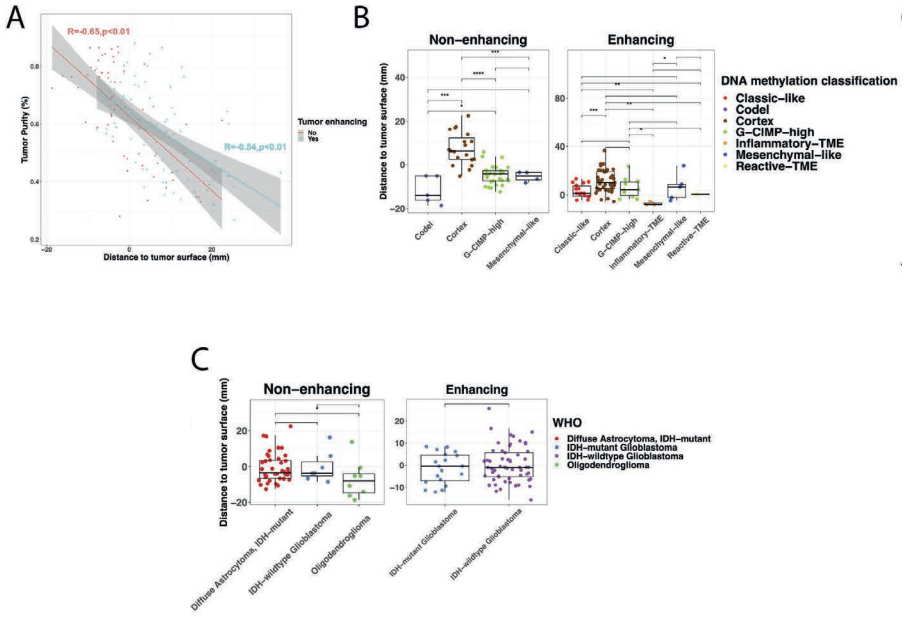
When we classified samples according to the DKFZ classifier, no heterogeneity was found for the main classes (**Figure 4-8b-c**). Heterogeneity of main class subtypes, so called family members, was found in eight patients (VUmc-02, VUmc-04, VUmc-07, VUmc-14, VUmc-17, Toronto-03, UCSF-01 and UCSF-18). Both tumor purity and prediction accuracy were lower in tumor samples of these eight patients with subtype heterogeneity (mean tumor purity 75.9 ± 0.09 and mean prediction accuracy 73.4 ± 14.9) compared to tumor samples of patients without heterogeneity (mean 78.1 ± 10.9, $P = 0.046$ and 87.8 ± 18.0, $P < 0.001$, respectively), suggesting tumor purity and prediction accuracy as potential confounding factors for the observed heterogeneity in these patients (**Figure 4-8d**). In two patients (VUmc-02 and VUmc-17) the discordant samples were the lowest tumor purity, although the absolute difference in tumor purity was small. In three patients (Toronto-03, UCSF-01 and UCSF-18) the discordant samples were the lowest prediction accuracy, although the absolute difference in prediction accuracy was small except for patient UCSF-01. Three patients (VUmc-04, VUmc-07 and VUmc-14) showed variable tumor purity and prediction accuracy of the tumor samples suggesting true heterogeneity. To validate the possible heterogeneity in these patients, we evaluated the histological and imaging data that was available for the exploration cohort. The only difference we found between subtypes was a higher cellularity ($P = 0.016$) and more frequent contrast enhancement on T1c ($P < 0.001$) in A IDH, HG samples compared to A IDH samples (**Figure 4-9a**). This confirmed the true heterogeneity in patient VUmc-04, with both A IDH, HG samples having the highest cellularity and showing contrast enhancement on T1c (**Figure 4-9b**). In the four glioblastoma patients with multiple subtypes, neither histology or imaging could confirm or reject the subtype heterogeneity (**Figure 4-9c**). Therefore, we conclude there is true DNA methylation subtype heterogeneity in patient VUmc-04, based on both DNA methylation, histology and imaging data, and patients VUmc-07 and VUmc-14, based on DNA methylation data. Based on all available data it remains unclear if there is true heterogeneity in the remaining five patients. Although not statistically significant, all true heterogenous patients had a glioblastoma (Chi-squared p=0.17), yet heterogeneity

was not related to IDH status (Chi-squared p=0.73). Since MGMT promotor methylation status is an important prognostic and predictive marker we assessed its spatial distribution. Heterogeneity was found in 30% of patients and was not related to tumor purity (Pearson correlation p=0.19), IDH status, histology or DNA methylation classification heterogeneity (Chi-squared p=1.0, p=0.76 and p=0.78, respectively). Overall, in our cohort, a low frequency of true spatial heterogeneity of DNA methylation-based subtypes was observed (14%, 3 of 22 patients), when taking tumor purity and prediction accuracy into account, yet possible heterogeneity cannot be ruled out in another 5 (19%) patients.

DNA methylation abnormalities extend beyond standard MRI boundaries

To understand the spatial distribution of glioma infiltration, we analyzed the correlation of tumor purity and subtypes with the distance to the tumor surface assessed with T1c for enhancing and FLAIR MRI for non-enhancing tumors. As expected, the distance of samples to the tumor surface showed a linear relationship with tumor purity (**Figure 4-10a**). In non-enhancing tumor, samples classified as Cortex were found further away from the radiological tumor boundaries than all the other subtypes (**Figure 4-10b**). Yet, in enhancing tumor this difference was not found, possibly indicating a more diffuse infiltration pattern of enhancing tumors. When evaluating the infiltration pattern in the histological subgroups, no difference was found except that distance to tumor surface in one oligodendroglioma (median -7.9mm, IQR -14.5 – -3.9mm) was lower than in IDH-mutant diffuse astrocytoma (-3.2mm, -6.3 – 3.6mm, p=0.047) (**Figure 4-10c**).

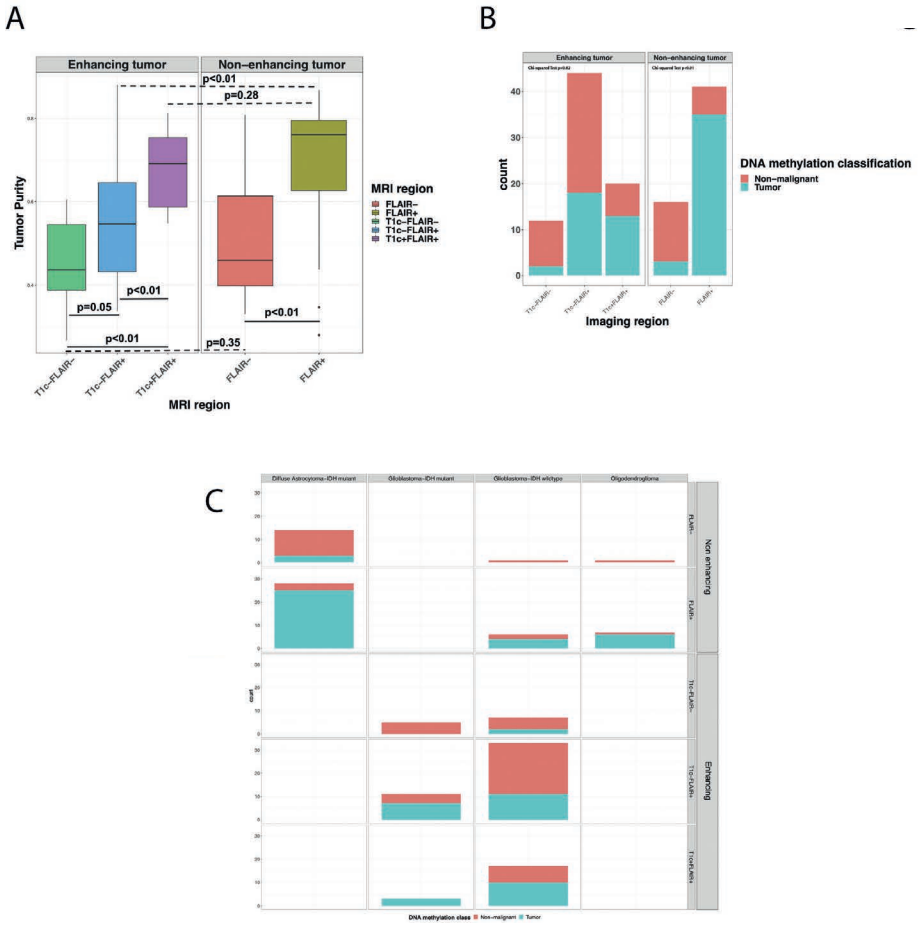
As anticipated, in enhancing tumors the T1c+/FLAIR+ region showed the highest tumor purity, followed by the T1c-/FLAIR+ and T1c-/FLAIR- (**Figure 4-11a**). In non-enhancing tumors, tumor purity was higher in the FLAIR+ than the FLAIR- region. Interestingly, samples taken from regions outside standard imaging abnormalities (FLAIR- in non-enhancing and T1c-/FLAIR- and T1c-/FLAIR+ in enhancing tumors) showed a tumor subtype in 36% and 17% of enhancing and non-enhancing gliomas, respectively (**Figure 4-11b**).



↑ **Figure 4-10. Distance to tumor surface in the exploration cohort.**

A. Scatterplot of distance to tumor surface and tumor purity with colors indicating enhancing (red) or non-enhancing tumor (green) with Pearson correlation coefficient. **B.** Boxplot of distance to tumor surface of TCGA subtypes with * indicating a significant difference using the Mann Whitney U test. **C.** Boxplot of distance to tumor surface of histological subgroups, according to the 2016 WHO classification with * indicating a significant difference using the Mann Whitney U tests.

Conversely, samples taken from within the standard imaging abnormalities showed a non-malignant subtype in 35% of enhancing tumors, which is most likely due to necrosis in these samples. In non-enhancing tumors 15% of samples within the FLAIR+ region showed a non-malignant subtype. Tumor presence, assessed as ratio of tumor and non-malignant TCGA subtypes, in the different imaging regions was comparable between histological subgroups (all Chi-squared tests p>0.05) (**Figure 4-11c**). Samples with tumor subtypes were found up to 24mm outside imaging abnormalities. These findings support the diffusely infiltrative nature of these tumors and corroborate the notion that standard MRI does not capture the true extent of diffuse glioma infiltration.



↑ **Figure 4-11. Imaging regions in the exploration cohort.**

A. Boxplot of tumor purity of the different imaging regions of enhancing and non-enhancing tumors with p-values of Mann Whitney U tests. **B.** Barplot of tumor presence for the different imaging regions for enhancing (left) and non-enhancing (right) tumors. Tumor presence was defined using the TCGA subypes with all but Cortex, Inflammatory and Reactive classified as tumor. The proportion of non-malignant and tumor samples of the imaging regions was compared using the Chi-squared test. **C.** Barplot of tumor presence for the different imaging regions (y axis) for the different histological subtypes (x axis) in enhancing and non-enhancing tumors (y-axis). Tumor presence was defined using the TCGA subypes, with all subtypes, except for the Cortex, Inflammatory and Reactive subtype, classified as tumor The proportion of non-malignant and tumor samples of the imaging regions per histological subtype was compared using the Chi-squared test.

Discussion

This study represents a comprehensive analysis of spatially-separated samples in diffuse glioma. The combination of histological, radiological and DNA-methylation data enabled us to explore the spatial contexts of tumor purity, epigenetic molecular subtypes and tumor heterogeneity. Our study demonstrates that in most tumors molecular subtypes are stable and homogeneous after considering tumor purity. Moreover, gliomas are diffusely infiltrative tumors and our data clearly shows that they indeed extend beyond the tumor boundaries found on MRI. Finally, in our study the extent of heterogeneity in tumor samples was predominantly equal to or less than in non-malignant samples.

Information on the spatial heterogeneity of epigenetic molecular subtypes in the literature is limited. A recent study reported intratumor DNA methylation-based subtype heterogeneity in five of twelve glioblastomas in their cohort [195]. We were unable to confirm this extent of heterogeneity in our study. The differences may be explained by our approach to account for tumor purity prior to determining intratumoral epigenetic subtype classification. The non-purity related heterogeneity that was found could be explained by focal malignant progression in patients with an IDH-mutant astrocytoma, while cellular state plasticity, as described by Neftel et al,[186] might be an explanation for the coexistence of both Mesenchymal and RTK II subtypes in one of the IDH-wildtype glioblastoma patients. Although technically difficult, future single cell DNA methylation profiling can be expected to drastically increase our understanding of DNA methylation classification heterogeneity [196]. MGMT promotor methylation status proved to be more heterogeneous than DNA methylation subtypes, which is most likely due to the use of 2 probes for the MGMT status compared to ~1000 probes for the DNA methylation subtype.

We observed that samples obtained outside standard imaging abnormalities, on FLAIR in non-enhancing and on T1c MRI in enhancing gliomas, displayed similar epigenetic molecular subtypes as the samples from the tumor core. The fraction of tumor cells per specimen varied between the different MRI regions, with lower fractions in regions outside of imaging abnormalities. The presence of tumor tissue outside standard imaging abnormalities is well known [182, 183]. Our results suggest that spatial imaging heterogeneity in glioma is associated with tumor purity and not with epigenetic heterogeneity, although heterogeneity at the genomic and mutational level as cause cannot be excluded and has recently been reported for progressive IDH-mutant astrocytoma [197]. This was confirmed by the strong correlation between tumor purity and the imaging score.

Our observations imply that a viable part of the tumor, especially in IDH-wildtype glioblastomas, is left behind after resection of standard imaging abnormalities.

A limitation of this study is the difficulty of assessing the true tumor purity since there is no gold standard. Still, we have compared many tumor purity metrics and choose the most representative of those. Another limitation is the evaluation of tumor heterogeneity by bulk DNA methylation analysis instead of single-cell analysis, although correlation imaging data with single cell analysis might proof difficult due to the different level of detail. Finally, the use of T1c and FLAIR MRI for the assessment of tumor to sample distance is a limitation, since these modalities are known to be less accurate for the detection of diffuse glioma infiltration [198]. However, they are still the current standard imaging and therefore represent the daily practice of glioma treatment.

In conclusion, we demonstrate that DNA methylation subtypes in diffuse glioma show little intratumoral heterogeneity and are uniform across the different imaging regions, underscoring the diffuse infiltrative nature of this disease and the robustness of DNA methylation subtypes.

CHAPTER 5

Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer

Published in print

Kocakavuk E, Anderson KJ, Varn FS, Johnson KC, Amin SB, Sulman EP, Lolkema MP, **Barthel FP** (co-senior author), Verhaak RGW. *Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer.* Nat Genet. 2021 Jul;53(7):1088-1096.

Abstract

Ionizing radiation causes DNA damage and is a mainstay for cancer treatment, but we have limited understanding of its genomic impact. We analyzed mutational spectra following radiotherapy in 190 paired primary and recurrent gliomas from the Glioma Longitudinal Analysis (GLASS) Consortium and 3,693 post-treatment metastatic tumors from the Hartwig Medical Foundation (HMF). We identified radiotherapy-associated significant increases in the burden of small deletions (1-20 bp) and large deletions (20+ bp to chromosome-arm length). Small deletions were characterized by a larger span size, lacking breakpoint microhomology and were genomically more dispersed when compared to pre-existing deletions and deletions in non-irradiated tumors. Mutational signature analysis implicated c-NHEJ-mediated DNA damage repair and APOBEC-mutagenesis following radiotherapy. A high radiation-associated deletion burden was associated with worse clinical outcomes, suggesting that effective repair of radiation-induced DNA damage is detrimental to patient survival. These results may be leveraged to predict sensitivity to radiation therapy in recurrent cancer.

Introduction

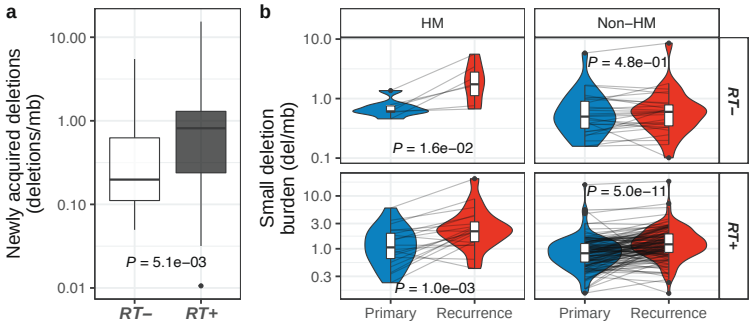
Radiation therapy or radiotherapy (RT) is used in the clinical management of more than half of all cancer patients [199, 200]. Ionizing radiation kills cells by inducing DNA damage such as double-strand breaks (DSBs), leading to cell death if left unrepaired or repaired in a manner that inhibits subsequent replication. DNA repair pathways are activated in response to DSBs, and these pathways include the error-free homologous recombination (HR) pathway and three error-prone pathways: classical non-homologous end joining (c-NHEJ), alternative end joining (a-EJ) or single strand annealing (SSA) [201]. In contrast to HR, c-NHEJ, a-EJ and SSA require different lengths of microhomologous sequences present on exposed DNA ends. Whereas c-NHEJ requires no microhomologies, a-EJ uses 2-20 bp length of microhomology (also called microhomology-mediated end joining, MMEJ) and SSA uses >20 bp of (micro-)homology. These repair processes may leave scars (mutations) in the post-treatment cancer genome that can be recognized through genome sequencing. The identification of therapy-associated mutations may imply an effect of therapy on the tumor and can aid in characterization of therapy-resistance mechanisms. A well-known example of such a process is hypermutation following treatment with DNA-alkylating agents, observed across cancers [202] and in particular following temozolomide (TMZ) chemotherapy of gliomas [185, 203]. Similarly, an increased burden of small deletions has been observed in rare radiation-induced malignancies [204]. Despite these advancements, the mutational footprints of therapeutic radiation in sporadic tumors are not well understood.

To address this gap in knowledge we leveraged pre- and post-treatment samples from the Glioma Longitudinal Analysis (GLASS) dataset as well as post-treatment metastatic tumor samples from the Hartwig Medical Foundation (HMF) dataset [137, 185, 205]. We identified a significant increase of small 5-15bp deletions, large deletions (+20 bp) and inversions in response to ionizing radiation, which we genomically characterized. Finally, we observed that the identified signatures were associated with worse clinical outcomes. This work has comprehensively characterized the genomic effects of ionizing radiation across cancers and provides a framework that may aid in increasing the therapeutic efficacy of radiation therapy for the clinical management of cancer.

Results

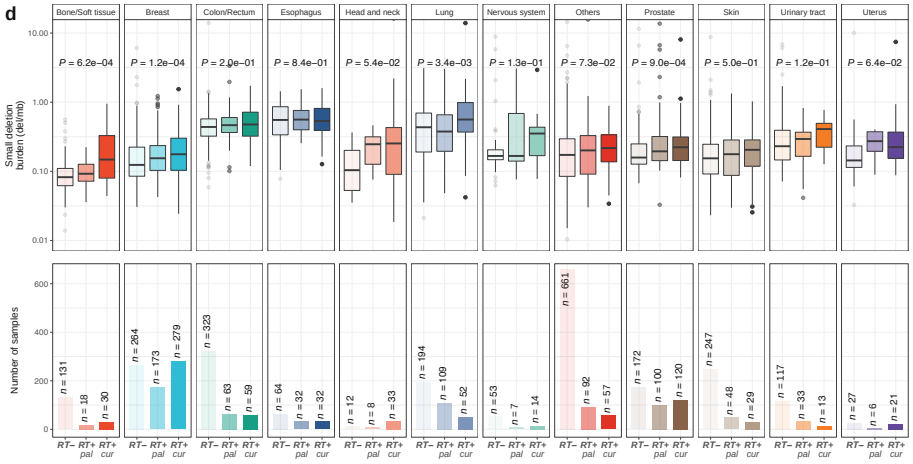
Radiotherapy treatment is associated with an increased small deletion burden

Somatic mutations can arise spontaneously and in response to anti-cancer treatment [206, 207]. Regimens involving radiotherapy (RT) and temozolomide (TMZ) are the post-surgical standard of care for patients with a diffuse glioma [208]. We assessed the impact of RT and TMZ on the burden of somatic mutations, including somatic single nucleotide variants (sSNVs) and small insertions/deletions (indels, length of 1 to 20bp), in matched pre- and post-treatment glioma samples (n = 190 patients). Of these, 119 (63%) patients received the combination of RT and TMZ, 19 (10%) received RT alone, 13 (7%) underwent only TMZ treatment, and 16 (8%) received no RT or TMZ treatment. For 23/190 (12%) cases, TMZ annotation was lacking with 18 of these having received RT. For each patient, we separated mutations into pre- (present in the primary tumor) and post-treatment (only present in the recurrent tumor). We then calculated the mutation burden (average mutation frequency per megabase (Mb)) of post-treatment mutations. A median of 0.68 new small deletions/Mb were acquired in recurrent RT-treated glioma which was significantly higher than the median of 0.19 new small deletions/Mb acquired in recurrent RT-naïve gliomas (**Figure 5-1a**, P = 5.1e-03, Mann-Whitney U test), and significantly higher than the small deletion burden detected at diagnosis (**Figure 5-1b**). RT was not associated with a significant increase in the sSNV burden (**Figure 5-2**a, P = 4.7e-01, Mann-Whitney U test) or small insertion burden (**Figure 5-2**, P = 6.7e-01, Mann-Whitney U test).



c

Variable		*n*	OR (95%CI)		*P*
Temozolomide	—	35		Reference	
	+	132		1.02 (0.82, 1.25)	0.885
Hypermutation	—	142		Reference	
	+	25		1.69 (1.38, 2.05)	**<0.001**
Radiotherapy	—	29		Reference	
	+	138		1.37 (1.11, 1.69)	**0.003**
Subtype	IDHmut−codel	18		Reference	
	IDHmut−noncodel	51		1.14 (0.89, 1.46)	0.293
	IDHwt	98		0.93 (0.68, 1.25)	0.616
Surgical Interval		167		1.00 (1.00, 1.00)	0.842



↑ **Figure 5-1. Radiotherapy is associated with an increased small deletion burden.**

a. Boxplot depicting the burden of newly acquired/post-treatment small deletions (deletions/Mb) in RT-naïve (*n* = 34) and RT-received (*n* = 156) patients from the GLASS cohort. Mann-Whitney *U* test was applied for statistical testing. **b.** Longitudinal comparison of small deletion burden between primary and recurrent glioma samples, separated by hypermutation (HM) and Radiotherapy (RT). Paired Wilcoxon signed-rank test was applied for statistical testing **c**. Forest plot showing multivariable log-linear regression model of newly acquired small deletion burden(deletions/Mb) including 1. TMZ- treatment, 2. Hypermutation, 3. RT-treatment, 4. Molecular subtype and 5. Surgical interval (in months) as variables. OR, Odds Ratio, CI, confidence interval. **d. Top**. Metastatic

cohort: Boxplots depicting small deletion burden (deletions/Mb) in RT-naïve (left), RT-treated with palliative intent (RT+ pal, middle) and RT-treated with curative intent (RT+ cur, right) tumor samples separated by primary tumor location. Kruskal-Wallis test was applied for statistical testing. **Bottom**. Sample sizes of metastatic cohort separated by primary tumor location.

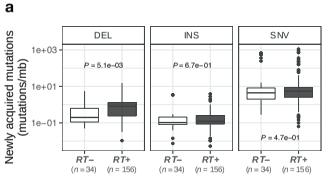
The increase in small deletions was particularly pronounced in the subset of gliomas marked by the presence of mutations in *IDH1* or *IDH2*, a clinically relevant subtype [61] predominantly consisting of grade II and III gliomas (**Figure 5-2b**, *P* = 1.4e-02, Mann-Whitney *U* test), while the number of RT-naive recurrent cases among IDH wild-type glioma was too small to test for differences (n = 2, vs n = 107 RT-treated cases). To ensure that these changes were not primarily due to TMZ-associated hypermutation (tumor mutation burden exceeding 10 mut/Mb at recurrence) [185], we stratified the cohort by hypermutation status. Hypermutation was associated with an increase in small deletions independent of RT-treatment, whereas amongst non-hypermutators only tumors from patients that received RT showed a significant increase in small deletions (**Figure 5-1b**, *P* = 5.0e-11, paired Wilcoxon signed-rank test), further implicating the observed increase in small deletions as a potential consequence of RT. To evaluate the independence of this finding from potential confounders, we fitted a multivariable log-linear regression model that included TMZ-treatment, glioma molecular subtype, time interval between surgeries and hypermutation. RT was independently associated with an increase in small deletions (**Figure 5-1c**, *P* = 3e-03, t-test), directly attributing the observed increase in small deletions to RT-treatment. We determined cancer cell fractions and found that post-treatment deletions in RT-treated patients did not show clonality differences compared to post-treatment deletions in RT-naïve patients, suggesting that these deletions were not more clonal/subclonal (**Figure 5-2c**, hypermutant: *P* = 9.3e-01, non-hypermutant: *P* = 8.7e-01, Mann-Whitney *U* test). Comparing the pre-treatment mutation burden and aneuploidy scores between glioma patients that acquired a high burden versus a low burden of post-treatment deletions revealed no significant differences, suggesting that these genomic characteristics of the pre-RT tumor are not predictive of the acquisition of small deletions in response to radiotherapy.

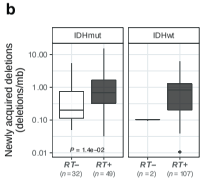
Importantly, 30% (41/136) of non-hypermutant samples gained more than 1 del/Mb following radiotherapy, whereas only 7% (2/27) of RT-naïve non-hypermutators acquired more than 1 del/Mb (*P* = 1.6e-02, Fisher's exact test). Among the samples treated with ionizing radiation, 35% (55/156) showed a doubling of the small deletion burden when compared with the primary tumor. The effect of RT

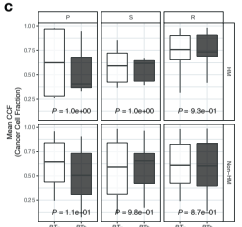
on mutational burden was significant for small deletions and not significant for other types of somatic mutations, such as insertions and sSNVs (**Figure 5-2d**). Conversely, TMZ-associated hypermutation was correlated with a significant increase in the burden of all types of mutations (**Figure 5-2d**).

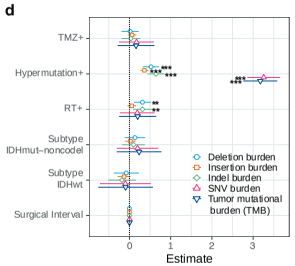
Following these observations, we hypothesized that radiotherapy may similarly increase the number of small deletions in other tumor types. To test this hypothesis, we evaluated whole genome sequencing-derived mutational profiles from 3693 metastatic tumors with complete treatment annotation (as described in **Figure 5-2e**), available via the Hartwig Medical Foundation (HMF) [205]. We separated tumors by site of origin and compared the small deletion burden between RT-treated and RT-naïve tumors (**Figure 5-2d**). RT-treated tumors were further stratified depending on whether the treatment intent was curative (RT+cur, n = 739) or palliative (RT+pal, n = 689), which generally differ in the applied cumulative dosage of ionizing radiation [209]. While this analysis was restricted to single time-point mutational profiles, we observed a significantly higher small deletion burden associated with curative RT in multiple tumor types, including bone/soft tissue (RT+cur: median 0.15 del/Mb, RT-: median 0.08 del/Mb, P = 6.2e-04, Kruskal-Wallis test), lung (RT+cur: median 0.56 del/Mb, RT-: median 0.43 del/Mb, P = 3.4e-03, Kruskal-Wallis test), and breast (RT+cur: median 0.18 del/Mb, RT-: median 0.12 del/Mb, P = 1.2e-04, Kruskal-Wallis test) cancers (**Figure 5-1d**). Further separation into tumor subtypes revealed that the observed patterns were present in both lung cancer types (**Figure 5-2f**, Non-small cell lung cancer: P = 6.9e-03 and small cell lung cancer: P = 6.0e-0.2, Kruskal-Wallis test), but in breast cancer these were restricted to ER-positive subtypes (**Figure 5-2f**, ER-positive/HER2-negative: P = 3.0e-04 and ER-positive/HER2-positive: P = 2.2e-02, Kruskal-Wallis test). Tumors treated palliatively with RT frequently presented an intermediate state in between the RT- and RT+cur cohorts, suggesting an association between RT-treatment derived small deletion burden and RT dose exposure.

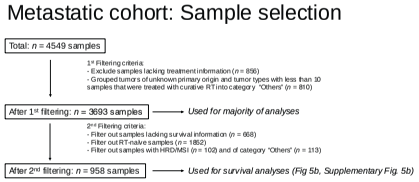
DNA repair deficient tumors were shown to harbor an increased mutational load [202]. To determine whether a DNA repair defective background has an impact on the small deletion burden in the HMF dataset, we derived information on microsatellite instability (MSI) and homologous recombination deficiency (HRD) in the HMF dataset from a recently published paper [210]. Notably, HRD+ and particularly MSI+ tumors harbored significantly more small deletions compared to samples that were HRD-/MSI- (**Figure 5-2g**, P < 2.2e-16, Kruskal-Wallis test).

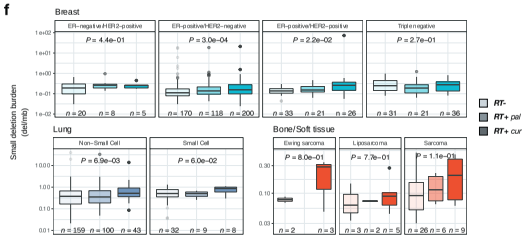

a
DEL
INS
SNV
Newly acquired mutations (mutations/mb)
P = 5.1e−03
P = 6.7e−01
P = 4.7e−01
RT− (n = 34)
RT+ (n = 156)

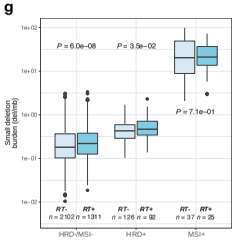

b
IDHmut
IDHwt
Newly acquired deletions (deletions/mb)
P = 1.4e−02
RT− (n = 32)
RT+ (n = 49)
RT− (n = 2)
RT+ (n = 107)

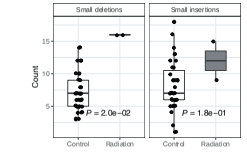

c
P
S
R
Mean CCF (Cancer Cell Fraction)
P = 1.0e+00
P = 1.0e+00
P = 9.3e−01
P = 1.1e−01
P = 9.8e−01
P = 8.7e−01
RT−
RT+

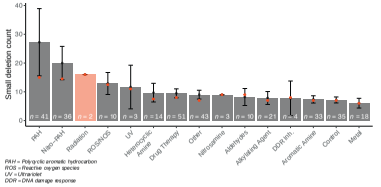

d
TMZ+
Hypermutation+
RT+
Subtype IDHmut−noncodel
Subtype IDHwt
Surgical Interval
Deletion burden
Insertion burden
Indel burden
SNV burden
Tumor mutational burden (TMB)
Estimate

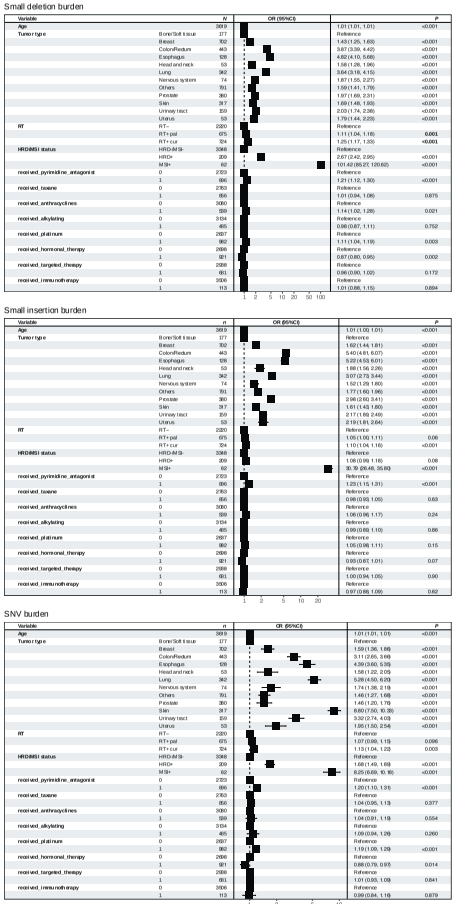

e
Metastatic cohort: Sample selection
Total: n = 4549 samples
1st Filtering criteria:
- Exclude samples lacking treatment information (n = 856)
- Grouped tumors of unknown primary origin and tumor types with less than 30 samples that were treated with curative RT into category "Others" (n = 810)
After 1st filtering: n = 3693 samples
Used for majority of analyses
2nd Filtering criteria:
- Filter out samples lacking survival information (n = 668)
- Filter out RT-naive samples (n = 1852)
- Filter out samples with HRD/MSI (n = 103) and of category "Others" (n = 113)
After 2nd filtering: n = 958 samples
Used for survival analyses (Fig 5b, Supplementary Fig. 5b)


f
Breast
ER−negative/HER2−positive
ER−positive/HER2−negative
ER−positive/HER2−positive
Triple negative
Small deletion burden (del/mb)
P = 4.4e−01
P = 3.0e−04
P = 2.2e−02
P = 2.7e−01
n = 20 n = 8 n = 5
n = 170 n = 118 n = 200
n = 33 n = 21 n = 26
n = 31 n = 21 n = 36
RT-
RT+ pal
RT+ cur
Lung
Non−Small Cell
Small Cell
P = 6.9e−03
P = 6.0e−02
n = 159 n = 100 n = 43
n = 32 n = 9 n = 8
Bone/Soft tissue
Ewing sarcoma
Liposarcoma
Sarcoma
P = 8.0e−01
P = 7.7e−01
P = 1.1e−01
n = 2 n = 3
n = 3 n = 2 n = 5
n = 26 n = 6 n = 9


g
Small deletion burden (del/mb)
P = 6.0e−08
P = 3.5e−02
P = 7.1e−01
RT- n = 2102
RT+ n = 1311
RT- n = 126
RT+ n = 92
RT- n = 37
RT+ n = 25
HRD-/MSI-
HRD+
MSI+


i
Small deletions
Small insertions
Count
P = 2.0e−02
P = 1.8e−01
Control
Radiation


k
Small deletion count
n = 41
n = 36
n = 5
n = 10
n = 3
n = 14
n = 51
n = 43
n = 3
n = 10
n = 21
n = 4
n = 33
n = 30
n = 18
HRD
Mismatch Repair
Radiation
ROS/DNS
UV
Homologous Recombination
Other
Nitrosamine
Alkylating Agent
Alkylating Agent
DDR inh.
Aromatic Amines
Control
Metal
PAH = Polycyclic aromatic hydrocarbon
ROS = Reactive oxygen species
UV = Ultraviolet
DDR = DNA damage response


h
Small deletion burden
Variable
N
OR (95%CI)
P
Age
3693
1.01 (1.01, 1.01)
<0.001
Tumor type
Bone/Soft tissue
177
Reference
Breast
702
1.43 (1.25, 1.63)
<0.001
Colon/Rectum
463
3.87 (3.39, 4.42)
<0.001
Esophagus
128
4.82 (4.10, 5.66)
<0.001
Head and neck
53
1.58 (1.28, 1.96)
<0.001
Lung
342
3.04 (3.18, 4.15)
<0.001
Nervous system
74
1.87 (1.55, 2.27)
<0.001
Others
761
1.69 (1.41, 1.79)
<0.001
Prostate
360
1.87 (1.66, 2.31)
<0.001
Skin
317
1.69 (1.48, 1.93)
<0.001
Urinary tract
359
2.03 (1.74, 2.38)
<0.001
Uterus
53
1.79 (1.44, 2.23)
<0.001
RT
RT−
2220
Reference
RT+ pal
675
1.11 (1.04, 1.18)
0.001
RT+ cur
304
1.25 (1.17, 1.33)
<0.001
HRD/MSI status
HRD-/MSI-
3368
Reference
HRD+
209
2.67 (2.42, 2.95)
<0.001
MSI+
62
101.40 (99.87, 1.0E+02)
<0.001
received_pyrimidine_antagonist
0
2533
Reference
1
896
1.21 (1.12, 1.30)
<0.001
received_taxane
0
2783
Reference
1
856
1.01 (0.94, 1.08)
0.875
received_anthracyclines
0
3080
Reference
1
599
1.14 (1.02, 1.28)
0.021
received_alkylating
0
3154
Reference
1
465
0.96 (0.87, 1.11)
0.752
received_platinum
0
2857
Reference
1
962
1.11 (1.04, 1.19)
0.003
received_hormonal_therapy
0
2956
Reference
1
921
0.87 (0.80, 0.95)
0.002
received_targeted_therapy
0
2888
Reference
1
891
0.96 (0.89, 1.02)
0.172
received_immunotherapy
0
3006
Reference
1
113
1.01 (0.88, 1.15)
0.894
Small insertion burden
Variable
n
OR (95%CI)
P
Age
3693
1.01 (1.00, 1.01)
<0.001
Tumor type
Bone/Soft tissue
177
Reference
Breast
702
1.62 (1.44, 1.81)
<0.001
Colon/Rectum
463
5.40 (4.81, 6.07)
<0.001
Esophagus
128
5.29 (4.59, 6.01)
<0.001
Head and neck
53
1.88 (1.56, 2.26)
<0.001
Lung
342
3.07 (2.73, 3.44)
<0.001
Nervous system
74
1.52 (1.20, 1.90)
<0.001
Others
761
1.77 (1.60, 1.96)
<0.001
Prostate
360
2.66 (2.40, 3.41)
<0.001
Skin
317
1.61 (1.43, 1.80)
<0.001
Urinary tract
359
2.17 (1.93, 2.46)
<0.001
Uterus
53
2.19 (1.81, 2.64)
<0.001
RT
RT−
2220
Reference
RT+ pal
675
1.05 (1.00, 1.11)
0.06
RT+ cur
304
1.13 (1.04, 1.16)
0.001
HRD/MSI status
HRD-/MSI-
3368
Reference
HRD+
209
1.09 (0.99, 1.19)
0.08
MSI+
62
30.70 (29.46, 35.60)
<0.001
received_pyrimidine_antagonist
0
2533
Reference
1
896
1.23 (1.15, 1.31)
<0.001
received_taxane
0
2783
Reference
1
856
0.99 (0.93, 1.05)
0.63
received_anthracyclines
0
3080
Reference
1
599
1.09 (0.99, 1.17)
0.24
received_alkylating
0
3154
Reference
1
465
0.99 (0.89, 1.10)
0.86
received_platinum
0
2857
Reference
1
962
1.05 (0.99, 1.11)
0.15
received_hormonal_therapy
0
2956
Reference
1
921
0.93 (0.87, 1.01)
0.07
received_targeted_therapy
0
2888
Reference
1
891
1.00 (0.94, 1.05)
0.90
received_immunotherapy
0
3006
Reference
1
113
0.97 (0.86, 1.09)
0.62
SNV burden
Variable
n
OR (95%CI)
P
Age
3693
1.01 (1.01, 1.01)
<0.001
Tumor type
Bone/Soft tissue
177
Reference
Breast
702
1.59 (1.36, 1.86)
<0.001
Colon/Rectum
463
3.11 (2.65, 3.66)
<0.001
Esophagus
128
4.39 (3.60, 5.30)
<0.001
Head and neck
53
1.58 (1.22, 2.05)
<0.001
Lung
342
5.26 (4.50, 6.03)
<0.001
Nervous system
74
1.74 (1.38, 2.19)
<0.001
Others
761
1.46 (1.27, 1.68)
<0.001
Prostate
360
1.46 (1.20, 1.79)
<0.001
Skin
317
8.60 (7.50, 10.38)
<0.001
Urinary tract
359
3.32 (2.74, 4.03)
<0.001
Uterus
53
1.95 (1.50, 2.54)
<0.001
RT
RT−
2220
Reference
RT+ pal
675
1.07 (0.99, 1.13)
0.066
RT+ cur
304
1.13 (1.04, 1.23)
0.003
HRD/MSI status
HRD-/MSI-
3368
Reference
HRD+
209
1.60 (1.49, 1.69)
<0.001
MSI+
62
8.25 (6.69, 10.16)
<0.001
received_pyrimidine_antagonist
0
2723
Reference
1
866
1.20 (1.10, 1.31)
<0.001
received_taxane
0
2763
Reference
1
856
1.04 (0.95, 1.13)
0.377
received_anthracyclines
0
3080
Reference
1
599
1.04 (0.91, 1.19)
0.554
received_alkylating
0
3154
Reference
1
465
1.09 (0.94, 1.26)
0.260
received_platinum
0
2857
Reference
1
962
1.19 (1.06, 1.29)
<0.001
received_hormonal_therapy
0
2956
Reference
1
921
0.98 (0.79, 0.97)
0.014
received_targeted_therapy
0
2888
Reference
1
891
1.01 (0.93, 1.09)
0.841
received_immunotherapy
0
3006
Reference
1
113
0.96 (0.84, 1.10)
0.879

← **Figure 5-2. Supplementary small deletion burden and multivariable analysis.**

a. Boxplot depicting the burden of newly acquired/post-treatment mutations (mutations/mb) in RT-naïve (n = 34) and RT-received (n = 156) patients from the GLASS cohort. Mutations were separated by DEL (small deletions), INS (small insertions) and SNV (single nucleotide variants). Mann-Whitney U test was applied for statistical testing. **b.** Comparison of newly acquired small deletion burden between RT-naïve and RT-received cases separated by molecular subtype. Mann Whitney U test was applied for statistical testing. **c.** Comparison of mean cancer cell fraction of small deletions per patient separated by P, primary only fraction (pre-treatment), S, shared fraction (pre-treatment) and R, recurrence only fraction (post-treatment) and by HM, hypermutation. Mann-Whitney U test was applied for statistical testing. d. Top. Metastatic cohort: Boxplots depicting small deletion burden (deletions/Mb) in RT-naïve (left), RT-treated with palliative intent (RT+ pal, middle) and RT-treated with curative intent (RT+ cur, right) tumor samples separated by primary tumor location. Kruskal-Wallis test was applied for statistical testing. Bottom. Sample sizes of metastatic cohort separated by primary tumor location. **d.** Forest plots showing multivariable log-linear regression model of newly acquired mutation burden (mutations/mb) including 1. TMZ-treatment, 2. Hypermutation, 3. RT-treatment, 4. Molecular subtype and 5. Surgical interval (in months) as variables. Mutation types were separated into small deletions, small insertions, indels (small deletions + small insertions), SNVs (single nucleotide variants) and overall tumor mutational burden (TMB, small indels + SNVs). Point indicates mean estimate of the model and lines showing 95 % confidence interval. Note that hypermutation is significantly associated with increased burden of all types of mutations and RT is associated with slightly increased burden of small deletions and indels (potentially driven by the large effect size of small deletions), specifically. **e.** Sample selection and filtering criteria for metastatic cohort including a detailed description of the usage for specific figures. **f.** Separation of lung, breast and bone/soft tissue cancers into their respective subtypes. Comparison of small deletion burden between RT-, RT+ pal and RT+ cur samples, respectively. Kruskal-Wallis test was applied for statistical testing. **g.** Boxplots depicting the burden of small deletions in HRD-/MSI- (n = 3,413), HRD+ (n = 218) and MSI+ (n = 62) samples from the HMF cohort separated by the radiotherapy treatment status. Mann-Whitney U test was applied for statistical testing. **h**. Forest plots depicting multivariable log-linear regression model for mutation burdens in the metastatic cohort. Mutations were separated into small deletions, small insertions and SNVs. Independent variables include age, tumor type (primary tumor location), DNA repair deficiency background and various treatment types including radiotherapy, taxane, alkylating agents, platin and others. **i.** Comparison of small deletion counts between control vs ionizing radiation groups. Mann-Whitney U test was used for statistical testing. **k.** Distribution of small deletion counts per treatment group. Bars indicate means, error bars reflect standard deviation and red dots indicate median count of small deletions. Note that ionizing radiation group displays highest median counts of small deletions.

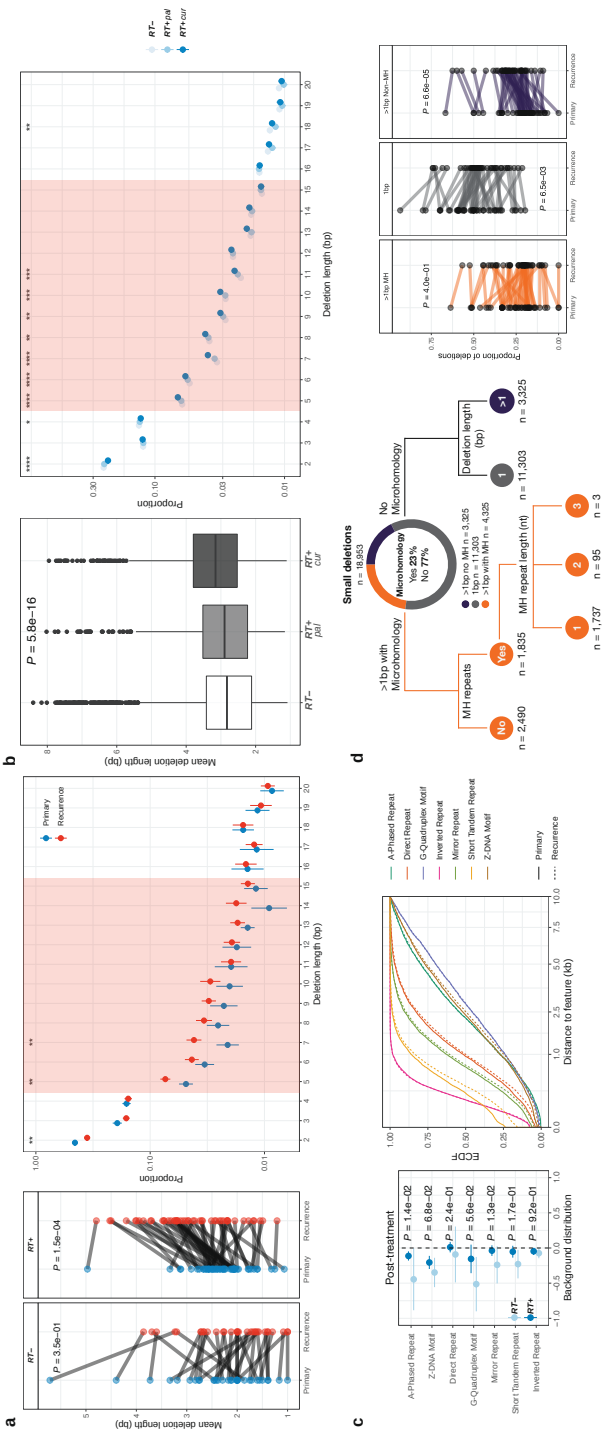

a
RT−
RT+
P = 3.5e−01
P = 1.5e−04
Mean deletion length (bp)
Primary
Recurrence
Proportion
Deletion length (bp)
b
P = 5.8e−16
Mean deletion length (bp)
RT−
RT+ pal
RT+ cur
Proportion
Deletion length (bp)
c
Post-treatment
P = 1.4e−02
P = 6.8e−02
P = 2.4e−01
P = 5.6e−02
P = 1.3e−02
P = 1.7e−01
P = 9.2e−01
A-Phased Repeat
Z-DNA Motif
Direct Repeat
G-Quadruplex Motif
Mirror Repeat
Short Tandem Repeat
Inverted Repeat
Background distribution
ECDF
Distance to feature (kb)
Primary
Recurrence
d
Small deletions
n = 18,953
Microhomology
Yes 23%
No 77%
No Microhomology
>1bp with Microhomology
MH repeats
Yes
No
n = 1,835
n = 2,490
>1bp no MH n = 3,325
1bp n = 11,303
>1bp with MH n = 4,325
Deletion length (bp)
1
>1
n = 11,303
n = 3,325
MH repeat length (nt)
1
2
3
n = 1,737
n = 95
n = 3
Proportion of deletions
>1bp MH
1bp
>1bp Non-MH
P = 4.0e−01
P = 6.5e−03
P = 6.6e−05
Primary
Recurrence

Figure 5-3. Distribution of small deletion characteristics.

a. Length distribution in GLASS. *Left:* Comparison of mean deletion lengths in primary vs recurrent IDH mutant glioma (*n* = 81), separated by RT-treatment (RT+, *n* = 32, RT-, *n* = 49). Paired Wilcoxon signed-rank test was applied for statistical testing. Note a significant increase in mean deletion lengths only for ionizing radiation treated samples. *Right:* Y-Axis indicating proportion of deletions and X-Axis reflecting deletion length >1bp. Proportions calculated for each patient, mean (point) and 95%-CI (line-range) compared between Primary and Recurrence in non-hypermutant glioma treated with RT (*n* = 44). Paired Wilcoxon signed-rank test was applied for statistical testing (* = p < 0.05, ** = p < 0.01). **b.** Length distribution metastatic cohort. *Left:* Comparison of mean deletion lengths in RT-naïve vs RT+pal vs. RT+cur samples. Kruskal-Wallis test was applied for statistical testing. *Right:* Y-Axis indicating proportion of deletions and X-Axis reflecting deletion length >1bp. Proportions calculated for each patient, mean (point) and 95%-CI (line-range) compared between RT-naive vs. RT+pal vs. RT+cur samples. Kruskal-Wallis test was applied for statistical testing (* = p < 0.05, ** = p < 0.01, *** = p < 0.001, **** = p < 0.0001). **c.** Relation to genomic features in GLASS. *Left:* Distribution of deletions in relation to genomic features. Y-Axis: non-B DNA genomic feature, X-Axis: Log10 ratio of mean distance of non-radiation-associated and radiation-associated post-treatment deletions to genomic feature over background distribution in non-hypermutated glioma samples (*n* = 69). Distribution of radiation-associated deletions shows little variability (narrow 95% CI) and resemble background distribution more closely (closer to 0). Significant differences between radiation-associated deletions and non-radiation-associated deletions seen in relation to repeats (Mann-Whitney *U* test). *Right:* Empirical cumulative distribution function (ECDF, Y-Axis) of distance to non-B DNA features in kb (X-Axis) reveals right-shift towards larger distances in post-radiated non-hypermutated recurrent samples (*n* = 44). Note that neither in hypermutated, nor in RT-naïve non-hypermutated glioma samples longitudinal differences were observed (**Figure 5-4**c)**. d.** Small-deletion categories in GLASS. *Left:* separation of small deletions in the GLASS cohort into 3 major categories: 1bp (gray), >1bp without microhomology (MH; purple) and >1bp with microhomology (orange) in IDH-mutant gliomas (n = 81). The microhomology category was further classified based on the occurrence of microhomology repeat sequences and length of repeats. nt, nucleotide. *Right:* a comparison of the proportion of deletions for each RT+ non-hypermutated glioma sample (n = 44) using a two-sided paired Wilcoxon signed-rank test.

RT-treatment was associated with an increase in small deletion burden in HRD-/MSI- (**Figure 5-2g**, *P* = 6.0e-08, Mann-Whitney *U* test) and HRD+ tumors (*P* = 3.5e-02), but not in MSI+ tumors (*P* = 7.1e-01). These results raised the possibility that DNA repair deficiencies like HRD and MSI confounded the association between RT-treatment and the small deletion burden. To address this, we have included HRD and MSI status into the multivariable log-linear regression analysis, which showed that RT-treatment is associated with an increase in the small deletion burden independent of a number of potential confounders, including a DNA repair defective background (**Figure 5-2h**).

Next, we assessed whether the small deletion burden was associated with mutations in selected genes (*ATM*, *ATR*, *CHEK1*, *CHEK2*, *PARP1*, *PRKDC*, *TP53* and *WEE1*) in-

volved in the DNA damage response (DDR). This analysis indicated that DDR mutations were universally associated with a significantly higher small deletion burden.

We used log-linear regression to adjust for potential confounding variables, including age, tumor type, DNA damage repair background, DRR gene mutations and various cytotoxic treatment regimens (e.g. taxane, platinum, anthracyclines, alkylating agents) that have been previously associated with increased mutation burdens [206]. Results from this analysis revealed a robust association for both palliative and curative radiotherapy treatment but not any other therapy and small deletions (**Figure 5-2h**, RT+cur vs. RT-naïve: odds ratio = 1.25, $P < 1e-03$, t-test). This confirmed that the increased small deletion burden associated with radiotherapy is independent of tumor type, HRD, MSI, DDR gene mutations or additional cytotoxic therapy.

To verify the causal association between RT and acquired small deletions, we re-analyzed a previously published dataset [207] consisting of whole-genome sequencing data from 324 human-induced pluripotent stem cells (iPSCs) exposed to known or suspected environmental carcinogens, including two iPSCs treated with ionizing radiation. We found that the small deletion burden was significantly higher in the RT-treated iPSCs compared to controls (**Figure 5-2i**, $P = 2.0e-02$, Mann-Whitney *U* test). In contrast, we did not observe a significant difference in small insertion burden ($P = 1.8e-01$). Strikingly, the ionizing radiation group showed the highest median burden of small deletions across all treatment modalities, further substantiating our human tissue analysis (**Figure 5-2k**).

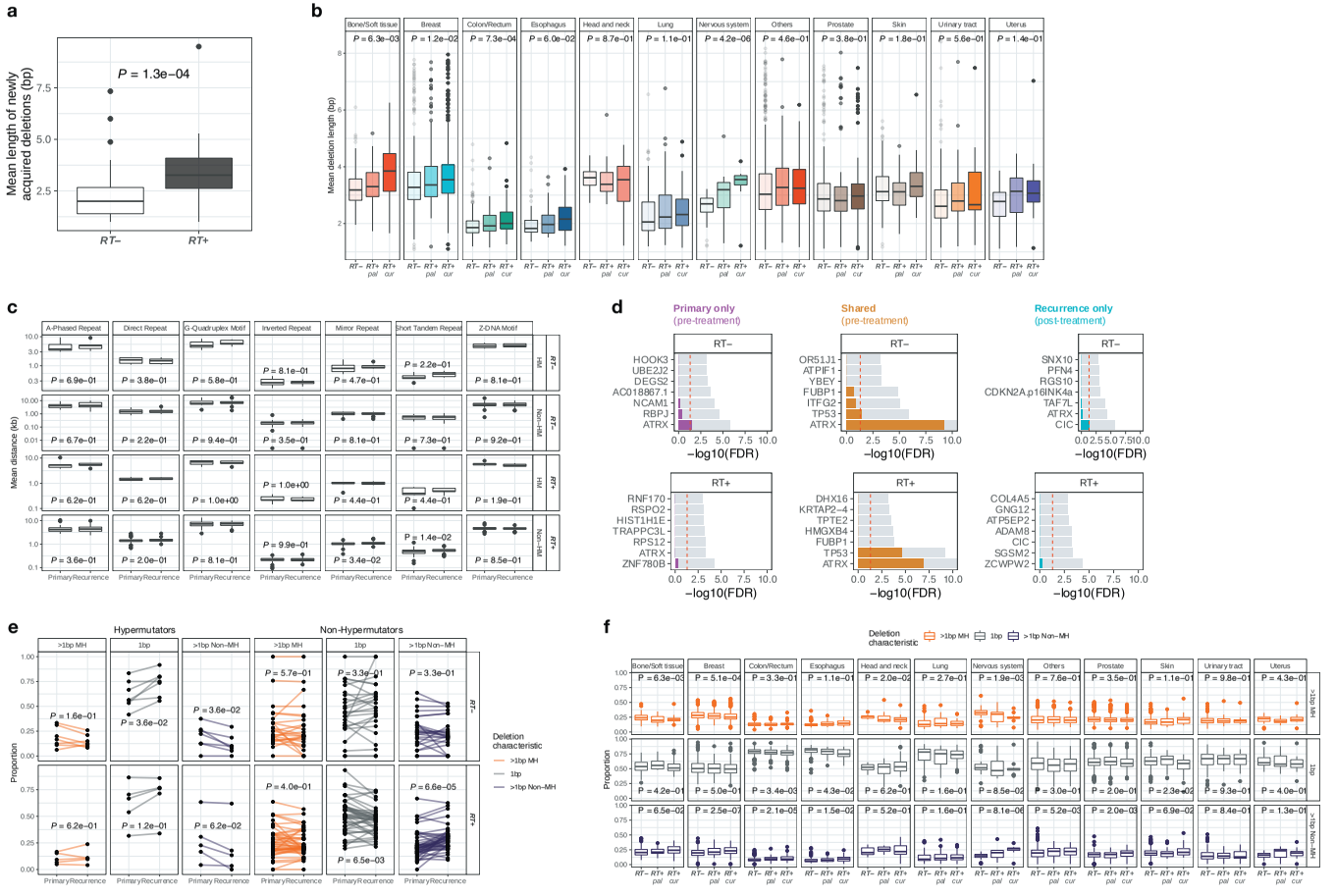
Radiotherapy-associated small deletions harbor a characteristic genomic signature

Characteristics of RT-associated small deletions, such as length distribution and breakpoint microhomology, may be able to provide insights on their etiology. We explored such features in the GLASS dataset, limiting the analysis to IDH mutant gliomas (RT+, n=49; RT-, n=32) due to the imbalance in radiation treatment amongst IDH wild-type gliomas in the GLASS cohort (RT+, n=107 vs RT-, n=2). Small deletions in recurrent tumor samples that were treated with RT showed increased deletion lengths (**Figure 5-3a**, left, RT+: $P = 1.5e-04$; RT-: $P = 3.5e-01$, paired Wilcoxon signed-rank test). This increase was particularly associated with new deletions that occurred after therapy (**Figure 5-4a**, $P = 1.3e-04$, Mann-Whitney *U* test), supporting the idea that RT leads to longer deletion lengths (**Figure 5-3a**, left). Moreover, a detailed analysis of the size distribution of deletions revealed a shift towards deletions of length ~5-15bp following RT-treatment in non-hypermutated gliomas (**Figure 5-3a**, right).

Comparing RT-treated and RT-naïve metastatic tumor samples from the single time-point HMF dataset showed a similar larger average deletion length for both palliative and curative RT-treated tumor samples (**Figure 5-3b**, **Figure 5-4b**). Moreover, we also observed a shift in deletion span from small 1-4 bp deletions towards medium-sized 5-15 bp deletions (**Figure 5-3b**). A stepwise increase in deletion length was observed for palliative and curative RT-treatment, respectively, providing further evidence for a dose and exposure association. Taken together, these results indicate not only an increased deletion burden, but also highlight distinct characteristics of RT-associated small deletions.

B-DNA is the common right-handed, double helical formation of DNA. Non-canonical non-B-DNA structures and fragile repeat-rich DNA may be more prone to acquiring mutations [211]. Therefore, we hypothesized that RT-induced deletions were more likely to occur in these fragile regions of the genome. We investigated the link between small deletions and these genomic features by adjusting for a random background distribution. Importantly, deletions following RT showed less variability and higher similarity to the random background distribution compared to non-RT-induced deletions (**Figure 5-3c**, left). Furthermore, comparison of GLASS pre- and post-treatment deletions indicated that deletions following radiotherapy showed larger distances to non-B DNA features (**Figure 5-3c**, right, **Figure 5-4c**). The lack of or reduced association between radiation-induced deletions and the analyzed genomic features, such as repeats and G-quadruplex motifs, suggests that ionizing radiation associated small deletions occur in a largely stochastic manner, occurring independently from the intrinsic mutagenicity of the fragile regions of the genome analyzed.

We assessed whether RT-associated small deletions showed enrichment in driver genes. We computed the covariate-adjusted normalized ratio between non-synonymous and synonymous mutations (dN/dS) in order to identify selection of mutations at the level of individual genes separately for GLASS pre- and post-treatment fractions (**Figure 5-4d**) [212]. Genes with dN/dS ratios strongly deviating from one are thought to be under selection and may be associated with the RT+ small deletion phenotype. We did not find evidence for significant selection for any genes in the post-treatment fraction following radiation therapy. As these analyses focused on acquired deletions only present after RT, we were unable to also perform them in the HMF set where pre-treatment samples are unavailable. Our results in IDH-mutant glioma further support the notion that RT-associated deletions do not occur at particular genomic loci, instead showing a more dispersed genomic spread.


a
Mean length of newly acquired deletions (bp)
P = 1.3e−04
RT−
RT+
b
Mean deletion length (bp)
Bone/Soft tissue
Breast
Colon/Rectum
Esophagus
Head and neck
Lung
Nervous system
Others
Prostate
Skin
Urinary tract
Uterus
c
Mean distance (kb)
A-Phased Repeat
Direct Repeat
G-Quadruplex Motif
Inverted Repeat
Mirror Repeat
Short Tandem Repeat
Z-DNA Motif
MH
Non−MH
RT−
RT+
PrimaryRecurrence
d
Primary only
(pre-treatment)
Shared
(pre-treatment)
Recurrence only
(post-treatment)
HOOK3
UBE2J2
DEGS2
AC018867.1
NCAM1
RBPJ
ATRX
OR51J1
ATP1F1
YBEY
FUBP1
ITFG2
TP53
ATRX
SNX10
PFN4
RGS10
CDKN2A.p16INK4a
TAF7L
ATRX
CIC
RNF170
RSPO2
HIST1H1E
TRAPPC3L
RPS12
ATRX
ZNF780B
DHX16
KRTAP2−4
TPTE2
HMGXB4
FUBP1
TP53
ATRX
COL4A5
GNG12
ATP5EP2
ADAM8
CIC
SGSM2
ZCWPW2
−log10(FDR)
e
Hypermutators
Non-Hypermutators
>1bp MH
1bp
>1bp Non−MH
Proportion
Deletion characteristic
f
Deletion characteristic
>1bp MH
1bp
>1bp Non−MH
Bone/Soft tissue
Breast
Colon/Rectum
Esophagus
Head and neck
Lung
Nervous system
Others
Prostate
Skin
Urinary tract
Uterus
Proportion
>1bp MH
1bp
>1bp Non−MH
RT−
RT+ pal
RT+ cur

← **Figure 5-4. Supplementary characteristics of RT-associated deletions.**

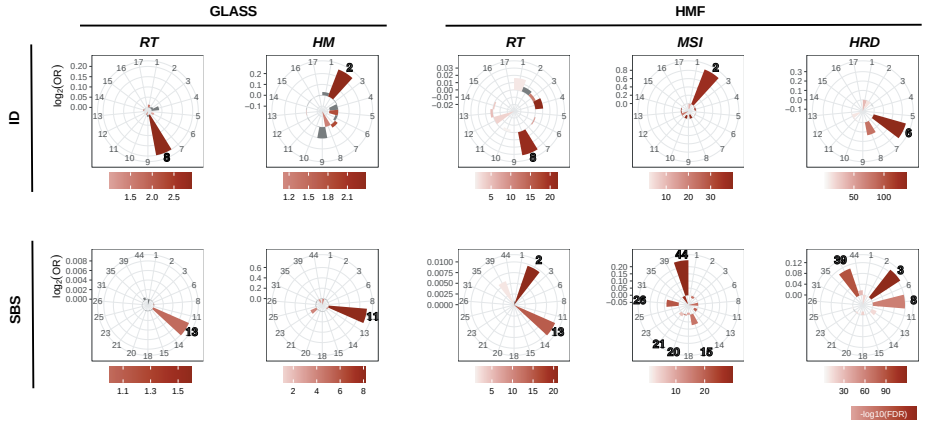
a. Comparison of mean deletion lengths of newly acquired deletions (post-treatment fraction) in RT- vs RT+ IDHmut glioma samples. Mann-Whitney U test was applied for statistical testing. **b.** Metastatic cohort: Boxplots depicting mean deletion lengths in RT-naïve (left) and palliative RT-treated (middle) and curative RT-treated (right) tumor samples separated by primary tumor location. Kruskal-Wallis test was applied for statistical testing. **c.** Longitudinal comparison (X-Axis) of mean distances of deletions non-B DNA features in kb (Y-Axis) in IDHmut glioma cases. Cases separated by radiation treatment and hypermutation. Note that neither in hypermutated, nor in RT-naïve non-hypermutated glioma samples significant longitudinal differences were observed. Paired Wilcoxon signed-rank test was applied for statistical testing. **d.** Gene-wise dN/dS estimates by radiation treatment (rows) and fraction (columns). Genes are sorted by Q value (Bonferroni adjusted P value) and P value. Q values indicated in color, whereas the P value are shown in light grey. The Q value threshold of 0.05 is indicated by a horizontal red line. Note that no genes show significant selection in the post-radiation fraction. **e.** Comparison of proportion of deletions for IDHmut glioma samples separated by radiation treatment and hypermutation using paired Wilcoxon signed-rank test. For each sample, the proportion of deletions with 1bp length, > 1bp length with microhomology and > 1bp length without microhomology add up to 1. Bottom right panels (RT-received non-hypermutators) presented in **Figure 5-3**d and shown here for comparison with other groups. Note the significant increase of 1bp deletions in hypermutated (radiation-naïve) cases. No significant differences were observed for radiation-naïve non-hypermutated cases. **f.** Comparison of proportion of deletions in metastatic cohort between RT-treated and RT-naïve cases using Kruskal-Wallis test. In bone/soft tissue, breast and head & neck and nervous system cancers, significantly lower proportions of deletions >1bp with microhomology were observed in RT-treated samples compared to RT-naïve samples. In contrast, RT-received breast, colon/rectum, esophagus, nervous system and prostate tumor samples showed significantly higher proportions in deletions > 1bp without microhomology.

Small deletions can be the result of error-prone DSB-repair utilizing mechanisms such as c-NHEJ and a-EJ/MMEJ [201]. To investigate whether a preference for DSB-repair pathway choice following RT exists, we computed microhomology sequences at breakpoints and characterized deletions based on size, microhomology and repeat content (**Figure 5-3d**, **Figure 5-4e**). Deletions without microhomology comprised the majority of deletions in the dataset (77%, **Figure 5-3d**). However, in non-hypermutant gliomas receiving ionizing radiation we observed a significant increase in > 1bp deletions without microhomology (**Figure 5-3d**, P = 6.6e-05, Paired Wilcoxon signed-rank test) and conversely a decrease in 1bp-deletions (**Figure 5-3d**, P = 6.5e-03, Paired Wilcoxon signed-rank test).

Using the same three categories described in **Figure 5-3d**, comparison of RT-treated and RT-naïve metastatic tumors from the HMF dataset demonstrated comparable results (**Figure 5-4f**). These data suggest that c-NHEJ is the preferred pathway for repairing radiation-induced DNA damage.

Distinct ID and SBS mutational signatures associated with radiotherapy

Cancer cells accumulate somatic mutations that are caused by intrinsic and/or extrinsic mechanisms. The different mutational processes can leave distinct genomic scars, termed mutational signatures.



↑ **Figure 5-5. ID8 and APOBEC-SBS signatures associated with radiotherapy.**

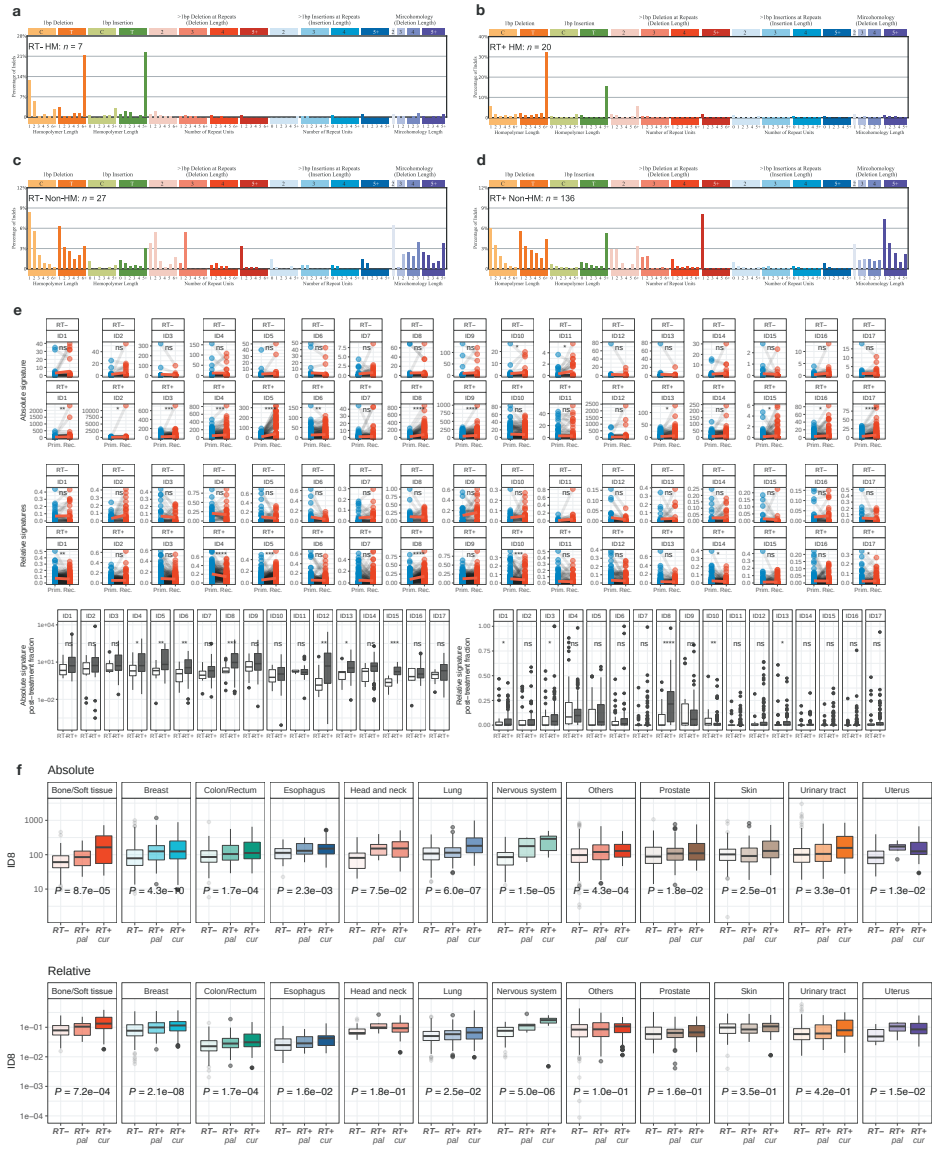
Indel (ID) and single base substitution (SBS) mutational signatures in the GLASS and HMF cohorts associated with RT (radiotherapy), Hypermutation (HM), Microsatellite instability (MSI) and homologous recombination deficiency (HRD). Mann-Whitney U test was applied for statistical testing and false discovery rate (FDR) correction was used to adjust for multiple testing. Bars in the petal plots not reaching statistical significance (defined as FDR < 0.01) are indicated in grey.

To validate the underlying mutational processes of radiotherapy, we compared pre-and post-treatment mutations in the GLASS dataset to previously defined mutational signatures [213]. The comparison of signature contributions between post-treatment mutations in RT-treated and RT-naïve IDH mutant glioma samples revealed a strong enrichment of indel signature 8 (ID8, **Figure 5-5**, **Figure 5-6d**, RT+, mean contribution = 0.22, vs. RT-, mean contribution, *P* = 7.4e-05, *Q* = 3.8e-03, Mann-Whitney *U* test and false discovery rate, respectively). Furthermore, in RT-treated patients but not RT-naïve patients comparing ID8 values before and after treatment resulted in significant increases in absolute (**Figure 5-6e**, *P* = 4.5e-07, Paired Wilcoxon rank-signed test) and relative (**Figure 5-6e**, *P* = 2.3e-03) ID8 contributions, respectively. Signature ID8 is composed of ≥ 5 bp deletions without microhomology and has previously been linked to DSB repair by c-NHEJ, providing further evidence that radiation-induced DSBs are primarily repaired via c-NHEJ [213]. Hypermutation in IDH mutant gliomas was associated with a

significant enrichment of indel signature 2 (ID2, **Figure 5-5**, **Figure 5-6a-b**). ID2 comprises 1-bp deletions at homopolymers and has been reported previously to be elevated in DNA mismatch repair deficient cancers [213].

A previous analysis of mutational signatures in the HMF dataset found that of all indel signatures the strongest association with radiotherapy treatment was with COSMIC-ID6 (corresponding to SignatureAnalyzer-ID12) [206]. In line with our findings in the GLASS cohort, we observed that the strongest association was with ID8, and significant but substantially less pronounced for ID6 (**Figure 5-5**). Both absolute and relative ID8 values were significantly higher in RT-treated samples when compared to RT-naïve samples, and a significant association was observed in nine of twelve tumor types (**Figure 5-6f**). Importantly, the comparison of HRD+ and HRD- samples showed a clear association of HR-deficiency with ID6. ID6 comprises > 5 bp deletions with microhomology at breakpoints and has been reported previously to be elevated in HR-defect breast cancers [214]. Additionally, we validated the findings from the GLASS cohort and previous observations that MSI samples were enriched for indel signature 2 (ID2, **Figure 5-5**).

Collectively, these results have important implications for differential mutational processes acting on the cancer genome. While MSI leads to an increased burden in small deletions due to hypermutability resulting from impaired DNA mismatch repair at microsatellites/homopolymers, the DSBs induced by RT and due to HRD are repaired via error prone DSB-repair mechanisms. Our results clearly suggest two different mechanisms for the repair of DSBs: the a-EJ pathway that utilizes microhomologies at breakpoints in HR-deficient samples (signature ID6) and c-NHEJ pathway that does not require microhomologies at breakpoints in RT-treated samples (signature ID8).



↑ **Figure 5-6. Mutational signatures of RT-associated deletions.**

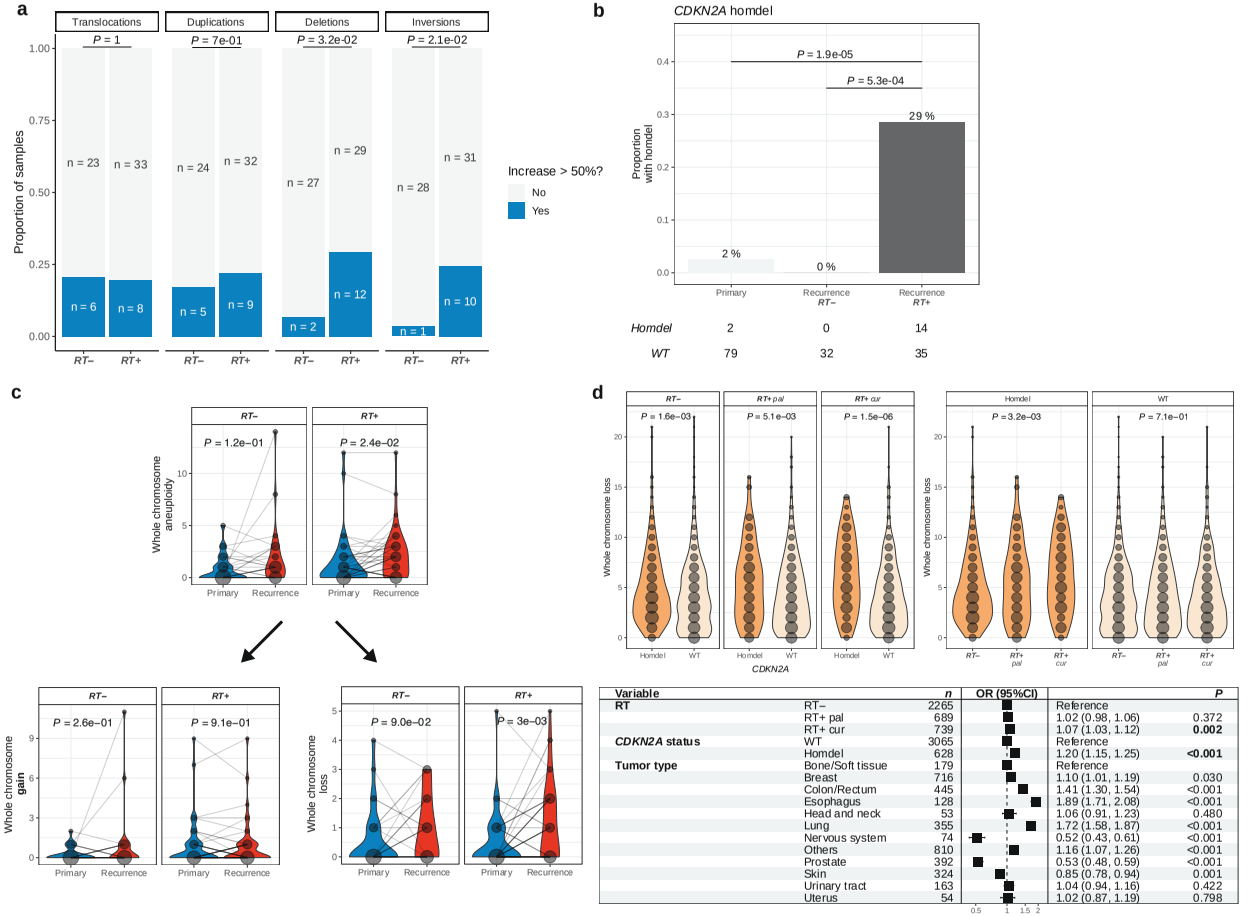
a-d. Distribution of indel types for post-treatment mutations in the GLASS cohort, separated by RT (a, c, RT-negative; b, d, RT-treated) and HM (a-b, Hypermutator; c-d, Non-Hypermutator). Note that patterns of indels in hypermutated samples resemble the previously identified MSI signature ID2, whereas RT-treated Non-Hypermutant gliomas harbor large similarities with ID8. Sample sizes for each subgroup are annotated. **e.** Comprehensive comparison of all 17 COSMIC indel (ID) signatures in IDHmut glioma. Top 2 panels display longitudinal comparison of absolute

signature contributions separated by radiation treatment (RT+ and RT-). Middle 2 panels display longitudinal comparison of relative signature contributions separated by radiation treatment. For these panels paired Wilcoxon signed-rank test was applied for statistical testing. Bottom panels display comparison of absolute (left) and relative (right) signatures of post-treatment indels between RT-treated and RT-naïve samples. For these panels Mann-Whitney U test was applied for statistical testing. (ns = not significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$). Note that ID8 is the only signature consistently associated with radiation therapy across different comparisons, nominating it as a robust signature of radiotherapy. **f.** Absolute (top) and relative (bottom) contribution of ID8 signature in metastatic cohort compared between cases with prior radiation treatment and cases without prior radiation treatment separated by tumor types. Note that most tumor types show significantly higher values of the signature in curative RT+ cases. Kruskal-Wallis test was applied for statistical testing.

In a next step, we sought to identify single base substitution (SBS) signature associations in both datasets. In line with previous reports, we found significant enrichment of SBS11 in hypermutant IDH mutant glioma samples [185, 215] and enrichment of signatures SBS44, SBS26, SBS21, SBS20 and SBS15 in MSI samples [213]. In HRD cases we observed enrichment of SBS3 and SBS8 which were previously described [213, 214, 216] along with a so far unreported enrichment of SBS39 (**Figure 5-5**).

In addition to these confirmatory results, we found several previously unreported associations suggesting the involvement of APOBEC in RT-associated DSB-repair. In the GLASS cohort, RT-treatment was significantly associated with SBS13 and in the HMF cohort with SBS2 and SBS13. SBS2 and SBS13 are APOBEC signatures [213, 217]. APOBEC cytosine deaminases are involved in retrovirus and retrotransposon restriction, and the enrichment of APOBEC signatures in RT-treated samples in both datasets implicated APOBEC-mediated mutagenesis in association with RT-associated DSB-repair [218-220].

Our results support the hypothesis that mutational signatures are shaped by cycles of DNA damage and DNA repair [221]. While extrinsic radiotherapy causes DNA DSBs, the intrinsic erroneous repair of the damage via c-NHEJ leads to specific small deletions (ID8), and APOBEC cytidine deaminases may be activated during the repair process leading to specific SNVs (SBS2/SBS13).



Variable		n	OR (95%CI)	P
RT	RT–	2265	Reference	
	RT+ pal	689	1.02 (0.98, 1.06)	0.372
	RT+ cur	739	1.07 (1.03, 1.12)	**0.002**
***CDKN2A* status**	WT	3065	Reference	
	Homdel	628	1.20 (1.15, 1.25)	**<0.001**
Tumor type	Bone/Soft tissue	179	Reference	
	Breast	716	1.10 (1.01, 1.19)	0.030
	Colon/Rectum	445	1.41 (1.30, 1.54)	<0.001
	Esophagus	128	1.89 (1.71, 2.08)	<0.001
	Head and neck	53	1.06 (0.91, 1.23)	0.480
	Lung	355	1.72 (1.58, 1.87)	<0.001
	Nervous system	74	0.52 (0.43, 0.61)	<0.001
	Others	810	1.16 (1.07, 1.26)	<0.001
	Prostate	392	0.53 (0.48, 0.59)	<0.001
	Skin	324	0.85 (0.78, 0.94)	0.001
	Urinary tract	163	1.04 (0.94, 1.16)	0.422
	Uterus	54	1.02 (0.87, 1.19)	0.798

← **Figure 5-7. RT associated with increase in large deletions and inversions.**

a. Analysis of structural variants (SVs) after RT in IDHmut glioma samples with sufficient quality for calling (n = 70): Translocations, Duplications, Deletions, Inversions. For each patient, number of SVs were calculated pre-and post-treatment. Based on the distribution of percent increase from primary to recurrence, cutoff was set for > 50% increase (**Figure 5-4a**). Comparison of proportion of samples with/without increase of given SVs between RT-treated vs RT-naïve. Fisher's exact test was applied for statistical testing. **b.** Depicted are proportions of IDHmut glioma samples (n = 81) harboring a homozygous deletion in CDKN2A. Using Fisher's exact test, proportions were compared between RT-received recurrence (RT+) vs. RT-naïve recurrence (RT-) and RT-received recurrence (RT+) vs. samples prior to treatment (Primary). Detailed distributions of whole chromosome deletion scores are provided in **Figure 5-4f**. **c.** *Upper*: Longitudinal comparison of whole chromosome aneuploidy scores separated by RT-treatment for IDHmut glioma samples with sufficient quality for calling and complete treatment annotation (total *n* = 69, RT-treated *n* = 42, RT-naïve *n* = 27). *Bottom*: Separation of whole chromosome aneuploidy into whole chromosome gain (left) and whole chromosome loss (right) scores, respectively. Note that the increase of whole chromosome aneuploidy in RT-treated samples is associated with whole chromosome losses. Dots are proportional to the frequency of whole chromosome loss integer for each subgroup. Paired Wilcoxon rank-signed test was applied for statistical testing. ***d.*** *Upper:* Comparison of whole chromosome deletion scores between RT-naïve vs RT+pal vs RT+cur and/or *CDKN2A* homdel vs. WT samples. Note that *CDKN2A* homdel is associated with higher whole chromosome deletion scores, independent of RT. Within samples with *CDKN2A* homdel, samples that were RT-treated with curative intent show the highest deletion scores. Dots are proportional to the frequency of whole chromosome loss integer for each subgroup. Kruskal-Wallis test was applied for statistical testing. Detailed distributions of whole chromosome deletion scores are provided in **Figure 5-4g**. *Bottom:* Multivariable poisson regression model for whole chromosome deletion scores integrating RT, *CDKN2A* and tumor types as variables. Note that curative radiotherapy and *CDKN2A* homozygous deletion are independently associated with higher levels of whole chromosome deletions.

Radiotherapy treatment is associated with aneuploidy and larger deletions

Having established an association with radiotherapy and small deletions as well as specific mutational signatures, we reasoned that repair of RT-associated DSBs may also result in other types of genomic variants. We detected large structural variants, including large deletions, duplications, inversions and translocations, in the longitudinal GLASS cohort. We observed that a significant number of RT-treated patients demonstrated an increase in large deletions (length > 20bp to chromosome arm length) post-therapy, compared to RT-naïve patients (**Figure 5-7a**, *P* = 3.2 e-02, Fisher's exact test). Interestingly, we observed a similar statistically significant increase in inversions (**Figure 5-7a**, *P* = 2.1e-02) and no differences were observed for translocations (**Figure 5-7a**, *P* = 1) and duplications (**Figure 5-7a**, *P* = 7e-01). These associations remained significant after accounting

for potentially confounding factors such as TMZ treatment and molecular subtype (**Figure 5-8b**). While radiation-associated secondary malignancies have been reported to contain increased rates of inversions [204], a concomitant increase in large deletions in association with RT has not been previously observed.

We next evaluated whether any specific deletions were associated with radiotherapy treatment. Comparing alteration frequencies before and after RT-treatment in the GLASS cohort identified a significant link between radiotherapy and the acquisition of *CDKN2A* homozygous deletions among IDH-mutant gliomas where *CDKN2A* loss at initial diagnosis is rare (**Figure 5-7b**) [222]. *CDKN2A* homozygous deletions occurred exclusively in RT-treated recurrences (**Figure 5-7b**) and occurred significantly more frequently than in pre-treatment samples (**Figure 5-7b**, 29% vs 2%, P = 1.9e-05, Fisher's exact test), nominating acquired *CDKN2A* homozygous loss as a potential novel biomarker for RT resistance among IDH-mutant gliomas, but not IDH-wild type gliomas where *CDKN2A* homozygous deletion at diagnosis is common.

Ionizing radiation can promote mitotic chromosome segregation errors causing aneuploidy [223-225]. Specifically, ionizing radiation can induce non-disjunction events during mitosis, leading to an imbalanced chromosomal copy number between two daughter cells [226]. We investigated the association of aneuploidy with radiation therapy, separating aneuploidy events into gains or losses of entire chromosomes, likely the result of segregation errors; and partial gains or losses, requiring additional DSBs (see Methods, **Figure 5-8c**). In an analysis of the IDH-mutant GLASS cohort, we observed a significant association between RT and chromosome losses, whereas no association was observed for simple gains or complex events (**Figure 5-7c**, **Figure 5-8d**). However, after adjusting for covariates in a multivariable Poisson regression model used to model integer counts of aneuploidy events, the effect of radiotherapy on chromosome losses in the GLASS cohort was no longer statistically significant. The analysis highlighted a significant association between chromosome losses and *CDKN2A* deletions (**Figure 5-8e**), implicating that the increase in chromosome loss frequency following RT is specific to RT-associated acquired *CDKN2A* deletions. Using the HMF metastatic tumor cohort, we confirmed the association between *CDKN2A* homozygous deletions and chromosome losses (**Figure 5-7d**, **Figure 5-8f**). Using Poisson regression to model associations of chromosome loss frequency, we found that both curative RT-treatment and *CDKN2A* homozygous deletions are independently associated with increased number of chromosomal losses in the HMF datasets (**Figure 5-7d**, **Figure 5-8f**). However, testing for interactions between *CDKN2A* deletions and

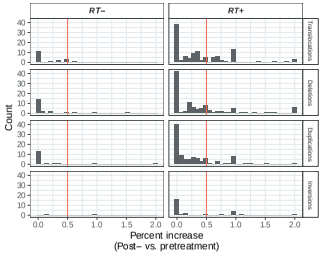
RT-treatment indicated a trend towards interaction between palliative/curative radiotherapy and *CDKN2A* deletions (**Table 5-1**, P = 9.75e-02 and P = 4.92e-02, respectively, t-test). These results raise the possibility that chromosome segregation may not directly be associated with radiotherapy but through interactions with *CDKN2A* deletions, requiring further investigation.

Table 5-1. Multivariable Poisson regression model for whole chromosome losses in metastatic cohort.

Variable	Levels	*n*	Odds Ratio (95% CI)	*P*-value
Tumor type	Bone/Soft tissue	179	(ref)	
	Breast	716	1.10 (1.01-1.20)	**2.50E-02**
	Colon/Rectum	445	1.42 (1.30-1.54)	**1.11E-15**
	Esophagus	128	1.89 (1.71-2.08)	**< 2E-16**
	Head and Neck	53	1.05 (0.90-1.22)	5.12E-01
	Lung	355	1.73 (1.59-1.88)	**< 2E-16**
	Nervous System	74	0.51 (0.43-0.60)	**2.45E-14**
	Others	810	1.16 (1.07-1.26)	**2.90E-04**
	Prostate	392	0.53 (0.48-0.59)	**< 2E-16**
	Skin	324	0.86 (0.78-0.94)	**1.50E-03**
	Urinary Tract	163	1.05 (0.94-1.17)	3.87E-01
	Uterus	54	1.03 (0.88-1.20)	7.37E-01
Radiotherapy	RT-	2265	(ref)	
	RT+ pal	689	1.04 (0.99-1.08)	1.30E-01
	RT+ cur	739	1.05 (1.01-1.10)	**2.40E-02**
CDKN2A status	WT	3065	(ref)	
	Homdel	628	1.20 (1.14-1.26)	**1.79E-12**
Interaction terms	RT+ pal : CDKN2A homdel			9.75E-02
	RT+ cur : CDKN2A homdel			6.92E-02

Including tumor type, RT, CDKN2A status and an interaction term between RT and CDKN2A as variables. Note that CDKN2A homdel and curative RT are independently associated with higher whole chromosome losses, while a trend towards interaction between CDKN2A deletion and palliative/curatuve radiotherapy treatment exists.

a



b

Deletions

Variable		n	Odds ratio	P
RT	–	27	Reference	
	+	124	5.76 (1.29, 41.40)	**0.04**
TMZ	–	28	Reference	
	+	123	3.63 (0.85, 20.05)	0.10
Surgical interval		151	0.99 (0.97, 1.01)	0.19
Subtype	IDHmut–codel	15	Reference	
	IDHmut–noncodel	45	1.44 (0.27, 11.08)	0.69
	IDHwt	91	0.33 (0.04, 3.10)	0.29

Inversions

Variable		n	Odds ratio	P
RT	–	27	Reference	
	+	124	10.29 (1.49, 219.58)	**0.047**
TMZ	–	28	Reference	
	+	123	1.58 (0.28, 11.39)	0.619
Surgical interval		151	1.01 (0.99, 1.04)	0.198
Subtype	IDHmut–codel	15	Reference	
	IDHmut–noncodel	45	0.23 (0.03, 1.36)	0.108
	IDHwt	91	0.20 (0.02, 1.78)	0.140

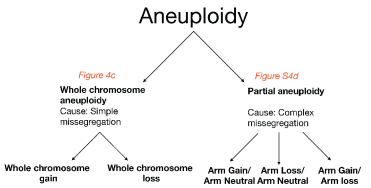
Duplications

Variable		n	Odds ratio	P
RT	–	27	Reference	
	+	124	1.56 (0.45, 6.05)	0.49
TMZ	–	28	Reference	
	+	123	1.45 (0.39, 5.95)	0.59
Surgical interval		151	1.01 (0.99, 1.03)	0.24
Subtype	IDHmut–codel	15	Reference	
	IDHmut–noncodel	45	0.48 (0.12, 2.02)	0.30
	IDHwt	91	0.36 (0.06, 2.30)	0.27

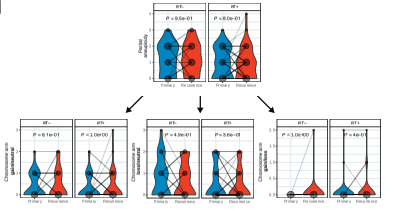
Translocations

Variable		n	Odds ratio	P
RT	–	27	Reference	
	+	124	1.16 (0.35, 4.03)	0.81
TMZ	–	28	Reference	
	+	123	0.77 (0.23, 2.60)	0.66
Surgical interval		151	1.00 (0.98, 1.02)	0.80
Subtype	IDHmut–codel	15	Reference	
	IDHmut–noncodel	45	1.41 (0.34, 7.43)	0.65
	IDHwt	91	1.45 (0.25, 10.80)	0.69

c



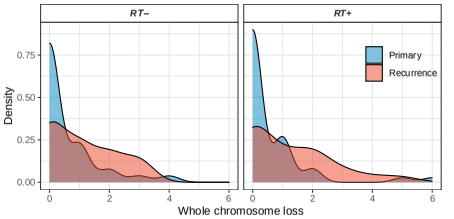
d



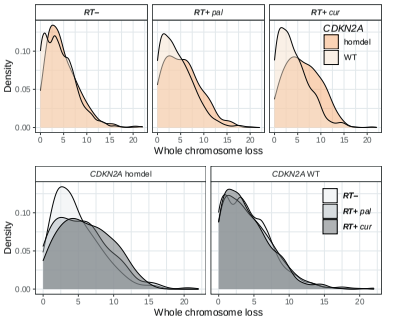
e

Variable		n	OR (95% CI)	P
Subtype	IDHmut–codel	18	Reference	
	IDHmut–noncodel	51	0.61 (0.36, 1.05)	0.068
Radiotherapy	–	27	Reference	
	+	42	0.92 (0.53, 1.58)	0.754
***CDKN2A* at recurrence**	WT	56	Reference	
	Homdel	13	2.25 (1.26, 4.00)	**0.006**
Temozolomide	–	31	Reference	
	+	38	0.80 (0.50, 1.29)	0.361
Surgical interval		69	0.99 (0.99, 1.00)	0.128

f



g



← **Figure 5-8. Aneuploidy and SV analysis.**

a. Analysis of structural variants (SVs) in glioma samples (Translocations, Duplications, Deletions, Inversions). For each patient, number of SVs were calculated pre-and post-treatment and the proportional increase after therapy for each SV-type was plotted separately for RT-naive and RT-treated samples. Based on the distribution of proportional increase from primary to recurrence, a cutoff was defined for > 50% increase that was further used for analyses in **Figure 5-7a**. **b.** To support analyses presented in **Figure 5-7a**, a multivariable logistic regression model was fitted for the >50% increase values of the structural variant types. This model includes radiation therapy, temozolomide therapy, molecular subtype and surgical interval as variables. Note that radiation therapy is independently associated with an increase in large deletions and inversions, but not duplications and translocations. **c.** Schematic overview of separation of aneuploidy events into whole chromosome aneuploidy as a result of simple segregation errors and partial aneuploidy as a result of complex segregation errors. **d.** Longitudinal analysis of partial aneuploidy in IDHmut glioma samples. Note that neither the general partial aneuploidy values, nor the detailed separation of partial aneuploidy values into gain of chromosome arms (chromosome arm gain/neutral), loss of chromosome arms (chromosome arm loss/neutral) and complex chromosome arm alterations (chromosome arm gain/loss) show significant differences for any radiation treatment group. Dots are proportional to the frequency of whole chromosome loss integer for each subgroup. Paired Wilcoxon rank-signed test was applied for statistical testing. **e.** Multivariable Poisson regression model for whole chromosome losses in IDHmut glioma including molecular subtype, RT, TMZ, surgical interval and CDKN2A status at recurrence as variables. Note that CDKN2A homdel, but not RT is independently associated with higher whole chromosome losses. **f**. Density plots over integers of whole chromosome deletion scores for comparison between primary vs recurrent glioma samples, separated by radiotherapy. **g.** Density plots over integers of whole chromosome deletion scores for comparison between RT-naïve vs RT+pal vs RT+cur and/or *CDKN2A* homdel vs. Wild-type (WT) samples from the HMF dataset. Note that *CDKN2A* homdel is associated with higher whole chromosome deletion scores, independent of RT. Within samples with *CDKN2A* homdel, samples that were RT-treated with curative intent show the highest deletion scores.

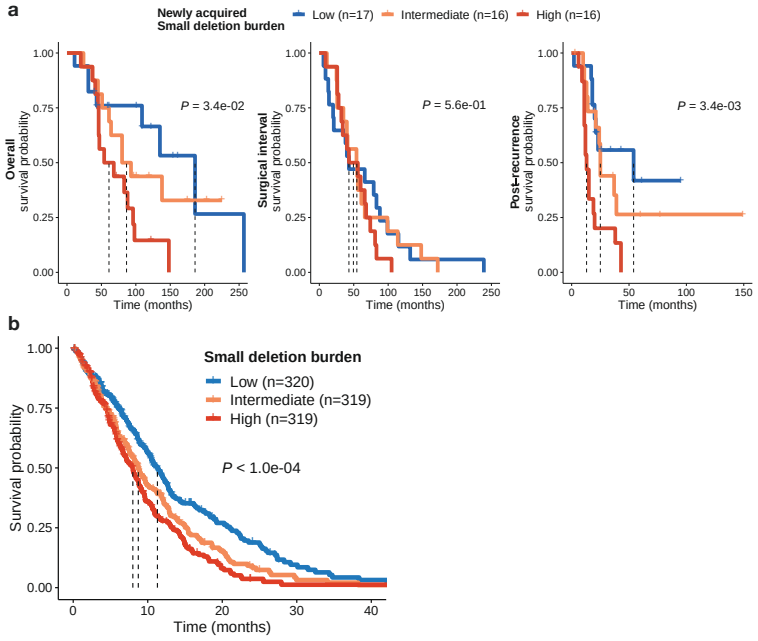
RT-associated genomic changes are linked to poor survival

Finally, we wanted to ascertain whether the genomic effects of radiotherapy were relevant to patient outcomes. A survival analysis confirmed previous observations [185] that *CDKN2A* homozygous deletion at recurrence was significantly associated with worse overall survival in IDH-mutant glioma samples (**Figure 5-10a**, *left*, $P < 1e\text{-}04$, log-rank test), independent of age, treatment or subtype (**Figure 5-10a**, *right*, $P = 1.4e\text{-}02$, Wald test). To test for a survival association of *CDKN2A* deletions amongst RT-treated patients in the HMF dataset we selected 958 samples that received RT and had sufficient survival information available from 11 tumor types (**Figure 5-2e**). Although *CDKN2A* homozygous deletions are less common outside glioma and HMF data lacks longitudinal samples to limit the analysis to cases with acquired *CDKN2A* deletion, patients whose tumors harbored a *CDKN2A* homozygous deletions showed worse outcomes compared to patients

with *CDKN2A* wild-type tumors (**Figure 5-10b**, left). Stratification of the cohort into tertials based on genome-wide aneuploidy frequency demonstrated that low aneuploidy was linked to favorable outcomes and high aneuploidy was linked to poor outcomes (**Figure 5-10b**, middle). In summary, these results nominate acquired *CDKN2A* homozygous deletion as a biomarker of RT resistance after recurrence and support the clinical reassessment of *CDKN2A* status at recurrence for optimizing treatment strategies.

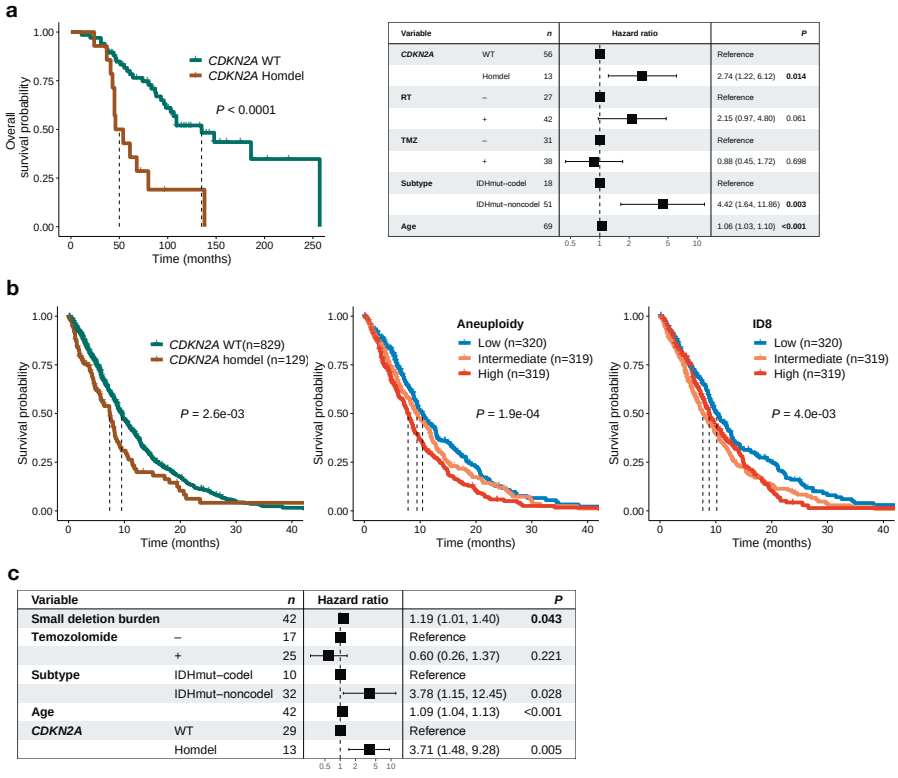
Independent of the poor prognostic implications of specific, RT-linked changes such as *CDKN2A* deletions, we found that GLASS patients with tumors carrying a high small deletion burden at recurrence (top tertial) had significantly shorter overall survival (**Figure 5-9a**, P = 3.4e-02, log rank-test). The association remained significant when accounting for the small deletion burden as a continuous variable and possible confounding variables, indicating a robust correlation (**Figure 5-10c**, HR = 1.19 [95% CI: 1.01 - 1.14]; P = 4.3e-02, Wald test). Multivariable modeling using a limited subset of patients with detailed dosage information in the GLASS cohort (n=21) further indicated that the association with the small deletion burden and survival is independent of dose, P = 2e-02). Separating the overall survival time into surgical interval and post-recurrence survival indicated that the association of high newly acquired small deletion burden with worse survival was limited to post-recurrence survival (**Figure 5-9a**, P = 3.4e-03, log-rank test). Surgical interval times did not differ significantly between the three tertials (**Figure 5-9a**, P = 5.6e-01), suggesting that glioma patients may initially benefit equally from RT, but after certain exposure to RT and acquisition of the deletion signature, patients may lose sensitivity to further radiotherapy. This pattern is reminiscent of the association between hypermutant glioma and temozolomide therapy [215].

We then determined the association between radiotherapy and patient outcome in the HMF metastatic tumor datasets using the 958 RT-treated patients described above (**Figure 5-10g**). Using this cohort, we found that patients harboring a high small deletion burden (top tertial) had significantly shorter survival than other RT-treated cases (**Figure 5-9b**, P < 4e-04, log rank-test). Similarly, we could stratify HMF patients into tertiles according to ID8 burden to show that an intermediate or high ID8 burden was associated with poor survival and a low ID8 burden was associated with improved outcomes (**Figure 5-10b**, right).



↑ **Figure 5-9. Association of RT-related deletions with survival in GLASS cohort.**

a. *Left:* Kaplan-Meier survival plots comparing overall survival dependent on deletion burden at recurrence using log-rank test in RT-treated IDH mutant glioma samples (n = 49 with available survival information). Samples were separated into 3 tertials based on deletion burden at recurrence: High (top tertial), Intermediate (middle tertial) and Low (bottom tertial). Dotted lines indicate median overall survival times. Note the stepwise association of tertials with survival. *Middle:* Kaplan-Meier survival plots comparing surgical interval/time to second surgery dependent on deletion burden at recurrence using log-rank test. *Right:* Kaplan-Meier survival plots comparing post-recurrence survival dependent on deletion burden at recurrence using log-rank test. **b**. Association of RT-related deletions with survival in metastatic cohort. Kaplan-Meier survival plots comparing survival time dependent on deletion burden at metastasis using log-rank test in RT-treated metastases (n = 958 with available survival information). Samples were separated into 3 tertials based on deletion burden: High (top tertial), Intermediate (middle tertial) and Low (bottom tertial). Dotted lines indicate median survival times. Note the stepwise association of tertials with survival.



↑ **Figure 5-10. Supplementary survival analysis of CDKN2A deletions, aneuploidy, ID8 and RT-associated deletions.**

a. *Left:* Kaplan-Meier survival plot comparing overall survival time dependent on *CDKN2A* status at recurrence using log-rank test in IDH mutant glioma samples. *Right:* Multivariable cox regression model including *CDKN2A* status at recurrence, TMZ-treatment, molecular subtype and Age as variables. **b.** *Left:* Kaplan Meier survival plot comparing survival time dependent on *CDKN2A* status at metastasis using log-rank test RT-treated metastases (n = 958 with available survival information). *Middle:* Kaplan Meier survival plot comparing survival time dependent on aneuploidy burden at metastasis using log-rank test in RT-treated metastases (n = 958 with available survival information). Samples were separated into 3 tertials based on whole chromosome loss aneuploidy scores: high (top tertial), intermediate (middle tertial) and low (bottom tertial). *Right:* Kaplan Meier survival plot comparing survival time dependent RT signature ID8 burden at metastasis using log-rank test in RT-treated metastases (n = 958 with available survival information). Samples were separated into 3 tertials based on whole chromosome loss aneuploidy scores: high (top tertial), intermediate (middle tertial) and low (bottom tertial). Note that a low ID8 burden is associated with better survival, indicating a better response to RT. **c.** Multivariable cox regression model including deletion burden at recurrence as continuous variable, *CDKN2A* homozygous deletion, Temozolomide-treatment, molecular subtype and age as variables in RT-treated samples.

The effect of RT-associated small deletion burden on the survival of the patients was further independent of mutations in DDR genes. This validation of a worse outcome association in a single-time point analysis suggests that the presence of a higher number of RT-associated small deletions implicates a tumor that has responded to therapy, but which may have lost some or all of the treatment sensitivity. Taken together, these results suggest that a higher deletion burden may reflect a scenario that is favorable to the tumor characterized by proficient DNA repair resulting in less tumor cell killing and decreased treatment efficacy.

Discussion

In this study we comprehensively evaluated the effects of ionizing radiation on the cancer genome using a cohort of pre- and post-treatment glioma samples and a cohort of metastatic tumor specimens. We identified a unique signature of RT-associated deletions that is reminiscent of double strand break repair by canonical non-homologous end joining. We found that *CDKN2A* homozygous deletions are acquired in RT-treated IDH-mutant gliomas but not in untreated recurrent IDH-mutant gliomas, raising the possibility that radiotherapy-induced DNA damage promotes the acquisition of this poor prognostic marker. Further, we found that a higher load of RT-induced deletions associated with worse patient outcomes, suggesting that the increased deletion burden reflects the repair of RT-induced DNA damage. The ability to effectively repair RT-induced damage implies that sensitivity to RT has been diminished or lost. Inhibiting these repair processes could therefore potentially sensitize tumors to the tumor-killing effect of ionizing radiation.

The genomic impact of therapeutic radiation has not been comprehensively shown, implicating a knowledge gap. Radiotherapy is used in the clinical management in over 50% of cancer patients [199, 200] and effectively the most widely used regiment for cancer treatment. Prior studies on radiation induced tumors have shown a wide range of genomic effects and have suggested the involvement of various DNA double strand break repair mechanisms [225, 227-230]. Our findings that RT is associated with significantly higher burden of small deletions harboring specific genomic signatures, large deletions, large inversion and whole chromosome loss-driven aneuploidy extends our knowledge base and provides direction for development of effective radiosensitizers.

Our work raises the possibility that these events are a consequence of RT-induced mutagenesis/repair cycles. While the nearly exclusive association between

acquired *CDKN2A* deletions and RT-treatment could indicate that these events are RT-induced, the datasets analyzed here were of insufficient coverage to determine whether the *CDKN2A* lesions resulted from selection of pre-existing lesions or were a direct consequence of radiotherapy.

The significant expansion of clones harboring RT-induced genomic events depends on clonal selection or drift [231]. Therefore, the increased small deletion burden in combination with poor outcomes may reflect the emergence of more competitive clones under RT-induced stress, innately active repair processes ensuring tumor maintenance, or a combination of these two. Thus, additional rounds of RT in patients with recurrent or metastatic tumors containing a significant increase in small deletion burden is unlikely to further extend progression-free survival. A biomarker able to readily detect increased small deletion burden may help safe cost of treatment and avoid RT-associated patient comorbidities and side-effects.

We note several limitations to our study. In contrast to the longitudinal GLASS dataset, the HMF metastatic dataset comprises samples from a single time point, complicating the attribution of alterations as post-treatment. Furthermore, the treatment annotation in HMF does not precisely describe whether only the initial tumor or additionally the metastatic site were irradiated. Considering these caveats, the effects of RT observed in this study might be more significant than what we have observed. Additionally, evolutionary pressures for local (lymph node) and distant metastases, which were analyzed homogenously in this study, might be fundamentally different [232]. We cannot exclude that RT treatment pressures have different impacts on anatomically distinct metastases, requiring further investigations in larger datasets in the future.

Numerous clinical and preclinical studies have shown efficacy in targeting DNA repair. Most famously, inhibitors directed at poly (ADP-ribose) polymerase (*PARP*) in homologous recombination-deficient tumors (synthetic lethality) were shown to be effective in the treatment of various cancer types [233-235]. Inhibitors of NHEJ are much less well studied but recent reports are suggesting efficacy in sensitizing tumors to radiotherapy. Inhibitors of ATM serine/threonine kinase (ATM), a protein kinase that sets off a cascade of steps leading to DSB repair; and DNA-dependent protein kinase catalytic subunit (DNA-PKc), a kinase that catalyzes repair at the DSB locus, were shown to be effective in pre-clinical studies and phase I trials of diffuse glioma [236-239]. The identification of enrichment of APOBEC associated mutational signatures may suggest a role for inhibitors of this class of cytosine deaminases as radiosensitizers [240]. Compounds that inhibit

DNA repair may improve the response of cancer cells to radiotherapy. The present study highlights the importance of effective DNA repair in therapy resistance.

Acknowledgements

This publication and the underlying study have been made possible partly on the basis of the data that Hartwig Medical Foundation and the Center of Personalized Cancer Treatment (CPCT) have made available to the study.

This work was supported by the following NIH grants: R01 CA190121, R01 CA237208, R21 NS114873 and Cancer Center Support Grant P30 CA034196, grants from the Musella Foundation, the B*CURED Foundation, the Brain Tumour Charity, and the Department of Defense W81XWH1910246 (R.G.W.V).

F.P.B. is supported by the JAX Scholar program and the National Cancer Institute (K99 CA226387). F.S.V. is supported by a postdoctoral fellowship from The Jane Coffin Childs Memorial Fund for Medical Research. K.C.J. is the recipient of an American Cancer Society Fellowship (130984-PF-17-141-01-DMC). E.K. is recipient of an MD-Fellowship by the Boehringer Ingelheim Fonds and is supported by the German National Academic Foundation.

Competing Interests

R.G.W.V. is a co-founder of Boundless Bio, Inc., which was not involved in the research presented here. R.G.W.V. is a member of Scientific Advisory Board of the Hartwig Medical Foundation. F.P.B. has performed consulting for Bristol Myers Squibb. R.G.W.V., E.K., K.J.A., and F.P.B. are listed as inventors on a patent application filed by The Jackson Laboratory, related to the findings described here.

Materials and Methods

Patient cohort

We curated a cohort of 190 patients with high-quality longitudinal DNA sequencing data, including treatment naïve primary and matched post-treatment first recurrence tumor samples from the GLASS dataset [185]. We classified paired samples according to the 2016 World Health Organization (WHO) classification into three subtypes: IDH mutant with 1p/19q co-deletion (IDHmut-codel), IDH mutant without 1p/19q co-deletion (IDHmut-noncodel) and IDH wild type (IDHwt)

[61]. The GLASS cohort used in this manuscript consists of *n* = 106 whole genome sequencing (WGS) samples (*n* = 53 primary samples, *n* = 53 matched first recurrence samples) and *n* = 274 whole exome sequencing (WES) samples (*n* = 106 primary samples, *n* = 106 matched first recurrence samples). Detailed information on sequence platforms, capture kits and read length information are outlined in the GLASS marker paper [185].

For validation analyses, we curated a metastatic cohort from the Hartwig Medical Foundation (HMF) comprising a total of 4549 samples [205]. The HMF cohort consists of metastatic tumor samples that were collected following local or systemic treatment as part of the CPCT-02 (NCT01855477) and DRUP (NCT02925234) clinical trials. Biopsy samples from a wide range of tumor types collected at various hospitals across the Netherlands were sequenced at the core facilities of the Hartwig Medical Foundation. Whole genome sequencing (WGS) was performed for each sample according to standardized protocols [241]. Detailed information on sequence platforms, capture kits and read length information are outlined in the HMF marker paper [205]. VCF files with mutations and associated metadata were downloaded from The Hartwig Medical Database (https://database.hartwigmedicalfoundation.nl) After application of filtering criteria (as described in detail in **Figure 5-2**e) a set of *n* = 3693 were defined and used for the majority of analyses throughout the manuscript. For survival analyses we selected curative RT-treated samples with sufficient survival information (*n* = 958). All prior radiotherapy data were extracted using clinical data as present in the CPCT-02 eCRF on December 08, 2020. These data were not cleaned and represent the data entered by the clinical sites. The prior radiotherapy was categorized as curative intent, palliative intent or other. All other instances were manually curated by the principal investigator. All adjuvant / neo-adjuvant or post-operative radiotherapy was considered curative intent radiotherapy. All local radiotherapy for pain relief or other symptom-directed goals were considered as palliative. Some items were not specified, and those events were not included in our analysis. We also excluded all radiotherapy for non-malignant disease states, specifically for gynecomastia treatment after castration. We cannot exclude over- or underrepresentation of the radiation signatures as we are unaware whether the metastases that were biopsied were not already present at the time of radiotherapy.

Variant Calling

Variant calling in the GLASS dataset was performed according to the GATK Best practices using GATK 4.1.0.0 and publicly released as part of a previous publication [185]. Briefly, GATK 4.1. was used for variant calling in tumor samples against a matched normal control. Additionally, panels of normals were constructed across multiple control samples from the same tissue source and sequencing center. Variants were broadly filtered for germline variants, cross-sample contamination, read orientation and sequence context. Variants were called across all samples for a given patient. Variants with a minimum coverage of 10 reads in both primary and recurrence and a minimum VAF of 10% for either the primary or the recurrence were included for further analysis. Variants were considered to be present if at least one mutant read was detected in a sample. Mutations directly overlapping with known repeat regions according to the repeatmasker database were removed. Specifically, we filtered out all variants in known repeat regions, including DNA satellites, microsatellites, long terminal repeats, transposable elements (LINE/SINE elements) and low complexity regions. Variant clonality was inferred for each patient individually using PyClone (v.0.13.1) and as described in more detail in the GLASS marker paper. Pipeline scripts can be found at https://github.com/fpbarthel/GLASS.

Mutation burden comparison

The mutation burden was calculated as the number of mutations per megabase (Mb) with at least 10x coverage and stratified by variant type. The overall tumor mutation burden (TMB) was calculated as the sum of the burden of small deletions, small insertions and single nucleotide variants. Recurrent tumors with greater than 10 mutations per Mb were considered hypermutated as previously described [185]. For the comparison of mutation burden between RT-treatment groups in the GLASS dataset, we calculated the burden of mutations unique to the recurrent tumors and therefore were acquired after treatment. To adjust for confounding covariables, we fitted a multivariable log-linear regression model using the *glm* function in R. In addition to RT-treatment, we included TMZ-treatment, hypermutation, surgical interval in months and molecular subtype as variables. The small deletion burden in the GLASS dataset was not confounded by batch effects. Accordingly, we included the full therapy and tumor type information for mutation burden analyses in the Hartwig metastatic cohort. To adjust for negative infinite values resulting from the log-transformation in the GLASS cohort, we added a constant value of 1 to the log function. For the metastatic

cohort, the log-transformation did not result in (negative) infinite values and therefore not necessitating the addition of a constant value.

Association of deletions with non-B DNA structures

The genomic locations of non-canonical DNA structures were derived from the Non-B DNA database [242]. We calculated for every variant position and, for comparison, for 250,000 randomly sampled positions from the reference genome, the distance to non-B features as a continuous (absolute distance to genomic feature in bp) or categorical (position in or up to 100 bp to genomic feature – yes/no) values. We used a Mann-Whitney *U* test for differences in the genomic properties of variants in radiation-induced and non-radiation-induced tumors after adjusting for random background distribution.

dNdScv

For quantification of selection processes at the level of individual genes dependent on radiation therapy, we calculated dN/dS ratios as previously described [185]. Briefly, the R package dNdScv [212] was run using the default and recommended parameters for each mutational fraction (private to primary, shared between primary and recurrence and private to recurrence). All analyses were conducted separately within radiotherapy-naïve and radiotherapy-treated groups.

Sequence microhomology

Sequence microhomology was determined by iteratively comparing the 3' end of the deleted sequence to the 5' flanking sequence. Any deletion demonstrating at least 2nt of homology was considered microhomology-mediated. The homologous sequence was characterized and further analyzed for the presence of 1nt, 2nt and 3nt repeats. The repeat unit and number of repeats were quantified.

Mutational signatures

SigProfiler was used to extract and plot mutational signatures of single base substitutions (SBS), double base substitutions (DBS) and indels (ID) [213]. Absolute and relative contributions of signatures were determined using modified functions from the MutationalPatterns R package [243]. Briefly, we fitted the mutational profile matrix generated with SigProfiler to the catalog of previously identified COSMIC mutational signatures (v3, May 2019) by solving the non-negative least squares problem. The single base substitution signatures SBS31 and SBS35 have been previously linked to platinum therapy [206, 213]. Analysis of

the HMF cohort using the signatures we extracted confirmed these previously established associations, providing further credence to the identified signatures. SigProfilerPlotting [244] was used to visualize the distribution of indel characteristics (**Figure 5-6**a-d).

Structural variants

For the GLASS dataset split reads and discordant read pairs were extracted from all tumor and normal BAM files using samtools 1.7 [169]. We used the *lumpyexpress* tool (from LUMPY 0.2.13) to call structural variants providing the data associated with the set of normal and tumor samples belonging to one patient [245]. CNV predictions inferred from read-depth using CNVnator 0.3.3 were additionally provided to garner further support for identified variants [246].The resulting call set was post-processed using SVtyper 0.6.0 to genotype structural variants for each individual sample belonging to a patient [247]. Finally, we used GATK VariantFiltration to filter all variants with less than four reads of support and those with quality scores less than ten [248]. Variants that showed any support in non-tumor samples were additionally removed. Variants were quantified per sample and further stratified according to type (translocation, duplication, deletion and inversion). We computed the change in frequencies for each patient by dividing the rate at recurrence by the rate at primary. Only variants spanning at least 20bp were considered.

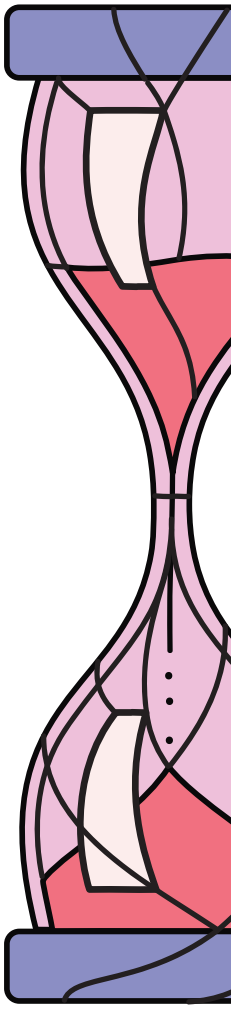
Aneuploidy calculation

Arm-level aneuploidy data from the GLASS dataset was obtained from a previous publication and copy number segmentation files from HMF were processed into arm-level copy number calls using the same methods [185]. Chromosomes demonstrating euploidy in both arms were considered euploid. Chromosomes with equidirectional aneuploidy in both arms or aneuploidy in a single arm and indeterminate ploidy in the other arm were considered “simple aneuploid”. Chromosomes with aneuploidy in one arm and incongruent ploidy in the other arm were considered “complex aneuploid”. Aneuploidy events were quantified for each tumor sample.

Statistical methods

All data analyses were conducted in R 3.6.1, Python 3.7.3 and PostgreSQL 10.5. All survival analyses including Kaplan–Meier plots and Cox proportional hazards models were conducted using the R packages *survival* and *survminer*. For un-

paired group comparisons the Mann-Whitney *U* test and Kruskal-Wallis test were used and for paired longitudinal comparisons the Wilcoxon signed-rank test was applied. Forest plots were generated using the R package *forestmodel*. Survival times for the GLASS dataset were calculated as described previously [185]. In the HMF metastatic cohort, we calculated survival starting from the date of biopsy to date of death. For patients that were alive, we used the last date of follow-up (date of treatment end) as censoring.



CHAPTER 6

Systematic analysis of telomere length and somatic alterations in 31 cancer types

Published in print

Barthel FP, Wei W, Tang M, Martinez-Ledesma E, Hu X, Amin SB, Akdemir KC, Seth S, Song X, Wang Q, Lichtenberg T, Hu J, Zhang J, Zheng S, Verhaak RG. *Systematic analysis of telomere length and somatic alterations in 31 cancer types.* Nat Genet. 2017 Mar;49(3):349-357.

Abstract

Cancer cells survive cellular crisis through telomere maintenance mechanisms. We report telomere lengths in 18,430 samples, including tumors and non-neoplastic samples, across 31 cancer types. Tumor telomeres were shorter compared to normal tissues, and longer in sarcomas and gliomas compared to other cancers. Amongst 6,835 cancers, 73% expressed telomerase reverse transcriptase *(TERT*), which was associated with *TERT* point mutations, rearrangements, DNA amplifications, and transcript fusions, and predicted telomerase activity. *TERT* promoter methylation provided an additional deregulatory *TERT* expression mechanism. Five percent of cases with undetectable *TERT* harbored *ATRX* or *DAXX* alterations, demonstrated elongated telomeres and increased telomeric repeat containing RNA (*TERRA*). The remaining 22% of tumors neither expressed *TERT*, nor harbored alterations in *ATRX/DAXX*. In this group, telomere length positively correlated with *TP53* and *RB1* mutations. Our analysis integrates telomere length, *TERT* abnormalities, telomerase activity and known ALT mechanisms with genomic determinants of telomere length in cancer.

Introduction

Telomeres make up the terminal ends of each chromosome and are composed of repetitive DNA sequence $(TTAGGG)^n$ and bound proteins [87]. These complexes function by protecting the chromosome ends from being recognized as DNA double strand breaks and preventing inadvertent activation of detrimental DNA damage response pathways [89]. Telomeres shorten with each cell division which eventually triggers cellular senescence resulting in growth arrest [88], a process that can be circumvented by the inactivation of p53 and Rb tumor suppressor proteins [249-252]. Further cell division leads to cellular crisis and ultimately cell death. Rare cells can overcome crisis through telomere maintenance.

Senescence and crisis are potent tumor suppressive mechanisms [90], and maintenance of telomere length is therefore an important step in oncogenesis. Telomere shortening can be counteracted by activating telomerase [253]. The telomerase enzymatic subunit is encoded by *TERT*, and while it is transcriptionally silent in most non-neoplastic cells, reactivation may endow a small population of cells with the ability to survive crisis, at which point they become immortalized [254]. It has been proposed that up to 90% of human cancers reactivate telomerase [255]. Several mechanisms have since been associated with *TERT* reactivation, including *TERT* promoter mutations, *TERT* promoter rearrangements, and *TERT* DNA copy number amplifications [104, 256-258]. Alternative lengthening of telomeres (ALT), a recombination-based process, is frequently observed in tumors with lack of telomerase activity and manifests with long but highly variable telomeres [259, 260]. Deactivating mutations in *ATRX* and its binding partner *DAXX* were found tightly correlated with long telomeres in pancreatic neuroendocrine tumors [261] and glioma [262]. Recent evidence suggested that loss of *ATRX* may contribute to ALT by promoting sustained sister telomere cohesion and chromatid exchange [263].

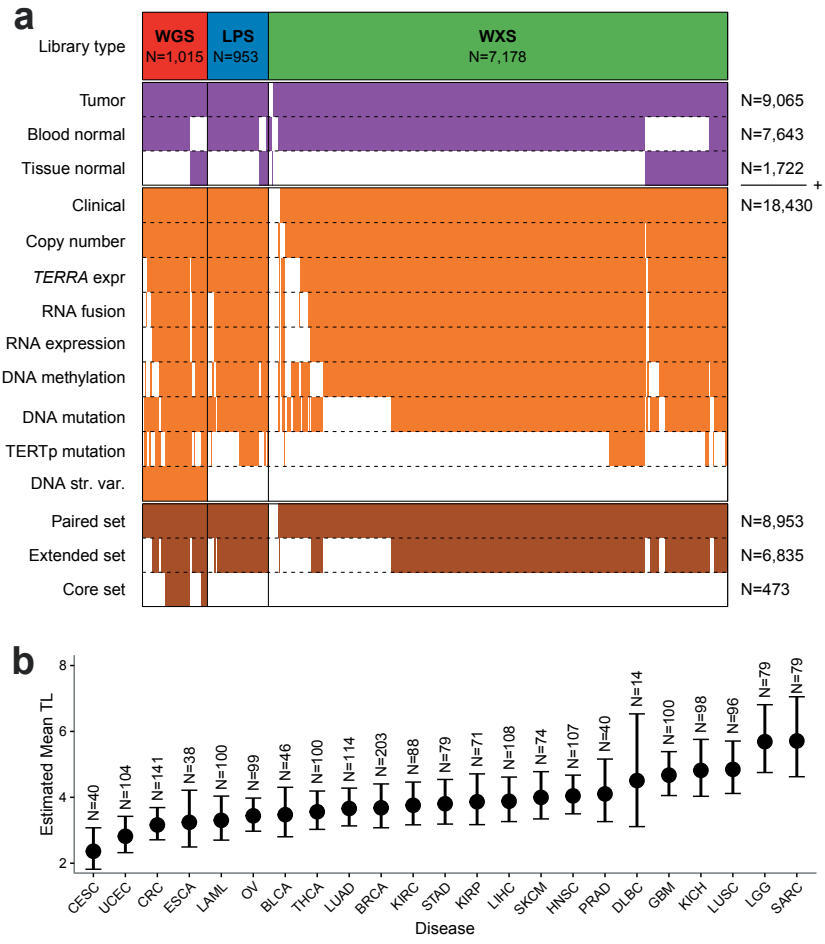
Here, we analyzed 18,430 unique samples, including tumors (n=9,065), blood controls (n=7,643) and solid tissue controls (n=1,722), from 9,127 patients across 31 cancer types, to identify genomic and transcriptomic characteristics of telomere length.

Results

Telomere length in human cancer and matching normal tissue

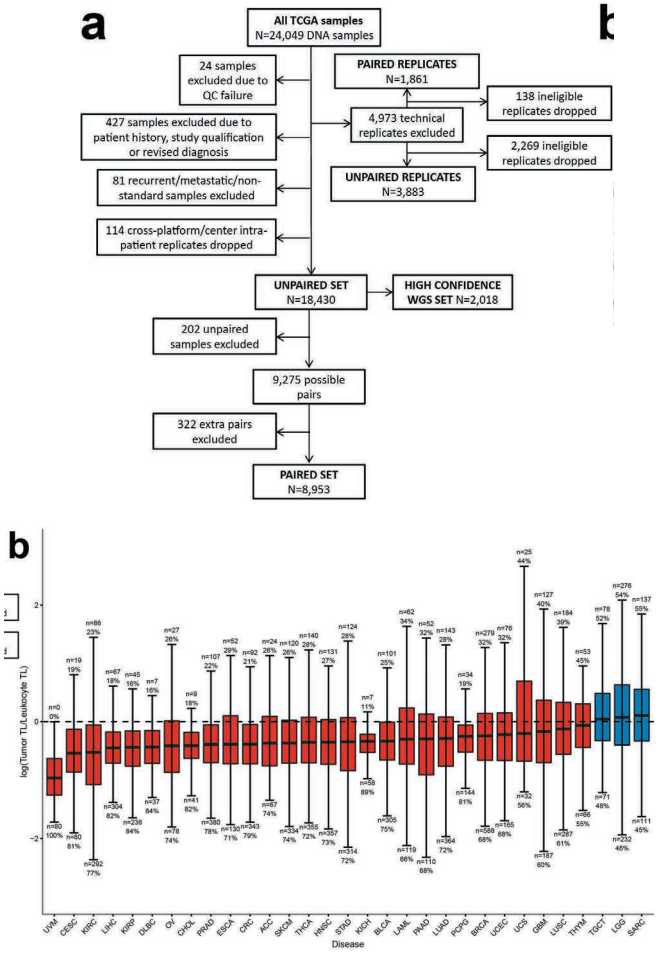
Telomere length (TL) is expected to vary between tumor types due to varying frequencies of ALT, different age distributions and variability of telomere lengths amongst cell of origin from different lineages. In order to quantify this heterogeneity, we estimated TL for 18,430 samples across 31 cancer cohorts available through The Cancer Genome Atlas (TCGA), including samples profiled using whole-genome sequencing (WGS, *n* = 2,018); low-pass whole-genome sequencing (LPS, *n* = 1,929); and whole-exome sequencing (WXS, *n* = 14,483) (**Figure 6-1a**, ***Supplementary Table 1****, available online with the full-text of this article at www.nature.com/ng*) [264]. The full dataset consisted of tumor samples, blood normal samples, and solid tissue controls (**Figure 6-2a**). Matching tumor and normal (T/N) samples were available from 8,953 unique patients (**Figure 6-2b**) and using tumor/normal TL ratios alleviated technical effects from differences in sequencing center and method (**Figure 6-3a-d**).

In order to compare TL across tumors and normal tissues, we used linear mixed modeling to adjust high-confidence WGS-based TL (*n* = 2,018) for confounding effects (***Supplementary Table 2****, available online with the full-text of this article at www.nature.com/ng*). In addition to 734 blood normal samples, this analysis included 213 normal tissue samples of five different tissue types, including liver (*n* = 21), lung (*n* = 46) and kidney (*n* = 81). We did not detect statistically significant differences between tissue types and confirmed considerable variability between samples from the same tissue type (**Figure 6-4a-b**) and negative correlation between TL and age (**Figure 6-4c**) [265].



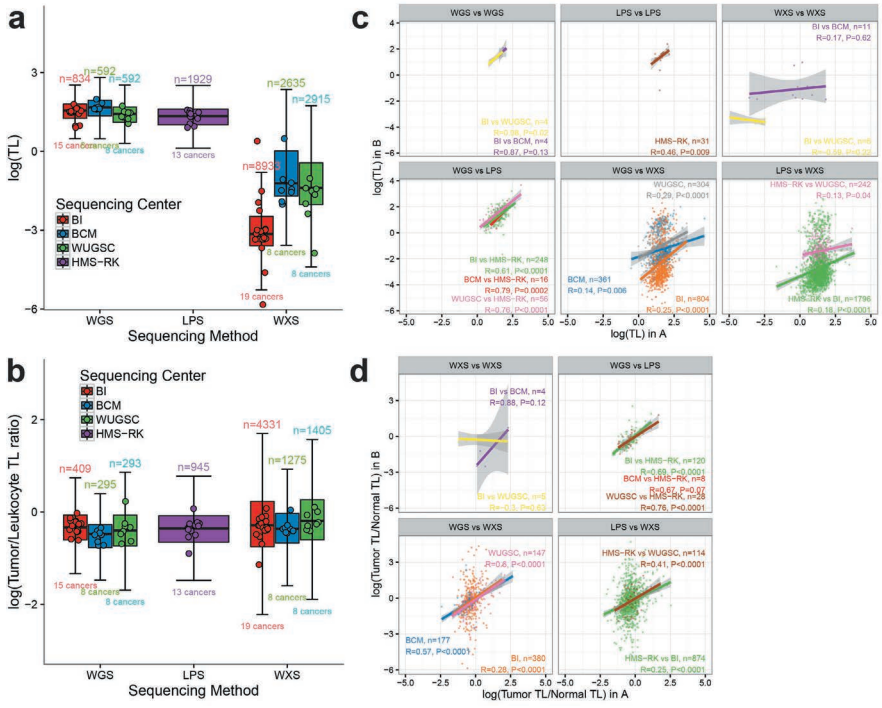
↑ **Figure 6-1. Telomere length in human cancer.**

a. Heatmap of patients in the unpaired set (in purple, *n* = 18,430) and fully paired set (in brown, *n* = 8,953). Each column represents a single patient. Rows in orange represent available data depending on platform. The extended set (*n* = 6,835) and core set (*n* = 473) are given in brown. **b.** Linear mixed model mean TL estimates using the high-confidence WGS set (*n* = 2,018) by sample type and for each tumor type. Error bars indicate 95% confidence interval. Estimates were adjusted for age, gender and sequencing center. Definitions for tumor type acronyms can be found in the Online Methods.



↑ **Figure 6-2. Sample selection procedure and TL ratio across cancer.**

a. Flowchart of sample selection. After exclusion of unsuitable samples 18,430 samples remained. Tumor and normal samples were subsequently paired and extra pairs per patient were dropped for a paired set consisting of *n* = 8,953 pairs. Further sample selection based on available data is presented in **Figure 6-1a**. The sample selection, tumor/normal pairing procedure and replicate analysis are explained in-depth in the supplementary methods. **b.** Boxplot of log T/N TL ratio 'log(Tumor TL/Control TL)' across cancer types for all *n* = 8,953 samples in the paired set. Boxes colored in blue indicate that the median TL ratio for that cancer type is greater than 1 (log ratio greater than 0), and thus more than 50% of samples show TL elongation. Numbers and percentages at the top and bottom whiskers represent cancer cases with TL longer and shorter than matched normal, respectively.

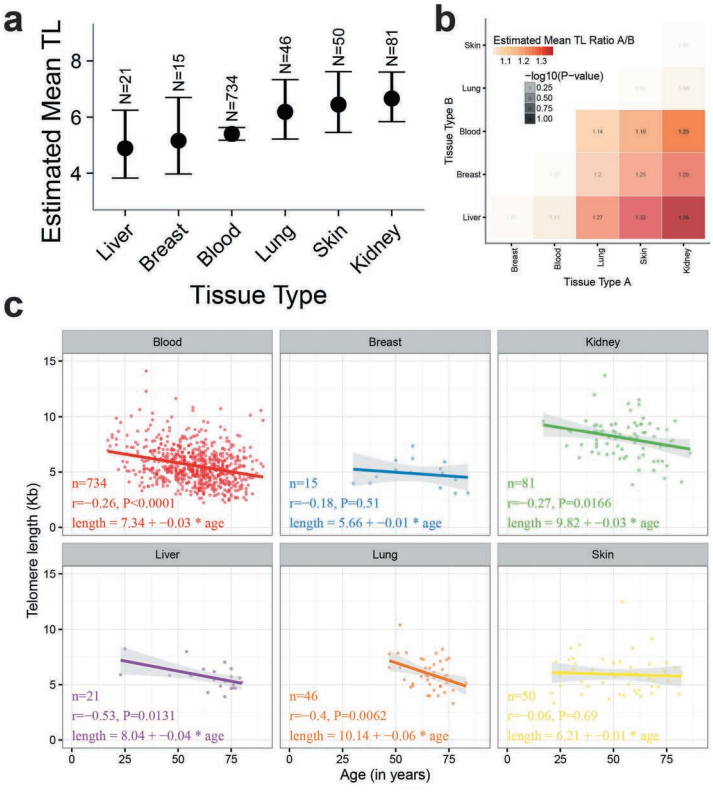


↑ **Figure 6-3. Benchmark of TL and TL ratio across centers and sequencing methods.**

a. Boxplots showing log(TL) across centers and sequencing methods. Each point indicates the median for a single cancer type. BI, Broad Institute sequencing center; BCM, Baylor College of Medicine sequencing center; WUGSC, Washington University sequencing center; HMS-RK, low pass sequencing. **b.** Boxplots showing the log(Tumor TL/Normal TL) ratio across centers and sequencing methods. Each point indicates the median for a single cancer type. **c.** Comparison of *n* = 3,883 log(TL) replicates. Sequencing center was kept constant for comparison across sequencing method and was variable for comparison within sequencing method. Replicate pairs that did not follow these criteria were dropped. **d.** Comparison of *n* = 1,861 T/N TL ratio replicates. Sequencing center was kept constant for comparison across sequencing method and was variable for comparison within sequencing method. Replicate pairs that did not follow these criteria were dropped.

Across neoplastic samples cervical (2.36 Kb, 95%CI 1.82-3.07 Kb) and endometrial cancer (2.82 Kb, 95%CI 2.32-3.42 Kb) showed the shortest average TL whereas glioma (5.69 Kb, 95%CI 4.75-6.81 Kb) and sarcoma (5.71 Kb, 95%CI 4.63-7.05 Kb) showed the longest (**Figure 6-1b**). Tumors showed relative TL shortening (tumor TL < normal TL) compared to matched normals in 70% and relative TL elonga-

tion (tumor TL > normal TL) in 30% of our cohort (**Figure 6-2b**). Tumor types that showed the highest rates of relative longer TL included testicular germ cell tumors (52%), lower grade glioma (54%) and sarcoma (55%), possibly reflecting the high telomerase activity seen in malignant testicular germ cell tumors [266] and the high frequency of ALT in lower-grade glioma and sarcoma. Conversely, uveal melanoma (100%), kidney chromophobe (89%), kidney papillary (84%) and lymphoma (84%) demonstrated the highest rate of relatively shorter TL.

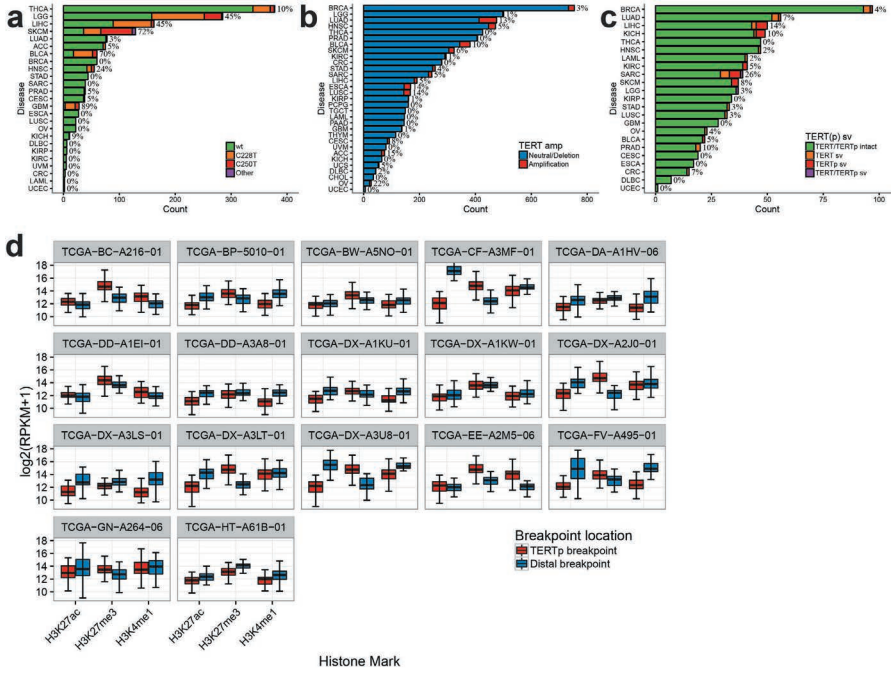


↑ **Figure 6-4. Telomere length in matching normal tissue.**

a. Linear mixed model mean TL estimates by for each tissue type. Error bars indicate 95% confidence interval. Estimates were adjusted for age, gender and sequencing center. **b.** Pairwise comparison of normal TL between tissue types. Tukey-Kramer adjustment was used for all pairwise comparisons. **c.** Scatter plots for age and normal TL across each normal tissue type (n > 10).

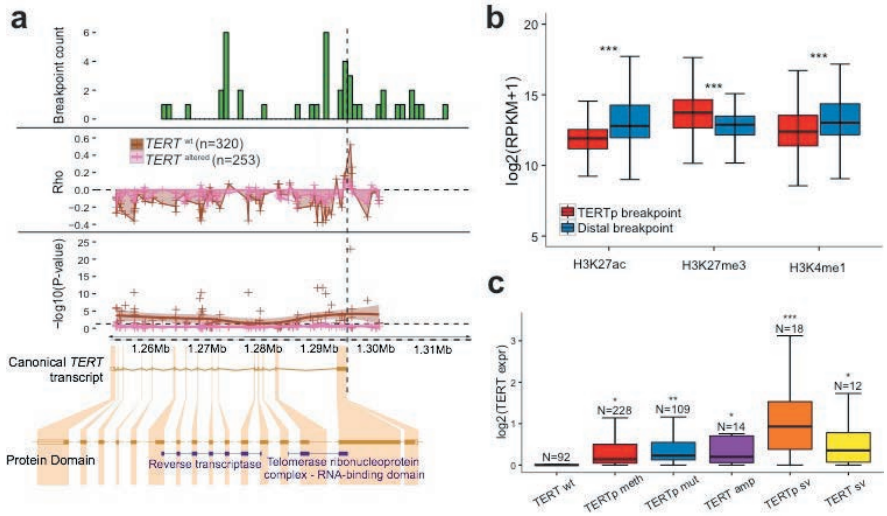
Multiple modalities associated with *TERT* overexpression

To catalogue somatic alterations that may lead to overexpression of *TERT* in cancer, we performed a genomic and epigenetic *TERT* survey. We curated a core sample set that consisted of 473 T/N pairs with the most comprehensive molecular profiling and an extended set that consisted of 6,835 T/N pairs with varying numbers of cases profiled by each individual platform (**Figure 6-1a**, Methods). *TERT* promoter (TERTp) mutations, predominantly C250T and C228T, were detected in 27% of the extended set for the cases where TERTp status could be determined (*n* = 1581). In agreement with previous reports [103, 267]; high incidence of TERTp mutations was found in bladder cancer (42/60, 70%), liver cancer (73/162, 45%), melanoma (93/129, 72%), lower grade glioma (127/285, 45%) and glioblastoma (25/28, 89%, **Figure 6-5a**). We found *TERT* focal amplifications in four percent of all samples and these events were most frequently observed in ovarian cancer (6/27, 22%), lung adenocarcinoma (63/476, 13%) and lung squamous cell carcinoma (23/167, 14%), esophageal carcinoma (23/168, 14%) and adrenocortical carcinoma (11/75, 15%, **Figure 6-5b**) [257]. Structural variants (SV) involving *TERT* or the *TERT* upstream regulatory region (TERTp) were detected in 15 samples (3%) respectively 17 samples (4%) of the core set (**Figure 6-6a** and ***Supplementary Table 3***, *available online with the full-text of this article at www.nature.com/ng*). *TERT*/TERTp structural variants were most frequent in sarcoma (10/39, 26%), hepatocellular carcinoma (7/50, 14%), kidney chromophobe (5/49, 10%) and prostate cancer (2/20, 10%)(**Figure 6-5c**).



↑ **Figure 6-5. TERT alterations in cancer (1).**

a. Circos plot of *TERT* fusion partners. Only segments of each chromosome are shown. Segment coordinates in Kbp are indicated in purple. The 5' fusion partner is shown in blue and the 3' fusion partner is shown in orange. **b.** Pairwise comparison of *TERT* promoter beta value (cg11625005) in tumor and normal. P-values were calculated using a two sided Mann–Whitney U test. **c.** Boxplot of TL length ratio in groups of *TERT* alterations. TL length ratio in *TERT* altered groups was compared to the *TERT* wt group using a two-sided t-test. TERTp meth, promoter methylation; TERTp mut, promoter hotspot mutation; TERT amp, copy number amplification; TERTp sv, promoter structural variation; TERT sv, gene body structural variation; TERT wt, cases without detectable evidences in all aforementioned groups. **d.** Boxplot of *TERC* expression in groups of *TERC* alterations. *TERC* expression in *TERC* altered groups was compared to the *TERC* wt group using a two-sided t-test. *** P<0.0001; ** P<0.001; * P<0.05; N.S. not significant.

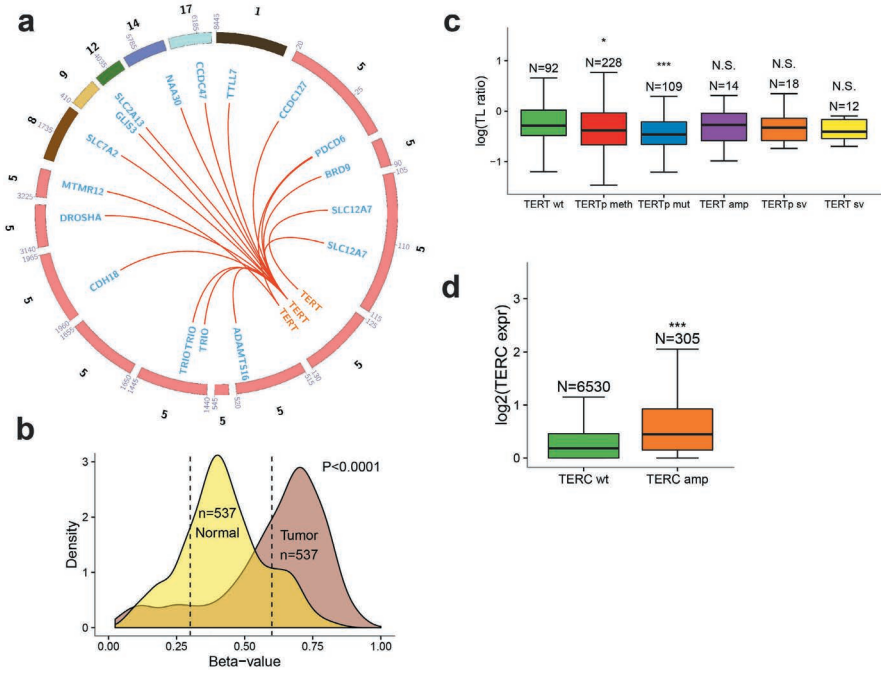


↑ **Figure 6-6. Multiple modalities associated with TERT overexpression.**

a. *Top:* Histogram of DNA breakpoints in *TERT* in the core set (n=473). SpeedSeq detected 44 breakpoints aligning to *TERT* or the *TERT* promoter in 30 samples. *Middle:* Smoothed scatter plot of the correlation coefficient and significance of probe-expression correlations for *TERT*. Each point represents an Illumina 450k probe. The vertical dashed line represents the transcription start site. The analysis was performed separately for *TERT* wt and *TERT* altered samples. *Bottom:* The x-axis is represented by the canonical *TERT* transcript visualized using the Ensembl browser. Protein domains shown according to Pfam. **b.** Boxplot of H3K27ac, H3K27me3 and H3K27me1 levels from the NIH Roadmap Epigenomics dataset at the locations of TERTp structural variant proximal and distal breakpoints (N=17). This dataset consists of 183 biological samples consolidated into 111 epigenomes (Suppl. Methods). P-values were calculated using two-sided t-tests. **c.** Boxplot of *TERT* expression in groups of various *TERT* alterations. *TERT* expression in each *TERT* altered group was compared to the *TERT* wt group using a two-sided t-test. *** P<0.0001; ** P<0.001; * P<0.05; N.S. not significant.

TERT promoter rearrangements have been proposed to lead to juxtaposition of enhancer elements in neuroblastoma [258]. To investigate this hypothesis using our dataset, we overlaid the genomic positions with a database of 65,950 super enhancers across 107 tumor and normal cell types [268]. In the majority of TERTp structural variants (65%), at least one predicted super enhancer was found to directly overlap with the juxtaposed position. We compared the enhancer marks histone H3K27 acetylation (H3K27ac) and H3K4 mono-methylation (H3K4me1) signal for each of the breakpoints in the *TERT* promoter and the juxtaposed, distal, breakpoints across 111 epigenomes available from Roadmap Epigenomics Consortium and found significant enrichment for enhancer marks in the juxtaposed

position compared to the *TERT* promoter (**Figure 6-6b**, **Figure 6-5d**) [269, 270]. Our data suggest that *TERT* promoter rearrangements may result in repositioning of enhancer elements that activate *TERT* transcription.



↑ **Figure 6-7. TERT alterations in cancer (2).**

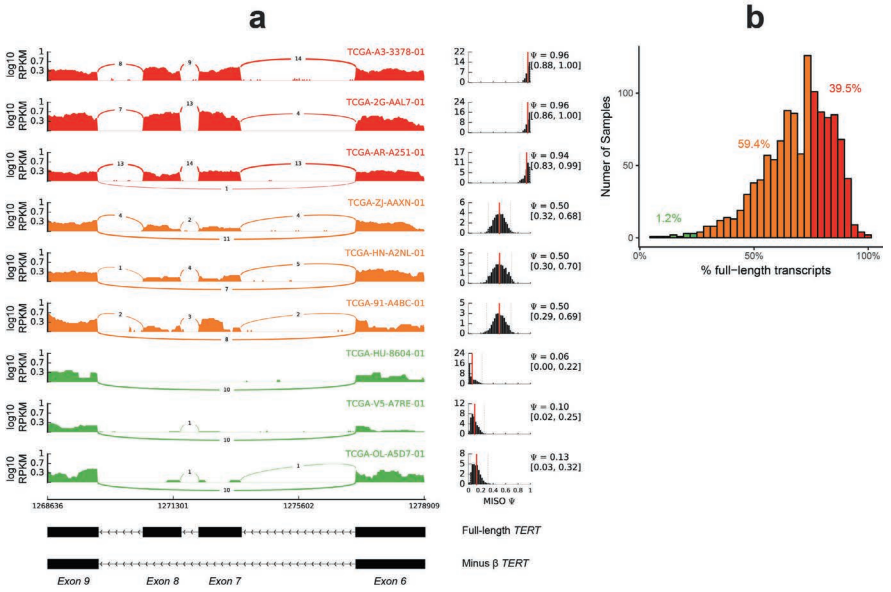
a. Circos plot of *TERT* fusion partners. Only segments of each chromosome are shown. Segment coordinates in Kbp are indicated in purple. The 5' fusion partner is shown in blue and the 3' fusion partner is shown in orange. **b.** Pairwise comparison of *TERT* promoter beta value (cg11625005) in tumor and normal. P-values were calculated using a two sided Mann–Whitney U test. **c.** Boxplot of TL length ratio in groups of *TERT* alterations. TL length ratio in *TERT* altered groups was compared to the *TERT* wt group using a two-sided t-test. TERTp meth, promoter methylation; TERTp mut, promoter hotspot mutation; TERT amp, copy number amplification; TERTp sv, promoter structural variation; TERT sv, gene body structural variation; TERT wt, cases without detectable evidences in all aforementioned groups. **d.** Boxplot of *TERC* expression in groups of *TERC* alterations. *TERC* expression in *TERC* altered groups was compared to the *TERC* wt group using a two-sided t-test. *** P<0.0001; ** P<0.001; * P<0.05; N.S. not significant.

We next performed a supervised search for *TERT* gene fusions [271, 272]. *TERT* fusion transcripts were detected in 3% of the extended set (**Figure 6-7a** and ***Supplementary Table 3***, *available online with the full-text of this article at www.*

nature.com/ng). *TERT* was always the 3' partner gene, and in 13 of 19 fusions the 5' partner gene resided on chromosome 5. All fusions demonstrated altered exon expression flanking the fusion point, and 16/19 TERT fusion breakpoints fell in the second intron. Both exons two and three map to the telomerase RNA binding domain and it is thus unlikely that the fusion product retains canonical *TERT* functionality [273, 274]. We observed read coverage on exon 2 in 15 of 16 samples, suggesting that additional *TERT* transcripts were expressed.

Taken together, we found somatic *TERT* alterations including *TERTp* mutations, *TERT* amplifications and *TERT* structural variants involving gene promoter or gene body in 32% of core set samples. Somatic *TERT* alterations associated with detectable *TERT* transcripts in 93%. Next, we evaluated whether epigenetic mechanisms could also be related to *TERT* transcriptional activation. We correlated *TERT* expression to DNA methylation probes mapping to the *TERT* gene body (n = 72) and promoter (n = 3). We observed moderate correlations between TERTp DNA methylation and *TERT* expression in samples carrying a somatic *TERT* alteration (FDR>0.05, abs(Rho)<0.3; **Figure 6-6a**). In contrast, samples lacking somatic *TERT* alterations showed significant negative correlation between gene body methylation and expression, and positive correlation between promoter methylation and expression (**Figure 6-6a**). As previously described in pediatric brain tumors, *TERT* promoter probe cg11625005 demonstrated a strong correlation with *TERT* expression (Rho=0.52, FDR<0.0001) [275]. Further comparison of this probe using 537 paired tumor and adjacent tissue normal samples showed a general absence of methylation of this probe in normal samples (**Figure 6-7b**).

We found that 63% of *TERT* wild type tumors in the core set expressed *TERT,* of which 91% showed promoter DNA methylation compared to 40% of the *TERT* non-expressing samples. Altogether, 95% of *TERT* expressing samples showed TERTp mutations (31%), *TERT* amplification (3%), *TERT* structural variants (3%), TERTp structural variants (5%) or TERTp methylation (53%). Amongst different types of *TERT* aberrations, the TERTp sv group showed the highest transcription levels (two-sided *t*-test $P < 0.05$, **Figure 6-6c**). TERTp methylation (two-sided *t*-test $P < 0.05$) and TERTp mutations (two-sided t-test P<0.0001) were associated with relative TL shortening compared to other types of *TERT* alterations (**Figure 6-7c**).



↑ **Figure 6-8. Abundance of the minus beta splice variant.**

a. Example of *TERT* isoform abundance. The figures shows three samples with abundant full-length transcripts (in red), three samples with mixed full length and minus-beta transcripts (orange) and three samples with predominantly minus beta transcripts (green). Full-length and minus beta exon models are shown for reference. The MISO ψ (psi, percent spliced in) is shown for each sample and indicates the percentage of full-length transcripts relative to minus beta transcripts. **b.** Histogram of the percentage of full-length transcripts relative to minus beta transcripts in N=1,201 samples. Samples with less than 25% full length transcripts (more than 75% minus beta transcripts) are shown in green, samples with between 25% and 75% full-length transcripts in orange and samples with more than 75% full-length transcripts in red. There is a significant enrichment of full-length transcripts relative to minus beta transcripts (one-sample t-test P < 0.0001; Mu = 50%).

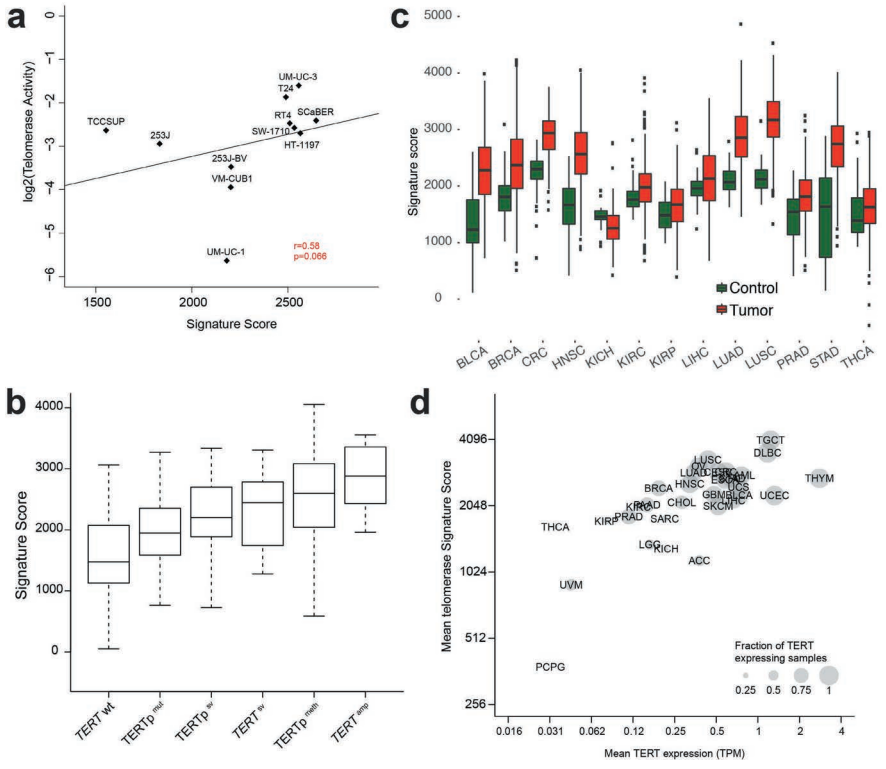
We next performed a similar analysis of *TERC*, the RNA subunit of the telomerase complex [276]. *TERC* was amplified in four percent of samples in the extended set, and most frequently in lung squamous cell carcinoma (*n* = 68/167, 41%), esophageal cancer (*n* = 36/168, 21%) and ovarian cancer (*n* = 6/27, 22%). We did not identify mutations or structural variants targeting *TERC*. Focal *TERC* amplifications were associated with increased *TERC* expression (two-sided *t*-test *P* < 0.0001) (**Figure 6-7d**), and were enriched in *TERT* expressing samples (Odds Ratio (OR) 2.59, Fisher's Exact *P* < 0.0001).

Inferring telomerase activity using a gene expression signature

Previous studies have shown a complex role for transcription of the *TERT* gene and its various isoforms in determining telomerase activity [126, 277, 278], in part because only the full length *TERT* transcript can encode enzymatically active protein [279]. The minus beta isoform skips exons 7 and 8 and prematurely stops at exon 10 and is thought to be the most abundant *TERT* splice variant in cancer (***Supplementary Table 4**, available online with the full-text of this article at www.nature.com/ng*)[280, 281]. We evaluated the presence of the minus beta transcripts in 1,201 samples with sufficient exon 6 to 9 coverage using MISO (**Figure 6-8a**) [282]. Full-length transcripts were detected in all samples and were significantly more abundant than minus beta transcripts (one-sample *t*-test $P < 0.0001$, Mu=50%; **Figure 6-8b**).

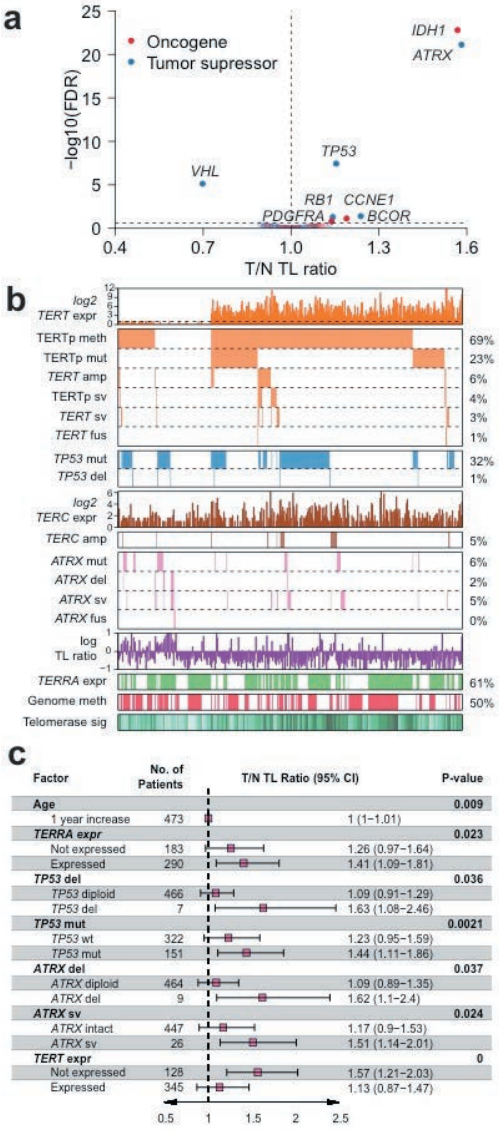
Given the limitations of *TERT* expression as measured by RNAseq in predicting telomerase activity, and lack of telomerase measurements in the TCGA proteomics datasets (data not shown), we inferred telomerase activity through a gene signature (***Supplementary Table 4**, available online with the full-text of this article at www.nature.com/ng*). The signature showed a positive correlation with telomerase enzymatic activities determined in urothelial cancer cell lines ($n = 11$, **Figure 6-9a**), without reaching statistical significance ($P = 0.07$) [126]. Telomerase activity scores were significantly higher in tumor samples compared to various solid tissues (**Figure 6-9b**), with kidney chromophobe carcinoma representing an exception. Both results suggested that the gene signature provided a general estimate of telomerase activity.

We found telomerase signature scores were significantly higher in *TERT* expressing cancers than non-expressing cancers in both core and extended set (both $P < 0.0001$, Wilcoxon rank sum test). Samples with *TERT* amplification scored highest, followed by promoter methylation, gene body and promoter structural variation, and promoter mutation (**Figure 6-9c**). *TERC* amplification was additionally associated with higher telomerase signature scores compared to non-amplified samples (two-sided *t*-test, $P < 0.0001$), which may in part be explained by the co-expression patterns of *TERT* and *TERC*. Pan-cancer analysis using all 31 cancer types revealed a positive correlation between *TERT* expression and telomerase signature score (Rho=0.69, $P < 0.0001$), with testicular germ cell tumors showing the highest average scores and pheochromocytoma and paraganglioma the lowest (**Figure 6-9d**).



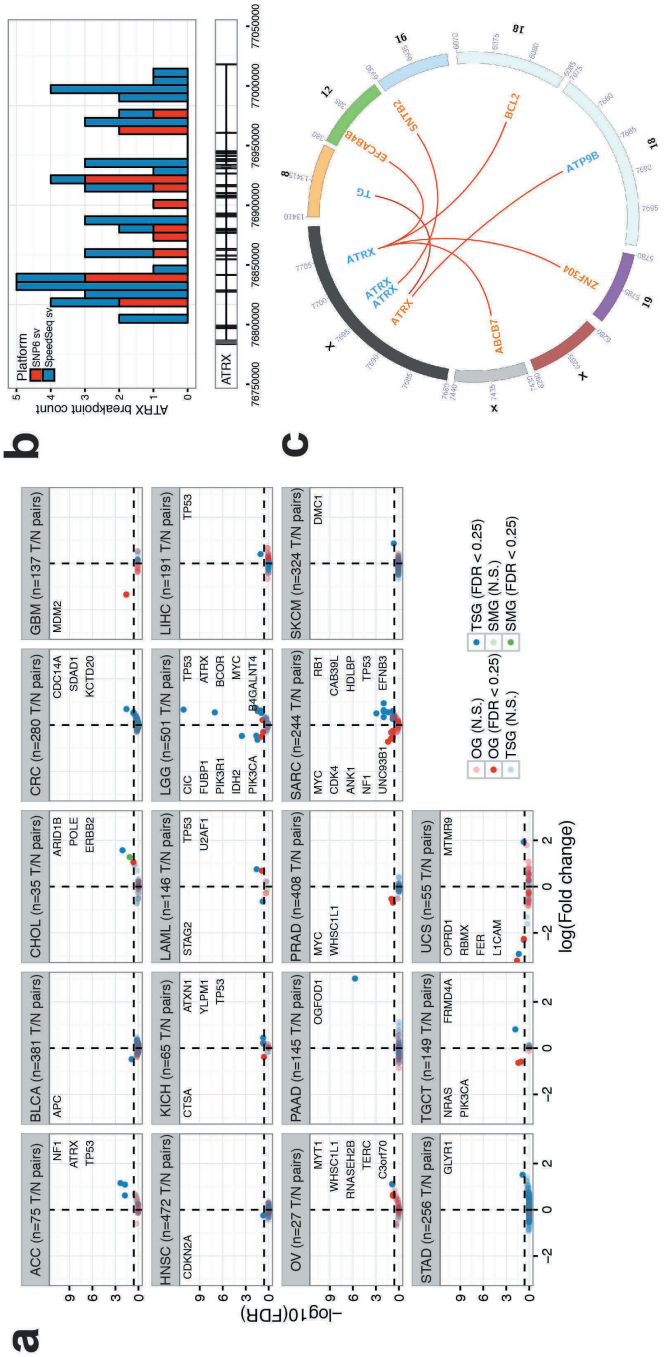
↑ **Figure 6-9. Inferring telomerase activity using a gene expression signature.**

a. The predicted telomerase activity is correlated with experimentally determined telomerase enzymatic activity in 11 urothelial cancer cell lines with a borderline significance (p=0.07, Spearman correlation). **b.** Telomerase signature score in tumor and normal samples in TCGA cohorts. Across all cancer types except KICH and THCA tumor scores are significantly higher than that of normal samples (P < 0.001, t test). **c.** Distribution of telomerase activity score across *TERT* alteration categories. All *TERT* aberrant groups are significantly higher than the TERT wt group (P<0.01). **d.** Telomerase activity score in 31 cancer types. X-axis represents mean *TERT* expression measured by TPM. Y-axis represents median telomerase score. The size of each dot is proportional to the percentage of *TERT* expressing samples in the corresponding cancer type. X and Y-axis are in log2 scale for better visualization.



↑ **Figure 6-10. Multivariable Genomic Determinants of Telomere Length.**

a. Scatterplot showing gene to TL ratio associations using the extended set (*n* = 6,835). P-values were calculated using a two-sided t-test and adjusted for multiple testing using FDR. **b.** Heatmap of *TERT*, *TERC*, *ATRX* and *TP53* expression and somatic alterations in the core set (*n* = 473). TL ratio, *TERRA* expression and telomerase signature score are also shown. Each column represents a sample. **c.** Linear regression analysis of TL ratio. Variables shown are independent predictors of TL in the core set (*n* = 473). Variables from panel b were selected using backwards elimination to derive the final model. R^2 = 0.16.


a
ACC (n=75 T/N pairs)
NF1
ATRX
TP53
BLCA (n=381 T/N pairs)
APC
CHOL (n=35 T/N pairs)
ARID1B
POLE
ERBB2
CRC (n=280 T/N pairs)
CDC14A
SDAD1
KCTD20
GBM (n=137 T/N pairs)
MDM2
HNSC (n=472 T/N pairs)
CDKN2A
KICH (n=65 T/N pairs)
CTSA
ATXN1
YLPM1
TP53
LAML (n=146 T/N pairs)
STAG2
TP53
U2AF1
LGG (n=501 T/N pairs)
CIC
FUBP1
PIK3R1
IDH2
PIK3CA
TP53
ATRX
BCOR
MYC
B4GALNT4
LIHC (n=191 T/N pairs)
TP53
OV (n=27 T/N pairs)
MYT1
WHSC1L1
RNASEH2B
TERC
C3orf70
PAAD (n=145 T/N pairs)
OGFOD1
PRAD (n=408 T/N pairs)
MYC
WHSC1L1
SARC (n=244 T/N pairs)
MYC
CDK4
ANK1
NF1
UNC93B1
RB1
CAB39L
HDLBP
TP53
EFNB3
SKCM (n=324 T/N pairs)
DMC1
STAD (n=256 T/N pairs)
GLYR1
TGCT (n=149 T/N pairs)
NRAS
PIK3CA
FRMD4A
UCS (n=55 T/N pairs)
OPRD1
RBMX
FER
L1CAM
MTMR9
−log10(FDR)
log(Fold change)
OG (N.S.)
OG (FDR < 0.25)
TSG (N.S.)
TSG (FDR < 0.25)
SMG (N.S.)
SMG (FDR < 0.25)
b
ATRX breakpoint count
Platform
SNP6 sv
SpeedSeq sv
ATRX
c
ATRX
ATRX
ATRX
TG
EFCAB4B
SNTB2
BCL2
ATP8B
ZNF304
ABCB7

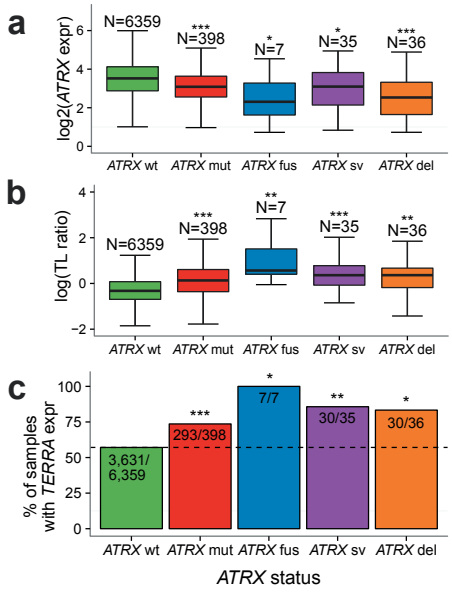
← **Figure 6-11. TL genomic associations, ATRX and TERRA.**

a. Scatterplot showing gene to TL ratio associations using the extended set (*n* = 6,835). Results are grouped by disease. P-values were calculated using a two-sided t-test and adjusted for multiple testing using FDR. Up to five negatively (left) or positively (right) associated genes with an FDR < 0.25 are listed in each plot. **b.** Histogram of DNA breakpoints in *ATRX* in the core set (N=473 samples). Bars are colored according to breakpoint detection method. **c.** Circos plot of *ATRX* fusion partners. Only segments of each chromosome are shown. Segment coordinates in Kbp are indicated. The 5' fusion partner is shown in blue and the 3' fusion partner is shown in orange.

ATRX altered tumors show profound telomere elongation

To identify *TERT*-independent mechanisms involved in TL regulation, we associated somatic alterations to TL ratio in the extended set. We reduced the search space by selecting genes significantly mutated, genes focally deleted or gained, and genes from a manually compiled list of telomere associated genes (*n* = 196; ***Supplementary Table 5**, available online with the full-text of this article at www.nature.com/ng*). We found alterations of *ATRX* and *IDH1* as the most significantly associated with relative TL elongation (both FDR<0.0001; **Figure 6-10a**). Since *IDH1* mutations frequently co-occur with *ATRX* in glioma, we tested a model with both tumor type and *IDH1* as covariates, and found *IDH1* no longer associated with TL ratio (two-sided t-test *P* = 0.15). Other hits (FDR<0.25) associated with relative TL elongation included *TP53* (TL ratio 1.15, 95%CI 1.1-1.21, FDR<0.0001), *BCOR* (TL ratio 1.24, 95%CI 1.08-1.42, FDR=0.04), *RB1* (TL ratio 1.14, 95%CI 1.04-1.25, FDR=0.05), *CCNE1* (TL ratio 1.19, 95%CI 1.05-1.35, FDR=0.07) and *TERC* (TL ratio 1.14, 95%CI 1.02-1.27, FDR=0.16). Alterations of the *VHL* were found to be associated with relative TL shortening (TL ratio 0.7, 95%CI 0.61-0.8, FDR<0.0001). We repeated the analysis in each cancer type and found *TP53* associated with relative TL elongation in six tumor types (**Figure 6-11a**). Given the role of p53 in apoptosis regulation and senescence bypass, direct involvement of *TP53* in telomere maintenance should be carefully tested in well-controlled conditions. A linear regression model showed that in addition to older age, positive *TERRA* expression, *TP53* deletion, *TP53* mutations, *ATRX* deletion, *ATRX* structural variants and absent/undetectable *TERT* expression were all independently associated with relative TL elongation (**Figure 6-10b-c**). Although *DAXX* has been linked to telomere length and ALT, *DAXX* mutations (*n* = 51/6,835) and deletions (*n* = 5/6,835) did not associate with TL [262, 283].

Because of the strong correlation between *ATRX* and TL, we performed a thorough investigation of alterations affecting this gene. In addition to non-synonymous mutations and deletions, we detected *ATRX* structural variants from WGS data in 5% of the core set samples (*n* = 26/473). In order to detect *ATRX* structural variants across the extended set, we used DNA copy number calls to detect breakpoints in *ATRX* (**Figure 6-11b**). Despite the high detection threshold of this method (34% sensitivity, 100% specificity), *ATRX* structural variants were predicted in 70 of 6,835 samples (1%). *ATRX* fusion transcripts were found in seven extended set samples (**Figure 6-11c**). *ATRX* was the 5' gene partner in five fusion genes and the fusions were predicted to result in retaining less than 10% of the *ATRX* protein sequence. *ATRX* was the 3' partner in two fusion transcripts and functional protein domains were retained in both.



↑ **Figure 6-12. ATRX altered tumors show profound telomere elongation.**

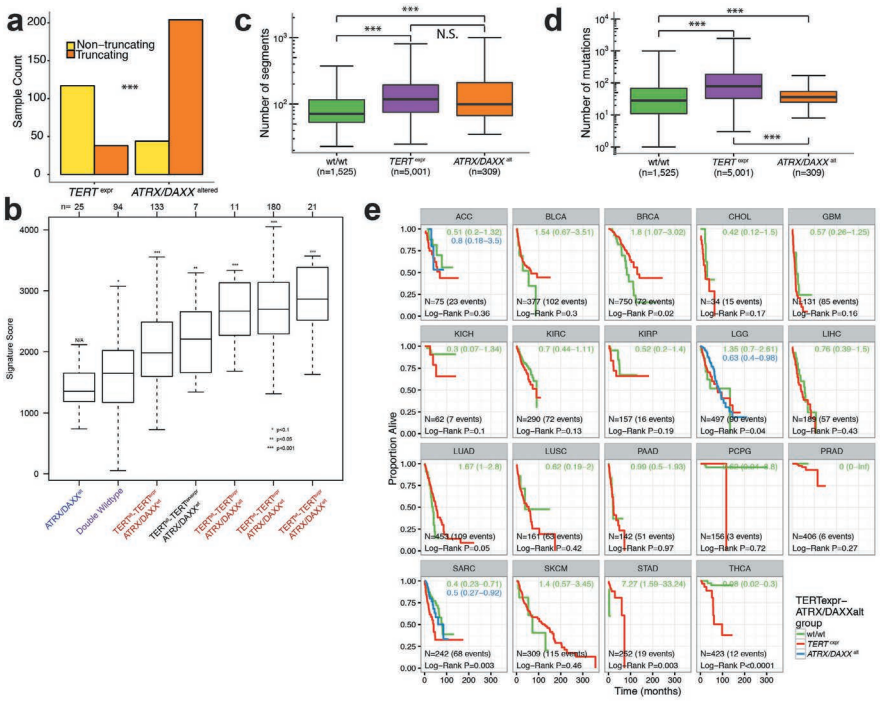
a. Boxplot of *ATRX* expression in groups of *ATRX* alterations. Only non-synonymous mutations are considered. *ATRX* expression in *ATRX* altered groups was compared to the *ATRX* wt group using a two-sided t-test. **b.** Boxplot of TL ratio in groups of *ATRX* alterations. TL ratio in *ATRX* altered groups was compared to the *ATRX* wt group using a two-sided t-test. **c.** Barplot indicating the percentage of *TERRA* expressing tumors per group of *ATRX* alterations. Proportion of *TERRA* expressing samples in *ATRX* altered samples was compared to *ATRX* wt samples using a two-sided Fisher's Exact test. *** P<0.0001; ** P<0.001; * P<0.05; N.S. not significant.

We observed a significant decrease in *ATRX* expression in samples showing mutations, deletions, fusions and structural variants compared to cases with wild type *ATRX* (**Figure 6-12**a). We found that all of types of *ATRX* alteration associated with significantly longer TL compared to wild type *ATRX*, consistent with the previously established association between *ATRX* deactivation and ALT (**Figure 6-12**b).

Recent studies found that *ATRX* knockdown resulted in elevated levels of telomeric repeat containing RNA (*TERRA*) [127]. We estimated *TERRA* levels using RNA sequencing in the extended set and in 566 non-neoplastic samples. Our results demonstrate a significantly higher fraction of *TERRA* expressing samples in all groups of *ATRX* altered samples (**Figure 6-12**, Fisher's Exact test $P < 0.05$) compared to the group of *ATRX* wild type samples. *TERRA* expression was associated with relative TL elongation (two-sided *t*-test TL ratio 1.13, 95%CI 1.07-1.017, $P < 0.0001$). Significantly more samples expressed *TERRA* in tumors lacking *TERT* expression, compared to samples expressing *TERT* (Fisher's Exact OR 0.73, 95%CI 0.65-0.82, P < 0.0001). In contrast, comparison of the *TERRA* levels of tumor samples to their matched normal controls ($n = 596$, Fisher's Exact test P = N.S.) showed no significant differences.

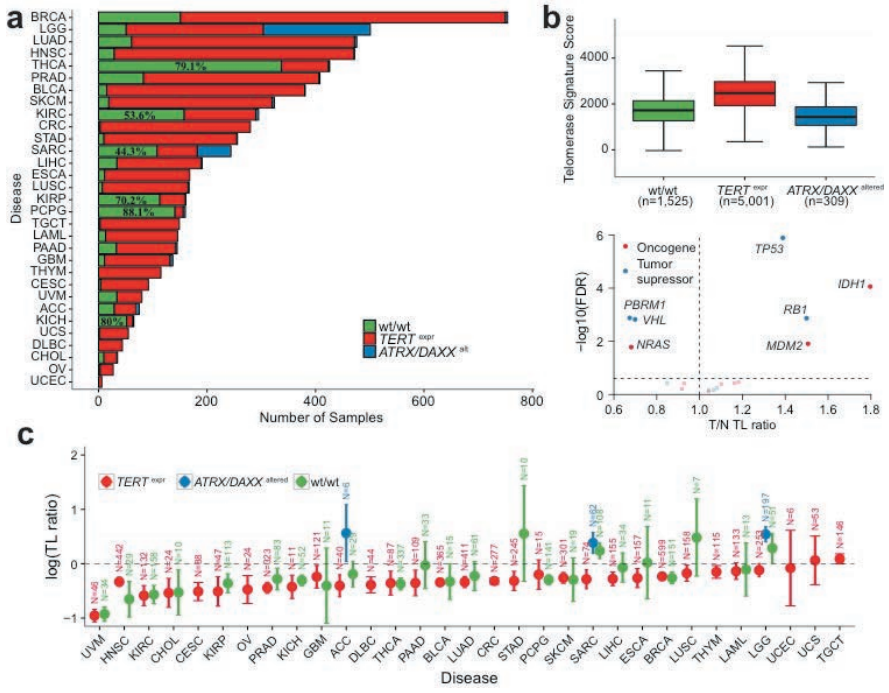
A substantial fraction of cancer samples lacks detectable *TERT* expression and mechanisms of *ATRX* deactivation

Due to the association between somatic *ATRX* and *TERT* alterations with TL, we grouped tumors in the extended set as *TERT* expressing ($TERT^{expr}$, $n = 5{,}001/6{,}835$, 73%) and *ATRX* or *DAXX* altered ($ATRX/DAXX^{alt}$, $n = 309/6{,}835$, 5%). *ATRX/DAXX* mutations were found in 210 *TERT* expressing samples, representing 3% of the cohort. These events were in majority non-truncating, while *ATRX/DAXX* mutations in TERT-negative cases were mostly truncating (**Figure 6-13a**). TERT-expressing *ATRX/DAXX* mutants showed higher telomerase signature scores compared to *ATRX/DAXX* altered samples lacking *TERT* expression (**Figure 6-13b**). Based on these observations we included these samples in the $TERT^{expr}$ category. The remaining 22% of samples had neither detectable *TERT* expression, nor somatic alterations in *ATRX* or *DAXX* (wt/wt, $n = 1{,}525/6{,}835$, **Figure 6-14a**). Both $TERT^{expr}$ and wt/wt groups showed significantly higher telomerase signature scores than the $ATRX/DAXX^{alt}$ group (**Figure 6-14b**, two-sided *t*-test $P < 0.0001$).



↑ **Figure 6-13. Classification of tumors based on TERT expression and ATRX/DAXX genomic alterations.**

a. Comparison of *ATRX* variant classification between samples classified as *TERT*^{expr} and samples classified as *ATRX/DAXX*^{alt}. Silent mutations, *DAXX* alterations, deletions and structural variants were omitted. Truncating variants consist of frameshift, nonsense and splice site variants. Non-truncating mutations consist of missense mutations and in-frame indels. P-value was calculated using a Fisher's Exact test. **b.** Distribution of telomerase signature score across *TERT/ATRX/DAXX* groups in the core set. A small group, *TERT*^{alt}*-TERT*^{expr}*-ATRX/DAXX*^{alt}, has only 2 cases thus were excluded from the comparison. All p values are derived from comparisons with the *ATRX/DAXX*^{alt} group (blue). Red are TERT expressing groups, and purple is the double wild type group. The group in black is TERT^{alt} but without expression. **c.** Number of copy number segments by *TERT*^{expr}*-ATRX/DAXX*^{alt} group. Number of segments between groups was compared using two-sided t-tests. **d.** Mutational burden by *TERT*^{expr}*-ATRX/DAXX*^{alt} groups. Mutational burden between groups was compared using two-sided t-tests. **e.** Survival differences between *TERT*^{expr}*-ATRX/DAXX*^{alt} groups by cancer type. Number of patients included, number of events (deaths) and univariable log-rank P-values are indicated in the bottom left corner for each tumor type. Hazard ratios and 95% confidence intervals comparing survival in the double wild-type group (green) and ATRX/DAXX^{alt} group (blue) relative to the *TERT*^{expr} group are shown in the top right corner of each tumor type. Groups with less than six samples were omitted. *** P<0.0001; ** P<0.001; * P<0.05; N.S. not significant.



↑ **Figure 6-14. A substantial fraction of cancer samples lacks detectable TERT expression and mechanisms of ATRX deactivation.**

a. Frequency of *TERT*^{expr}-*ATRX/DAXX*^{alt} groups across cancer types in the extended set. Definitions for tumor type acronyms can be found in the Online Methods. **b.** Telomerase signature score by *TERT*^{expr}-*ATRX/DAXX*^{alt} groups. Telomerase signature score between groups was compared using two-sided t-tests. **c.** T/N TL ratio by cancer type. *TERT*^{expr}-*ATRX/DAXX*^{alt} groups are shown in different colors. Groups with less than 6 samples were omitted. Error bars indicate 95% confidence interval. Definitions for tumor type acronyms can be found in the Online Methods. **d.** Scatterplot showing gene to T/N TL ratio associations within the wt/wt group (N=1,525). P-values were calculated using a two-sided t-test and adjusted for multiple testing using FDR. *** P<0.0001; ** P<0.001; * P<0.01; N.S. not significant.

TERT^{expr} samples showed relative TL attrition in most cancer types, ATRX^{alt} samples tended to show relative TL elongation and wt/wt samples demonstrated cancer type dependent patterns (**Figure 6-14c**). For example, wt/wt sarcoma (TL ratio 1.27, 95%CI 1.1-1.47) and glioma (TL ratio 1.33, 95%CI 1.01-1.75) samples demonstrated relative TL elongation, whereas wt/wt thyroid (TL ratio 0.69, 95%CI 0.62-0.77), kidney chromophobe (TL ratio 0.74, 95%CI 0.67-0.81) and kidney clear cell (TL ratio 0.57, 95%CI 0.48-0.67) cancer samples showed relative TL attrition. Patterns of relative TL change from the wt/wt group may indicate that some

tumors were detected before having acquired immortalized cells [284], or that other mechanisms to develop ALT exist.

Because this analysis suggests that the prevalence of ALT may be underestimated by *ATRX*/*DAXX* inactivation alone, we sought to compare the prevalence of *ATRX*/*DAXX* alteration to published prevalence of ALT across 26 cancer types, including 40 histological subtypes (***Supplementary Table 6**, available online with the full-text of this article at www.nature.com/ng*) [285]. The prevalence of ALT exceeded that of *ATRX*/*DAXX* alterations in seven histological subtypes and was reduced in two, reinforcing the notion of cancer type specific and *ATRX*/*DAXX* independent ALT mechanisms.

The intermediate telomerase activity levels and variable TLs suggested that the wt/wt group comprises a heterogeneous set of tumors, some of which may have undergone ALT through mechanisms independent of *ATRX* and *DAXX*. The wt/wt samples were most frequent amongst pheochromocytoma (*n* = 141/160, 88%), kidney papillary (*n* = 113/161, 70%), kidney chromophobe (*n* = 52/65, 80%) and thyroid cancer (*n* = 343/449, 79%). Seven of the 101 wt/wt samples in the core set showed a somatic *TERT* alteration in absence of *TERT* expression, including a TERTp mutation (*n* = 1), *TERT* amplification (*n* = 2), TERTp or *TERT* structural variant (*n* = 4).

To better understand alternative mechanisms that may contribute to telomere maintenance, we conducted an unsupervised search for TL ratio associated somatic alterations within the wt/wt group in the extended set. This analysis recovered strong associations between relatively long TL, *IDH1* (TL ratio 1.8, 95%CI 1.39-2.33, FDR=0.0001), and *TP53* (TL ratio 1.39, 95%CI 1.23-1.57, FDR<0.0001), both of which were also identified in the analysis of TL length in the entire dataset. The finding of *IDH1* may reflect a glioma-specific effect. In addition to *IDH1* and *TP53*, *RB1* (TL ratio 1.5, 95%CI 1.2-1.87, FDR=0.001) and *MDM2* (TL ratio 1.51, 95%CI 1.14-2, FDR=0.01) were revealed to associate with relatively long TLs. The finding of *RB1* is consistent with experimental data demonstrating markedly elongated telomeres in Rb1 deficient mice independent of telomerase [286, 287]. In the opposite direction, somatic alterations in *PBRM1* (TL ratio 0.67, 95%CI 0.55-0.83, FDR=0.001), *NRAS* (TL ratio 0.68, 95%CI 0.52-0.9, FDR=0.02) and *VHL* (TL ratio 0.7, 95%CI 0.57-0.85, FDR=0.002) were associated with relative TL shortening (**Figure 6-14d**).

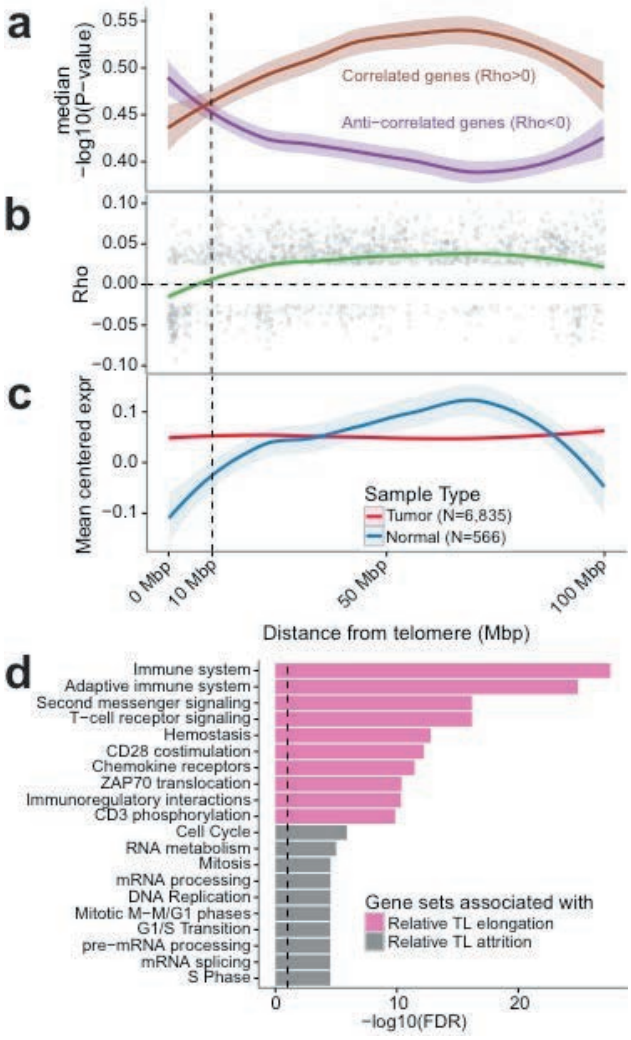
We compared global genomic characteristics between the three groups and found that wt/wt samples contained a lower number of copy number segments (two-sided t-test $P < 0.0001$, **Figure 6-13c**), and a reduced mutation rate (two-sided t-test $P < 0.0001$, **Figure 6-13d**) compared to $TERT^{expr}$ and $ATRX/DAXX^{alt}$ sam-

ples. Survival analyses within cancer types showed that wt/wt samples perform better than TERT^{expr} samples in sarcoma and thyroid carcinoma (**Figure 6-13e**). Telomerase activity score in wt/wt samples in these two cancer types were significantly lower (two-sided *t*-test *P* = 7.0e-7 and *P* = 7.8e-3, respectively), suggesting telomerase activity might represent a relevant prognostic factor in these cancers.

Telomere position effect

Provided the quantification of TL, we sought to relate TL to gene expression changes using the extended set and 566 adjacent tissue normal samples (***Supplementary Table 7**, available online with the full-text of this article at www.nature.com/ng*). We first investigated the telomere position effect (TPE), a phenomenon that describes TL mediated transcriptional regulation [288]. We found expression of genes close to telomeres to be negatively correlated with TL, and such effect attenuated with increased distance to telomeres (**Figure 6-15a-b**). Since tumors on average had shorter telomeres than matched normal samples, genes close to telomere showed higher expression in tumor than normal (**Figure 6-15c**). The TPE was no longer detectable beyond the distance of 10 million base pairs from telomeres.

To identify genesets associated with TL attrition and elongation in an unsupervised fashion, we associated gene expression to TL attrition and elongation. Pathway analysis of the top 500 correlated genes indicate that relative TL attrition and elongation were associated with immunoreactive and proliferative signatures, respectively (**Figure 6-15d**). These patterns may reflect differences in telomere maintenance rather than telomere attrition and elongation, and describe frequently observed patterns of expression across cancer, potentially related to more and less aggressive tumors [289].



↑ **Figure 6-15. Telomere position effect.**

a. Negative log10(P-value) for genes showing a correlation or an anti-correlation to tumor TL shown relative to the distance to the respective telomere. Spearman correlation tests were conducted individually within cancer types. P-values for *n* = 3,477 genes were averaged across cancer types. **b.** Spearman correlation coefficient across all cancer types for *n* = 2,016 genes significantly associated to TL (FDR < 0.25) shown relative to the distance to the respective telomere. **c.** Average gene expression (in TPM, transcripts per million) for tumor and normal samples relative to distance to the respective telomere. Mean expression was calculated for *n* = 3,477 genes. **d.** Reactome gene expression pathway analysis on the extended set (*n* = 6,835 samples) of the top 500 genes most associated with relative TL elongation and relative TL shortening, respectively. Top ten pathways for each set of 500 genes shown.

Discussion

This study represents the largest in-depth integrative analysis of telomere length and related somatic alterations performed to date. As expected, tumor TL was shorter compared to normal tissue TL, and TL was inversely correlated with age in both tumor and non-neoplastic tissues. Amongst *TERT* expressing tumors, 32% carried at least one of three *TERT* abnormalities: promoter mutation, amplification or chromosomal rearrangement. Of *TERT* expressing and wild type tumors, 91% was *TERT* promoter methylated compared to 40% of tumors lacking *TERT* transcription. This paradoxical association between *TERT* promoter methylation and increased *TERT* expression may result from loss of CTCF binding, a transcriptional repressor reported to bind to the unmethylated *TERT* promoter [290]. Structural *TERT* variants have been documented [258, 291, 292] and we detected these across several novel cancer types, including sarcoma, prostate and liver carcinoma. Hepatitis B virus and adeno-associated virus type 2 integration in *TERT* was found in about 5% of hepatocellular carcinomas [293]. The absence of similar observations in other virus-related cancer types, including head and neck, bladder, and gastric cancers, suggest it is a liver cancer specific mechanism thus was not considered in our study.

Whether *TERT* expression directly translates to active telomerase activity is unclear. Only the full-length transcript and not known isoforms were found to activate telomerase [274, 281]. *TERC* expression is additionally needed for telomerase activity [294]. An estimated 5-10% of *TERC* and *TERRA* carry the poly-A tails required for oligo-dT primer-based RNA sequencing quantification [295-297]. We detected *TERC* and *TERRA* expression across our cohort, suggesting that if expression is sufficiently high, transcripts can be detected using conventional RNA sequencing approaches. Lacking telomerase activity data, we sought to infer telomerase activity using a gene signature driven approach. Our results suggested a positive correlation between *TERT* expression and telomerase activity, corroborating recent findings in bladder cancer [126]. We also observed a positive correlation between *TERC* expression and telomerase activity, as well as increased telomerase activity in *TERT* expressing tumors compared to *ATRX/DAXX* altered tumors and double wild-type tumors. This gene signature may serve as a useful proxy for estimating telomerase activity using transcriptional profiles.

Deactivation of *ATRX* and/or *DAXX* has been related to ALT [262, 263, 298], and was observed in five percent of the cases in our core set. A detailed review of *ATRX* somatic changes revealed a large spectrum of potentially protein truncating changes, including inactivating mutations, deletions and structural variants. *DAXX*

alterations were much less frequent. Our analysis reinforced the association of inactivate *ATRX/DAXX* and ALT, demonstrating relative TL elongation in samples affected by somatic alterations in one of these two genes and a higher frequency of *TERRA* expression in tumors with these alterations.

The twenty two percent of cases that lacked detectable *TERT* expression and known ALT related abnormalities provide an important result from our analysis. It is plausible that *TERT* transcription below the currently applied detection threshold may be sufficient for telomere maintenance, that not all tumors harbor immortalized cells with a telomere maintenance mechanism [284], or that additional telomere maintenance mechanisms exist. Such mechanisms may involve some *RB1* and *TP53* alterations, as somatic changes in these genes were associated with telomere elongation within this group. Interestingly, double wild-type tumors were mostly pheochromocytoma and paraganglioma, kidney chromophobe and papillary thyroid tumors. Pheochromocytoma is generally (>90%) labeled benign. Chromophobe renal cell carcinoma and papillary thyroid cancer are malignant tumors but are generally well differentiated, infrequently metastasize and demonstrate a favorable outcome compared to other cancer types [299-302]. Future studies are needed to elucidate the telomere maintenance mechanisms, or lack thereof, in these tumor types.

In summary, our analysis has broadened the scope of potential *TERT* activating changes, completed the spectrum of *ATRX* truncating alterations and provided new insights into the telomere biology of tumors lacking these classic alterations. Combined these findings help us better understand telomere biology, and opens avenues for functional studies to help us understand how to target this crucial pathway in oncogenesis.

Methods

Sample selection

The 31 cancer types included in this study were: **LAML** acute myeloid leukemia; **ACC** Adrenocortical carcinoma; **BLCA** Bladder Urothelial Carcinoma; **LGG** Brain Lower Grade Glioma; **BRCA** Breast invasive carcinoma; **CESC** Cervical squamous cell carcinoma and endocervical adenocarcinoma; **CHOL** Cholangiocarcinoma; **COAD** Colon adenocarcinoma; **ESCA** Esophageal carcinoma; **GBM** Glioblastoma multiforme; **HNSC** Head and Neck squamous cell carcinoma; **KICH** Kidney Chro-

mophobe; KIRC Kidney renal clear cell carcinoma; **KIRP** Kidney renal papillary cell carcinoma; **LIHC** Liver hepatocellular carcinoma; **LUAD** Lung adenocarcinoma; **LUSC** Lung squamous cell carcinoma; **DLBC** Lymphoid Neoplasm Diffuse Large B-cell Lymphoma; **OV** Ovarian serous cystadenocarcinoma; **PAAD** Pancreatic adenocarcinoma; **PCPG** Pheochromocytoma and Paraganglioma; **PRAD** Prostate adenocarcinoma; **READ** Rectum adenocarcinoma; **SARC** Sarcoma; **SKCM** Skin Cutaneous Melanoma; **STAD** Stomach adenocarcinoma; **TGCT** Testicular Germ Cell Tumors; **THYM** Thymoma; **THCA** Thyroid carcinoma; **UCS** Uterine Carcinosarcoma; **UCEC** Uterine Corpus Endometrial Carcinoma; **UVM** Uveal Melanoma.

We removed cases that were annotated as having a bad DNA quality or failing a QC step. We also removed technical replicates and samples from patients who had prior systematic treatment or revised pathological diagnosis. Tumor and control samples from each patient were paired, selecting a blood-derived control if both blood and solid tissue control were available. A comprehensive sample selection procedure can be found in the Supplementary Note (*available online with the full-text of this article at www.nature.com/ng)* and a schematic is shown in **Figure 6-2a**.

Data Generation

Raw RNA and DNA sequencing data (n = 35,978 samples) was downloaded from CGHub and processed through a flowr pipeline consisting of various components depending on the data type [303, 304]. Processed clinical, gene expression, DNA methylation, copy number segmentation and mutation data were downloaded and compiled from previously published TCGA papers, the Firehose data portal (Broad Institute, Cambridge, MI) and the TCGA data portal (NIH, Washington, DC). GISTIC was employed on the processed segmentation files to infer somatic copy number changes [170]. A more detailed description of the data collection and generation process is described in the Supplementary Note.

Telomere length quantification

Quantification of telomere length (TL) was performed using the TelSeq tool [264]. Briefly, this tool counts the number of reads containing any (range k to ∞) amount of telomeric repeats (n_k), or TTAGGG $_{[k,\infty)}$, and the GC-adjusted coverage s. The quotient of the number of telomeric repeats and the GC-adjusted coverage is then multiplied by the average chromosome length c (c=332720800/46000, in Kbp), resulting in the estimated telomere length in Kbp l.

$$l = c * \frac{n_k}{s}$$

We used a k of 7, as per the author's recommendations. Using the default settings from TelSeq, this calculation is done individually for each read group within a sample. In order to calculate the average TL for each sample, the weighted average length was used, supplying the total number of reads in each read group as weight.

RNA-seq BAM files were also processed using TelSeq (k = 7) to obtain an estimate of *TERRA* expression. Because *TERRA* expression demonstrated a bimodal distribution, and over 40% of samples lacked any detectable *TERRA* expression, *TERRA* expression was dichotomized using a cutoff of zero.

TERT promoter mutation detection

We used GATK pileup to determine bases mapping to each position 200 bp upstream of the *TERT* transcription start site (chr5:1295162-1295404, hg19) in 1,771 tumor samples and matched normal controls from 20 cancer types where whole genome or low-pass whole genome sequencing data was available [305].
We first performed an unsupervised screen, testing each of 200 sites in every tumor and matched normal. For each site we required a minimum coverage of 10 reads, minimum variant allele fraction of 25% in tumor and maximum variant allele fraction of 2% in normal. Fisher's test was used to determine significance and a threshold of 0.05 was used. Samples with *TERT* expression below a threshold of 2 reads were filtered from the analysis. Four sites demonstrated significance in at least one sample: C228, C250, C242/243 and C169. We then determined the nucleotides called in each of the affected sites, and found that all affected sites underwent a C>T transition.

Next, we performed a supervised analysis using the four sites above in all tumor samples to 5nd papillary thyroid carcinoma (n = 384) for a grand total of n = 1807 *TERTp* mutation calls. In cases where WGS and targeted sequencing-based calls disagreed we selected the mutated variant.

Structural variant detection

We used SpeedSeq in order to call structural variants in WGS-based BAM files [247]. The SpeedSeq pipeline was built using our in-house pipeline building tool Flowr, and ran on a high performance computing cluster at the UT MD Anderson Cancer Center [303]. BAM files were realigned using BWA-MEM [306]. The re-

sulting calls in VCF format were filtered against matching normal. Only somatic events with at least 3 supporting reads were retained.

Because we were only interested in variants involving *TERT* (chr5:1253287-1315162, including 20kb upstream of the TSS) and *ATRX* (chrX:76760356-77041719), variants not overlapping one of these regions were excluded. An overview of all included variants can be found in ***Supplementary Table 2*** *(available online with the full-text of this article at www.nature.com/ng).*

Fusion transcript detection

All fusion transcripts were detected using the pipeline of RNA-seq Data Analysis (PRADA) as previously described [271, 272]. Briefly, chimeric fusion transcripts were detected from the realigned bam files mapped to a combined genome and transcriptome reference, based on the evidence of both discordant read pairs and fusion-spanning reads, where discordant read pairs represent the paired read-ends that map uniquely to two protein coding genes, fusion-spanning reads mapped to the exon–exon junctions between two coding genes. Then the confidence of the detected fusions were evaluated based on the number of junction spanning reads and discordant read pairs, gene partner uniqueness, gene homology, open reading frame preservation, transcript allele fraction and presence of DNA breakpoints in adjacent distal regions. A complete overview of all gene fusions can be found on our fusion web portal (http://tumorfusions.org/). All fusions harboring *TERT* or *ATRX* transcripts were considered bona fide fusion calls based on the criteria described previously and therefore included in this study. A summary of included fusions can be found in ***Supplementary Table 2*** *(available online with the full-text of this article at www.nature.com/ng).*

TERT isoform detection

PRADA aligned RNA-seq BAM files were available for *n* = 6,625 (97% of the extended set) samples, missing data on ovarian carcinoma (OV, *n* = 27) and prostate cancer (PRAD, *n* = 175).

We applied the mixture of isoforms (MISO) model to each BAM to infer the relative abundance of full-length and minus beta *TERT* transcripts in each sample [282]. Because the detection of isoforms is limited by the high GC-content of *TERT* combined with a 3' bias by regular polyA enrichment RNAseq protocol and unevenly spread expression across exons, we limited the analysis to exons 6-9 only and applied a coverage threshold of at least twelve reads within this region. These filters reduced the sample cohort to 24% (*n* = 1,201) of 5,001 tumors classified

as *TERT* expressing in the extended set. The exon model used to infer isoform abundance has been included in ***Supplementary Table 4*** *(available online with the full-text of this article at www.nature.com/ng).* Briefly, the minus beta isoform skips exons 7 and 8, and samples demonstrating predominantly minus beta transcripts should therefore demonstrate low coverage in exons 7 and 8 relative to 6 and 9, and show junction spanning reads between exons 6 and 9.

Structural variant breakpoint super enhancer and ChIP-seq analysis

Super enhancer coordinates were downloaded from dbSuper as of March 15th, 2016 [268]. At this time, the database consisted of 65,950 super-enhancers across 107 tissue and/or cell types. After excluding low-complexity regions and telo- and centromeres, regions in this database spanned approximately 18% of the genome. We used the UCSC genome browser to browse the database, and inspected the position of each of the juxtaposed genomic coordinates [307]. This analysis found overlapping super-enhancers in 11/17 TERTp structural variants, and this was significantly more than expected by chance (Chi-Square test P=0.001).

Publicly available ChIP sequencing data was downloaded from the NIH Roadmap Epigenomics Mapping Consortium [308]. This dataset consists of 183 biological samples, from multiple individuals, sequencing centers, tissue and/or cell types and was further consolidated into 111 unique epigenomes. We identified n = 44 tissue and/or cell types with H3K27ac (enhancer) sequencing, n = 53 tissue and/or cell types with H3K4me1 (enhancer) sequencing and n = 55 tissue and/or cell types with H3K27me3 sequencing (inactive promoter). Altogether we collected data from 71 distinct tissue and/or cell types across these three marks. Collected BED files were converted to BAM format and a total number of reads was computed for each sample and mark.

ChIP-seq data was downloaded as previously described. For each structural variant proximal (adjacent to the *TERT* promoter) and distal breakpoint (juxtaposed to the *TERT* promoter) we calculated the number of reads mapping to each of three histone marks (H3K27ac, H3K27me3 and H3Kme1) individually within each tissue and cell/type. Breakpoint coordinates were flanked on either side to form a 2kbp bin. Read counts were normalized to RPKM based on the total number of reads for each mark and sample. We then compared the RPKM between distal and proximal breakpoints for each histone mark for each set of breakpoint coordinates individually (**Figure 6-5d**), and subsequently with all breakpoints pooled together (**Figure 6-6b**). P-values were calculated using a two-sided *t*-test for each mark individually.

Telomerase activity signature inference

Microarray gene expression data from eight dedifferentiated liposarcoma samples were downloaded from the Gene Expression Ominibus (GEO, GSE20559). We performed a differential expression analysis comparing four telomerase positive and four telomerase negative (ALT) samples and identified 1302 genes associated with telomerase positive tumors (fold-change ≥ 1.5). Intersecting with 420 genes associated with embryonic stem cells [309], which are known to be telomerase positive, further refined this geneset and resulted in a list of 43 genes (***Supplementary Table 4***, *available online with the full-text of this article at www.nature.com/ng*). Validation of the resulting gene signature using matched telomerase activity and RNA sequencing data from eleven urothelial cell carcinoma cell lines provided some evidence that this gene signature may be able to predict telomerase activity (Rho=0.58, P-value = 0.07, **Figure 6-9a**). A detailed description of the strategy used to infer a telomerase activity signature may be found in the Supplementary Note.

Candidate gene selection and telomere length association

In order to narrow down the list of candidate gene and increase the FDR adjusted significance threshold, we took the union of the MutSig2CV gene list (FDR < 0.05, downloaded and compiled from publicly available Firehose analyses) from all cancer types, and combined this list with genes within significant GISTIC 2.0 peaks (FDR < 0.10) after filtering peaks larger than 1 Mbp [310].

Each gene was annotated as putative tumor suppressor or putative oncogene depending on the mutational patterns (see Supplementary Note) or whether it was found within an amplification or deletion peak. In those cases where the mutation-based classification did not agree with the copy number based classification (eg. gene with highly frequent hotspot mutations was found in a deletion peak), we selected to prefer the mutation-based classification. Genes lacking classification and genes where the mutation-based classification demonstrated contradicting evidence (eg. genes that could be classified as either tumor suppressor or oncogene) were dropped, unless the gene was present on a list of telomere related genes, in which case it was annotated as significantly mutated gene. The list of candidate genes was further reduced by removing multiple genes from the same peak, additionally leaving only those genes closely tied to telomere function when multiple genes present in the peak (eg. shelterin complex genes) and genes that were also found in the MutSig2CV gene list. The final list of genes

can be found in ***Supplementary Table 5*** *(available online with the full-text of this article at www.nature.com/ng).*

For each candidate gene and sample we determined whether it was altered or not depending on the classification. For genes classified as tumor suppressors we classified samples as altered when they showed a somatic mutation, a focal deletion or both. For genes classified as oncogenes we classified the sample as altered when they showed a somatic mutation, focal amplification or both. For genes classified as significantly mutated gene only somatic mutations were counted. Each candidate gene was then correlated to the TL ratio in all samples and within diseases individually. This analysis was repeated specifically within samples classified as double wild type.

Statistical analysis

Summary statistics of telomere length was provided in mean, SD, and range by tumor type and tissue type. Telomere length was transformed to the logarithmic scale due to skewness.

A linear mixed model was used to estimate mean telomere length of tumor and normal samples for each tumor type. Age, center and gender were included as a covariate in all models. Interaction between disease type and center, gender, and age were assessed and none was significant. Patient was modeled as a random effect in the mixed model to account for correlation between samples from the same patient.

Linear regression was used to correlate TL ratio to various biomarkers. Backwards elimination was then used to select the best model with only significant factors. All tests were two-sided and p-values of 0.05 or less were considered statistically significant.

Spearman correlation was used to associate *TERT* expression and methylation, and to correlate gene expression and TL. A two-sided t-test was used to correlate somatic alterations and TL. P-values were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate (FDR). An FDR of 0.25 or lower was considered statistically significant.

Survival curves were estimated and plotted using the Kaplan-Meier method. Log-rank tests were used to compare curves between groups. Univariate and multivariate Cox modeling were used to compute hazard ratios and confidence intervals. All statistical analyses were carried out using SAS version 9 (SAS Institute, Cary, NC) and R (R Foundation for Statistical Computing, Vienna, Austria).

URLs

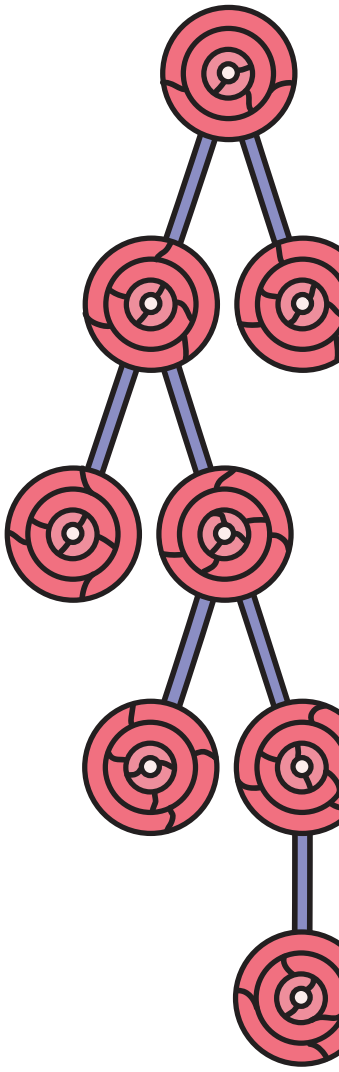
PRADA fusion portal www.tumorfusions.org

Acknowledgements

We thank Drs. Dian Jiao and Daniel Jackson at the High Performance Computing facility, MD Anderson Cancer Center for assistance with data downloading and storage. We thank Drs. Ian R Watson and Terence Wu for insights on TCGA melanoma project; Dr. David Wheeler for insights on TCGA liver cancer project. We thank Drs. Sumit Borah and Thomas R. Cech for sharing telomerase enzymatic activity data of the 23 urothelial cancer cell lines. We thank Dr. Dominique Broccoli for sharing the 14 gene signature from GSE20559. We thank Dr. Xiaojing Wang for help with CPTAC data and Andrea Barbo, M.Sc for statistical advice. We thank Bradley Murray for suggestions regarding the analysis of copy number data. We acknowledge the contributions from the TCGA Research Network. This project is supported by grants from the National Institutes of Health P50 CA127001, R01 CA190121, CA143883 and P01 CA085878; and the Cancer Prevention & Research Institute of Texas (CPRIT) R140606. This work is dedicated to all patients who consented to have their samples analyzed by The Cancer Genome Atlas Network.

Author contributions

F.P.B. was involved in all aspects of data analysis. W.W. performed linear mixed modeling. M.T. and S.B.A. analyzed whole genome sequencing data. E.M. analyzed gene expression data. X.H. and Q.W. performed gene fusion analysis. K.A. was involved in methylation and epigenetics analysis. S.S., X.S. and J.Z. collected and analyzed low pass sequencing data. T.L. collected clinic data of the solid tissue samples. Z.S. designed the telomerase signature score. F.P.B., Z.S. and R.G.W.V conceived the study and wrote the paper. Z.S and R.G.W.V supervised the study.



CHAPTER 7

Reconstructing the Molecular Life History of Gliomas

This chapter is based on the following review paper (published in print)

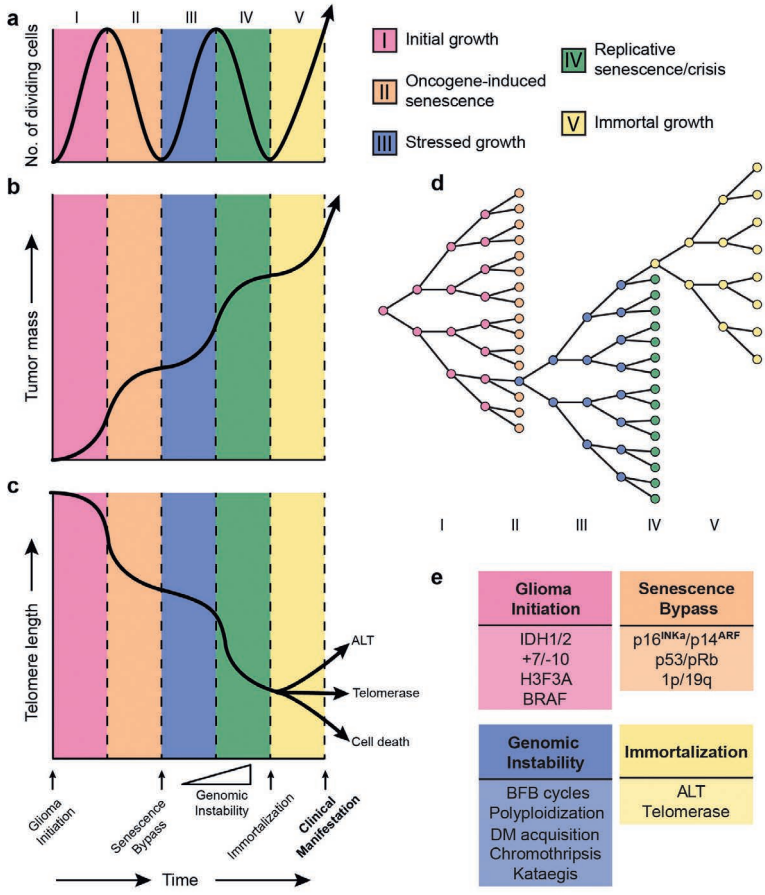
Barthel FP, Wesseling P, Verhaak RGW. *Reconstructing the molecular life history of gliomas.* Acta Neuropathol. 2018 May;135(5):649-670. doi: 10.1007/s00401-018-1842-y.

Abstract

At the time of their clinical manifestation, the heterogeneous group of adult and pediatric gliomas carry a wide range of somatic genomic alterations, ranging from somatic single nucleotide variants to structural chromosomal rearrangements. Somatic abnormalities may have functional consequences, such as a decrease, increase or change in mRNA transcripts, and cells pay a penalty for maintaining them. These abnormalities therefore must provide cells with a competitive advantage to become engrained into the glioma genome. Here, we propose a model of gliomagenesis consisting of the following five consecutive phases that glioma cells have traversed prior to clinical manifestation: I initial growth; II oncogene-induced senescence; III stressed growth; IV replicative senescence/crisis; V immortal growth. We have integrated the findings from a large number of studies in biology and (neuro)oncology and relate somatic alterations and other results discussed in these papers to each of these five phases. Understanding the story that each glioma tells at presentation may ultimately facilitate the design of novel, more effective therapeutic approaches.

In this charter, I will reflect on what I've learned about the molecular anatomy of these tumors to propose five phases in gliomagenesis that occur more or less sequentially and ultimately lead to its clinical manifestation (**Figure 7-1**). Each phase is characterized by distinct molecular alterations and phenotypic characteristics, such as differences in growth dynamics and evolutionary mechanisms. A critical assumption in this model is the existence of two growth barriers, which we refer to as oncogene-induced and replicative senescence. Similar barriers have been described in detail in the context of cultured epithelial cells and fibroblasts and much of this work has paved the road for our understanding of these mechanisms in gliomagenesis [311-313].

In the model we propose, the first phase of initial growth (phase I) follows the acquisition of a glioma initiation event and is characterized by aberrant proliferation of pre-tumor cells. Continued oncogenic exposure may impede tumor growth and trigger a durable form of cell cycle arrest termed oncogene-induced senescence (phase II) in a majority of tumor cells. Some phase I/II cells may acquire molecular changes to bypass oncogene-induced senescence and continue growth in spite of unfavorable and stressful conditions including DNA damage and dysfunctional telomeres. Continued growth despite incremental genomic instability marks the third phase of stressed growth (phase III). This second round of glioma cell growth under harsh conditions triggers a second round of durable growth arrest termed replicative senescence (phase IV) and in some cases, brings forth a state of cellular crisis characterized by widespread cell death. Rare cells may acquire stem-like characteristics and a means to continue growth indefinitely, giving rise to a final phase of immortal growth (phase V). Such phase V glioma stem-like cells uphold the tumor progenitor cell population *via* their capacity for self-renewal and may also give rise to more differentiated and growth arrested stage IV cells, losing their stem-like properties. In this manuscript, we systematically review the evidence for this model and propose candidate mechanisms where definitive evidence is lacking. Acknowledging that this model is an abstract simplification of gliomagenesis, we also provide some examples of exceptions and conflicting evidence. While the focus of this chapter is on various molecular categories of diffuse glioma recognized by the most recent WHO classification, a few examples in the realm of non-diffuse gliomas are touched upon as well.



↑ **Figure 7-1. Model of the molecular life history of gliomas, prior to becoming clinically manifest.**

The temporal sequence of events can be subdivided in five phases (I – V) represented in different colors. **a.** The number of dividing cells (or proliferation rate) across each phase. Proliferation peaks towards the end of growth phases and dips going into senescence phases. **b.** The tumor mass across each phase. Tumor mass increases exponentially during growth phases and logarithmically during senescence phases. **c.** Telomere length across each phase. Telomere length over time follows a pattern that is inverse to tumor mass. **d.** Cell doubling diagram indicating the growth barriers (senescence phases) and resulting selection bottleneck. **e.** Somatic alterations associated with different phases in gliomagenesis. The timing of each event is indicated on the x-axis of panel C. Genomic Instability events are accumulated during phase III-IV. Of note, this model is a simplified representation of true gliomagenesis. The x-axis is not drawn to scale, in part because the duration of the phases likely varies from cell to cell and between various tumor types. Furthermore, the position of the curves is arbitrary as cells in a tumor may not be in sync. BFB = breakage-fusion-bridge, DM = double minute, ALT = Alternative lengthening of telomeres

A Model for the Temporal Molecular Pathogenesis of Gliomas

Phase I: Initial Growth

The theory that cancer results from accumulation of mutations over time, in a subset of patients combined with contribution of inherited risk factors, has been around for over six decades and has been refined over the years [314-319]. For the purpose of this review we will consider a glioma initiation event to be the first acquired (somatic) event towards developing glioma. This event should provide a competitive growth advantage, either by directly increasing proliferation or by creating the conditions in which increased proliferation may happen. Cells and their progeny characterized by such an event will be primed to outcompete neighboring cells giving rise to an initial submicroscopic tumor mass. While germline events that contribute to glioma risk precede such glioma initiation events, their incomplete penetrance suggests that they cannot be considered as causal for glioma formation and instead prime the environment for tumor formation [320]. Furthermore, even if a glioma initiation event marks the first somatic event in the formation of a tumor it may not be responsible for initiating growth directly. Instead, this event may promote tumorigenesis indirectly *via* stochastic activation of oncogenes or repression of tumor suppressors. Genetic or epigenetic selection pressures will prioritize daughter cells with growth advantages over those without and daughter cells with lethal genotypes will rapidly disappear [51, 321, 322].

IDH-mutant diffuse gliomas

Mutations in *IDH1*/*IDH2* are commonly considered to be glioma initiating. Several studies have shown that they are amongst the few alterations highly shared amongst gliomas at first presentation and their recurrences [99, 323, 324] which is explained by their presence in the cell of origin and all cells derived thereof. Comparing multiple biopsies from the same tumor, IDH mutation can be confidently detected in each tumor segment and thus fit the proposed criteria of a glioma initiation event [73, 99, 325]. *In vitro* experiments have demonstrated that IDH mutations alone are sufficient to reprogram the transcriptome and epigenome of normal cells to prevent these cells from entering a terminally differentiated state [82-84].

IDH dysregulation likely contributes to gliomagenesis *via* the accumulation of the oncogenic metabolite R(-)-2-hydroxygulatarate (2HG) [326, 327]. Wildtype IDH enzymatically converts isocitrate into a-ketoglutarate (a-KG) as part of the

citric acid cycle, whereas mutant IDH metabolizes a-KG into 2HG [328, 329]. Both mutant and wildtype IDH alleles are therefore essential for the oncogenic function of IDH. IDH mutations in glioma result in genome-wide hypermethylation [23, 84], most likely due to effects of 2HG on the Ten-eleven translocation methylcytosine dioxygenase (Tet) family of proteins [81, 82, 330]. This hypermethylation may provide a growth advantage to cancer cells due to the epigenetic activation of oncogenes *via* stochastic activation of alternative gene regulatory programs, some conferring added fitness [51]. One such mechanism in glioma may be through methylation induced disruption of a CCCTC-binding factor (*CTCF*) binding site, resulting in aberrant activation of Platelet-Derived Growth Factor Receptor Alpha (*PDGRFA*) [50].

Although most experimental models of IDH involve overexpression of mutant IDH *in vitro*, several transgenic mouse models have been described [331, 332]. Early transgenic models showed that conditional knock-in of mutant IDH in the murine brain led to perinatal lethality [333]. A more recent inducible model demonstrated that mutant IDH led to increased proliferation and infiltration in the CNS parenchyma of murine neural stem cells [334]. Though these mice eventually died due to hydrocephalus and did not develop malignant tumors, they showed symptoms of the initial phases of gliomagenesis. Thus, while experimental models of mutant IDH are generally insufficient to cause glioma, mutant IDH leads to changes that could be interpreted as early tumor development.

IDH-wildtype diffuse astrocytomas

Approximately 70% of IDH-wildtype diffuse astrocytomas are characterized at the molecular level by a single copy loss of chromosome 10 and gain of chromosome 7 (+7/-10) [15, 335]. Loss or diploid loss of heterozygosity of chromosome 10, and of chromosome arm 10p in particular have been reported to occur more frequent and may precede gain of chromosome 7 [336]. Based on evolutionary modeling using primary-recurrent tumor pairs and multisector tumor sampling, several independent groups have found that +7/-10 is homogeneous and longitudinally preserved and thus likely the first and glioma initiation event in a large fraction of IDH-wildtype diffuse astrocytomas/glioblastomas [71, 73, 105, 139, 337]. A recent study suggested that gains of chromosome 7 likely occur early in tumorigenesis, amongst the first 10% of driver events [337]. In the past several years there has been a lot of interest in the role of *TERT* promoter mutations in oncogenesis and an increasing body of evidence suggests that these mutations precede +7/-10 [103]. Nevertheless, a potential role for *TERT* promoter mutations

to promote proliferation in the initial growth phase and as a glioma initiation event is speculative and will be discussed later in this review.

Chromosome 7 is home to several oncogenes that have been implicated in gliomagenesis such as Cyclin Dependent Kinase 6 (*CDK6*), MET Proto-Oncogene (*MET*) and Epidermal Growth Factor Receptor (*EGFR*), while chromosome 10 hosts several tumor suppressor genes, including Tet family member Tet Methylcytosine Dioxygenase 1 (*TET1*) and Phosphatase and Tensin Homolog *(PTEN)*. Though these genes comprise the prime suspects, it is unlikely that they alone are responsible for initiating glioma development [105]. Studies across different diseases and in various model organisms have shown that large chromosomal copy number changes led to gross gene dosage fluctuations impacting various specific and general cellular functions [338]. Such changes may therefore act in concert to promote tumor development.

Some IDH-wildtype diffuse gliomas show cytogenetically intact chromosomes 7 and 10, implying that other initiation events give rise to these tumors as well. Such events may include activating or inactivating alterations in the phosphoinositide 3-kinase (PI3K), receptor tyrosine kinase (RTK) and mitogen-activated protein kinase (MAPK) pathways [339]. PI3K-pathway alterations include mutations in PI3-Kinase Subunit Alpha (*PIK3CA*), PI3-Kinase Subunit P85-Alpha (*PIK3R1*), or inactivation the aforementioned tumor suppressor *PTEN* [340-342]. Glioma initiation events may also include point mutations in RTK-pathway genes such as *EGFR* and *PDGFRA* or in MAPK-pathway genes such as Neurofibromin 1 (*NF1*) [22]. Much is already known about the effect of these mutations on cancer growth but additional research is needed to secure their potential role as glioma initiation events.

A particular subgroup of diffuse IDH-wildtype gliomas is characterized by mutations in H3 histone family members and these gliomas occur most often in children [86, 283, 343]. The diffuse midline glioma, H3 K27M-mutant, shows a lysine to methionine substitution at position 27 of the H3 Histone Family Member 3A (*H3F3A*) or Histone Cluster 1 H3 Family Member 3B (*HIST1H3B*) gene and is included in the WHO 2016 classification as a separate entity. Other H3-mutant diffuse gliomas in children and adolescents occurring predominantly in the cerebral hemispheres often show H3 G34R/V mutation (implying a glycine 34 to arginine or valine substitution) [86, 106, 343, 344]. In contrast to hypermethylated IDH-mutant gliomas, these H3-mutant gliomas have a general DNA hypomethylation phenotype [345]. A recent study showed that expression of mutant H3 K27M in neonatal mice brains led to ectopic proliferation, indicating a possible

pre-cancerous change [346]. Although additional support is needed, combined with their apparent mutual exclusivity with mutations in IDH and changes characteristic for IDH-wildtype tumors, these findings indicate that H3 K27M and H3 G34R/V mutations may be glioma initiation events [347].

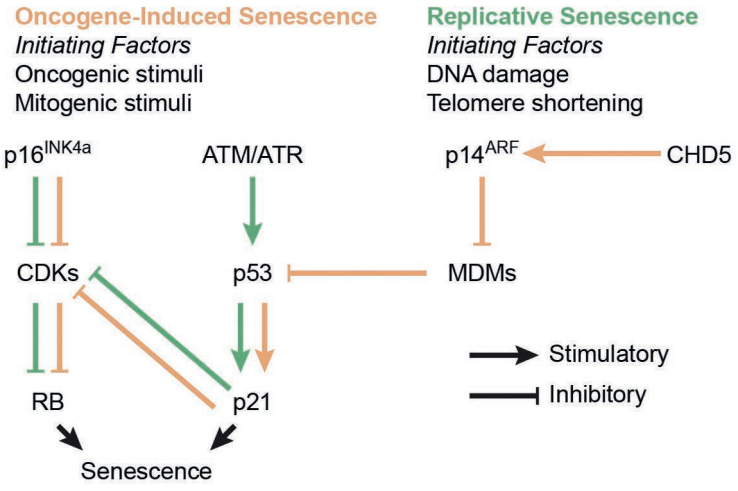
Non-diffuse gliomas

Recent studies have shown that pilocytic astrocytomas near universally harbor abnormalities in the MAPK-pathway, and most commonly a tandem duplication targeting chromosome 7q, which gives rise to a *KIAA1549-BRAF* fusion gene consisting of the N-terminus of *KIAA1549* and the kinase domain of v-RAF murine sarcoma viral oncogene homolog B1 (*BRAF*) [19]. Alternative alterations include the oncogenic V600E missense mutation also targeting *BRAF* [17]. The *BRAF* V600E mutation results in an activating change due to a substitution of valine with glutamic acid at codon 600. In a non-cancer setting, *BRAF* activates kinases MEK and ERK, which in turn activate transcriptional machinery to promote differentiation, proliferation, growth and apoptosis [348]. Both *BRAF* V600E mutation and *BRAF* fusion genes contribute to tumorigenesis by constitutively activating the kinase domain of *BRAF*, resulting in overactive signaling activity and a selective growth advantage for affected cells [349-351]. In most pilocytic astrocytomas (even after thorough analysis) an activating change in *BRAF* or other MAPK-pathway members is the only genomic change that can be confidently detected, implying that it is the glioma initiation event in this disease [352].

Phase II: Oncogene-induced senescence

Continued oncogenic signaling in the initial growth phase prompts the activation of tumor suppressive signaling *via* activation of the $p16^{INK4a}$/$p14^{ARF}$-RB-p53 cell cycle and cell stress pathways (**Figure 7-2**), slowing tumor growth and transitioning a majority of cells with intact pathways into a terminal state called oncogene-induced senescence [353]. First discovered over five decades ago in cultured fibroblasts, senescence is a stress-induced durable cell-cycle arrest [354]. The role of senescence in cancer has been reviewed extensively [355-360]. Briefly, senescence provides a major tumor suppressive barrier and dividing tumor cells are put under selection pressure to acquire molecular events to prevent or overcome its onset. Hallmarks of senescence include durable growth arrest; short, dysfunctional telomeres; and a marked increase in DNA damage and stress signaling [357]. Although senescent cells are growth arrested, they are metabolically active and release a plethora of signaling molecules to the microenvironment,

also known as the senescence-associated secretory phenotype [360]. Distinction must be made between oncogene-induced senescence which is discussed here and is triggered by chronic oncogenic signaling, and replicative senescence discussed later, describing senescence triggered by telomere dysfunction following extensive replicative cycles [358]. Oncogene-induced senescence poses a significant growth barrier, and most cells will not acquire molecular alterations that allow them to bypass this barrier and will therefore become senescent [353]. However, rare cells may acquire such alterations, eliciting a selective sweep by a subclone that will rapidly dominate the neoplastic cell population.



↑ **Figure 7-2. Process diagram indicating the p16INK4a/p14ARF-RB-p53-pathway in normal conditions.**

Disruption of one or multiple components through mutation or copy number change may prevent or suppress the onset of senescence. Various stimuli use different routes to activate the senescence response, leaving compensatory mechanisms in place in case components fail. For example, if oncogene-induced senescence is repressed via *CDKN2A/B* inactivation, DNA damage and telomere shortening could still trigger replicative senescence via *ATM* and *ATR*. CDKs = Cyclin Dependent Kinases (eg. CDK2), MDMs = Murine Double Minutes (eg. MDM2)

In some cancer types a senescent precursor stadium can be identified, such as intestinal polyps in colon cancer and dysplastic naevi in melanoma [361]. In contrast to the catastrophic karyotypes demonstrated by their later stage derivatives, these growth-arrested senescent precursor lesions entered a senescent state *via* the activation of single oncogene such as *BRAF* and are otherwise genetically unremarkable [362-365]. This begs the question if such premalignant

precursor changes exist for diffuse gliomas as well. If so, such precursor lesions may however well be (sub)microscopic in size and go unnoticed in imaging or autopsy studies.

The tumor suppressor proteins $p16^{INK4a}$, $p14^{ARF}$, RB and p53 can be considered as gatekeepers of senescence. The INK4a/ARF locus on chromosome 9 contains both Cyclin Dependent Kinase Inhibitor 2A (*CDKN2A*) and 2B (*CDKN2B*), combined encoding for both $p16^{INK4a}$ and $p14^{ARF}$ *via* alternative splicing. Tumor suppressors RB and p53 on the other hand are encoded for by the genes Retinoblastoma 1 (*RB1*) and Tumor Protein 53 (*TP53*) on chromosomes 13 and 17, respectively. Inactivation of one or multiple of these genes *via* genomic deletion and/or inactivating mutations has been linked to repression of senescence signaling and are a common event in all cancers including gliomas [249, 355, 366, 367]. For the purpose of this review, we define the term 'senescence bypass event' as any molecular alteration that suppresses the onset of oncogene-induced senescence.

IDH-mutant diffuse astrocytomas

IDH-mutant astrocytomas are often characterized by loss of one allele of *TP53*, combined with a loss-of-function mutation in the remaining allele. The frequency of *TP53* mutations in IDH-mutant ('secondary') glioblastomas is comparable to that in lower grade IDH-mutant astrocytomas from which these glioblastomas are derived via malignant progression, suggesting that *TP53* aberrations are early lesions in these tumors [368]. Furthermore, and in contrast to IDH-wildtype gliomas, *TP53* mutations were shared between all *TP53*-mutant cases of primary and recurrent tumors in a recent study [323]. Analysis of multiple biopsies from the same IDH-mutant tumors indicated that samples mutant for *TP53* were always IDH mutant, while some IDH-mutant samples lacked *TP53* mutations, suggesting that IDH mutations precede *TP53* inactivation [49]. IDH mutation and *TP53* inactivation therefore both comprise early events in gliomagenesis, with *TP53* inactivation generally following mutation in IDH.

The p53 tumor suppressor protein is involved in many different functions, and especially its role in cell cycle arrest and senescence is very well understood [369]. While enzymatically active wildtype p53 triggers senescence in response to oncogenic stress, mutant p53 inadequately blocks proliferation thereby bypassing the onset of oncogene-induced senescence. In addition, p53 takes a prominent role in the senescence pathway (**Figure 7-2**) and loss of its enzymatic activity impacts replicative senescence, triggering crisis [249]. This is discussed further in phase IV (paragraph 2.4).

IDH-mutant oligodendrogliomas, 1p/19q-codeleted

The majority of IDH-mutant tumors wildtype for *TP53* demonstrate a combined single copy loss of the complete chromosome arms 1p and 19q (complete 1p/19q-codeletion) [11, 370]. These tumors are the canonical oligodendrogliomas according to the revised WHO criteria [60]. Though it has been suggested that 1p/19q-codeletions are the result of an unbalanced translocation, much about the contribution of this event to oncogenesis remains to be resolved [371]. 1p/19q-codeletion was found to be stable across longitudinal samples and multiple biopsies, suggesting that they are early events [372-374]. The finding that codeleted tumors are almost exclusively IDH-mutant, while the reverse is not true, suggests that mutations in IDH precede codeletion. This suggestion may implicate a common cell of origin for both astrocytoma (IDH-mutant, non-codeleted) and oligodendroglioma (IDH-mutant and 1p/19q-codeleted). Indeed, recent evidence was provided that substantiates this hypothesis, demonstrating that differences in inferred cell identity between oligodendrogliomas and astrocytomas may be entirely explained by different microenvironment makeup and the impact of the 1p/19q-codeletion on the expression of genes on these chromosome arms [132]. The authors found that although these histological tumor types differ in morphology, these histological subtypes share a comparable developmental hierarchy and glial lineage. Transcription factors involved in oligodendrocyte differentiation are also expressed in both histologies [375]. Several case reports have highlighted 'dual genotype' oligoastrocytomas, demonstrating molecular features of both bona fide astrocytic and bona fide oligodendroglial tumor cells [376, 377]. Such a dual genotype may be explained by assuming that very early in gliomagenesis a subset of IDH-mutant cells experiences a complete 1p/19q-codeletion and a *TERT* promoter mutation, while another IDH-mutant subset acquires a mutation in *TP53* and/or *ATRX*. Detailed and sufficiently powered longitudinal studies of primary and recurrent diffuse gliomas may help to elucidate the actual frequency of such 'dual genotypes', how they are related to mixed histological appearances, and how they evolve over time [378].

Additional loss-of-function mutations in far-upstream element binding protein (*FUBP1*) on 1p31.1 and capicua transcriptional repressor (*CIC*) on 19q13.2 are observed in over 60% of 1p/19q-codeleted gliomas [129]. A paired analysis of primary and recurrent 1p/19q-codeleted oligodendrogliomas described distinct alterations in *CIC* and *FUBP1* in the primary and the recurrent tumor [324]. Another report described that these events were frequently private to either primary or recurrence, but not both [374]. Both findings corroborate observations of *CIC* and

FUBP1 mutation heterogeneity across nine distinct samples from the same oligodendroglioma, including finding five distinct *CIC* mutants across nine tumor samples [99]. Loss-of-function mutations in these genes led to a loss of protein expression and the *FUBP1* mutation was associated with adverse survival compared to wild type tumors [379, 380]. These findings suggest an important role for these genes and indicate that in cases in which these events were not found in one sample of the tumor, they still might be present elsewhere. Despite their apparent importance, the role of these events in gliomagenesis remains to be understood.

It is not clear whether 1p/19q-codeleted oligodendrogliomas undergo senescence or acquire mechanisms to bypass senescence. 1p/19q-codeleted gliomas generally lack alterations in genes associated with oncogene-induced senescence such as *BRAF* or senescence bypass such as *TP53*, *RB1* or *CDKN2A*. Nevertheless, given their continued clinical growth it appears that these tumors have somehow evaded growth arrest and senescence barriers. Perhaps the 1p/19q-codeletion allows pre-cancerous cells to avoid senescence through the mono-allelic inactivation of tumor suppressor genes on these chromosome arms [381]. One mechanism that was previously proposed may be *via* a dosage-dependent repression of Chromodomain Helicase DNA Binding Domain 5 (*CHD5*) on 1p36 [382-384]. A study that used genetic engineering to create mouse models with gains and losses of a region corresponding to human 1p36 found that duplication of this region led to decreased proliferation and senescence whereas a single-copy deletion led to immortalization [384].

IDH-wildtype diffuse astrocytomas

Amongst IDH-wildtype astrocytomas/glioblastomas, one of the most frequent alterations is a homozygous loss of *CDKN2A* and *CDKN2B* [385]. Mathematical modelling has suggested that homozygous *CDKN2A/B* loss occurs after +7/-10 but before other molecular events [105]. Homozygous *CDKN2A/B* loss alone is insufficient for tumor formation in mice, requiring the activation of an oncogene to generate tumors *in vivo* [386]. As such, homozygous *CDKN2A/B* loss is likely a second event in the tumorigenesis of IDH-wildtype astrocytoma or glioblastoma. The role of protein products $p16^{INK4a}$ and $p14^{ARF}$ in senescence are very well understood. Indeed, astrocytes with a homozygous deletion of *CDKN2A/B* can grow indefinitely in culture, and introduction of $p16^{INK4a}$ in immortal human glioma cell lines with this deletion leads to cell cycle arrest and senescence [387, 388]. Taken together, these results indicate that loss of *CDKN2A/B may* provide adult IDH-wildtype astrocytomas with a reliable means for senescence bypass.

Mutations in *TP53* sometimes co-occur with homozygous *CDKN2A/B* loss in IDH-wildtype glioma [389]. While *TP53* mutations are often shared across all tumor cells in IDH-mutant astrocytoma, *TP53* mutations in IDH-wildtype astrocytomas are frequently unique to one or a few tumor subclones [73]. In this same study, it was found that amongst IDH-wildtype astrocytomas, whereas *CDKN2A/B* is consistently deleted, *TP53* mutations are frequently lost or gained at tumor recurrence. These observations suggest that, in IDH-wildtype astrocytoma, *CDKN2A/B* may be primarily important for senescence regulation

Pediatric H3-mutant/IDH-wildtype diffuse gliomas are *TP53* mutant in about 50% of cases, which may act as a senescence bypass event in these tumors [283]. Amongst remaining H3-mutant tumors, approximately 20-30% of H3 K27 mutant diffuse gliomas show mutations in Activin A Receptor Type 1 (*ACVR1*) [390-393]. There is no evidence that *ACVR1* has any role in sentence regulation, however *ACVR1* alterations were found to be mostly mutually exclusive with alterations in *TP53* and *PPM1D*. *PPM1D* is a protein phosphatase downstream of p53 involved in apoptosis regulation following DNA damage and thus likely involved in senescence [394]. Expression of *ACVR1* mutants in $Tp53^{null}$ murine astrocytes implanted in mouse brains failed to induce tumors, likely because H3 K27 mutations are required as a tumor initiation event and thus suggesting that *ACVR1* does not meet the criteria for a tumor initiation event [393]. Thus, while about half of pediatric H3-mutant gliomas demonstrate *TP53* mutations that may act to bypass senescence, it remains unclear if and how *TP53*-wildtype H3-mutant pediatric gliomas bypass senescence.

Non-diffuse gliomas

Several lines of evidence suggest that pilocytic astrocytomas (WHO grade I, IDH-wildtype) are arrested in a senescent phase II state and do not advance to later phases. First, these tumors were found to frequently demonstrate several biomarkers of senescence at tumor detection, including widespread b-galactosidase activity and $p16^{INK4a}$ staining [395-397]. Second, these tumors are very quiescent genetically, often demonstrating but a single activated oncogene, such as a *BRAF* fusion, *BRAF* V600E mutation or rarely an activating mutation in *FGFR1* or *PTPN11* [116, 398, 399]. Third, these tumors generally grow slowly, have excellent outcomes and sometimes regress, perhaps because the tumor cells do not immortalize [400-403]. Fourth, expression of activated *BRAF* V600E alone does not lead to tumor development in *in vitro* and in *in vivo* mouse models, while the combined activation of *BRAF* and loss of *CDKN2A/B* is transforming, suggest-

ing that additional mutations to bypass senescence are required to advance to phase III [404-406]. The clinical presentation of pilocytic astrocytoma may be the result of senescence-mitigating circumstances, such as a cell-of-origin with proliferative potential in absence of senescence-bypass.

Phase III: Stressed Growth

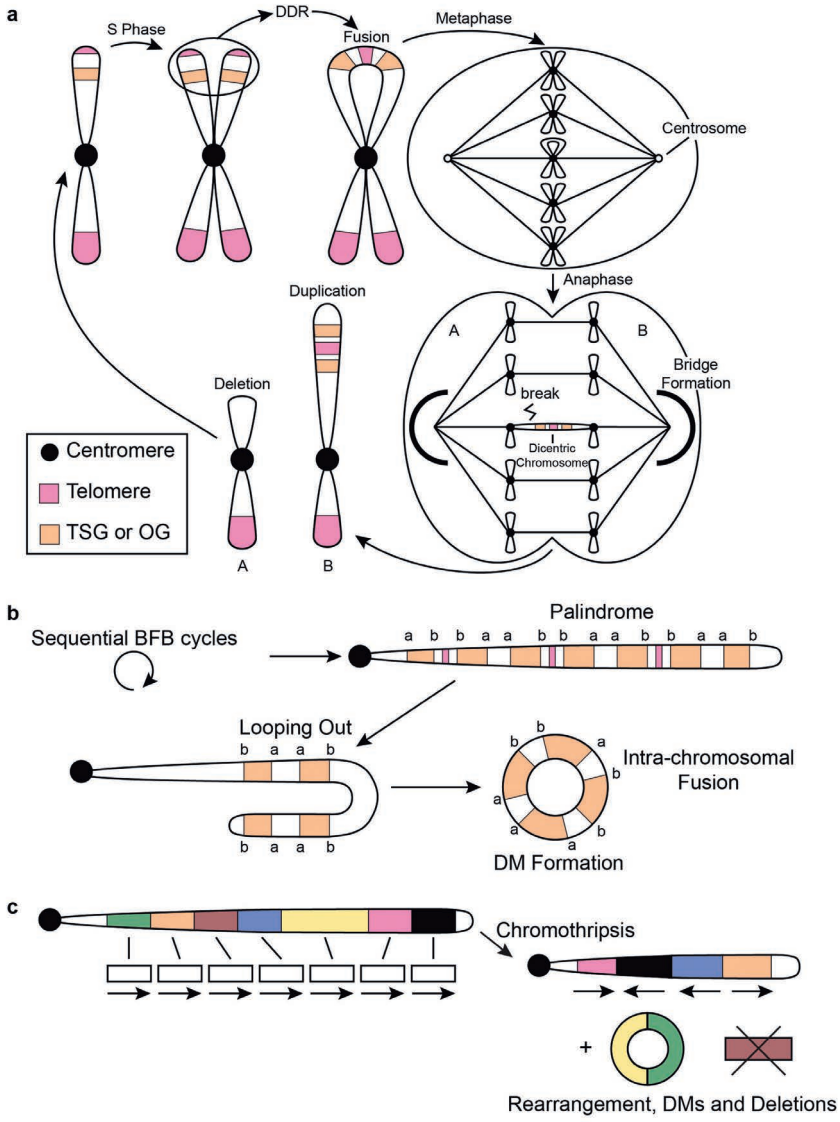
Cells presenting with continued proliferative signaling beyond the oncogene-induced senescence barrier are generally characterized by defective DNA damage response signaling and continued growth in a stressed environment. During this phase the repetitive DNA at the telomeric terminal ends of chromosomes become increasingly important [87]. Telomeres progressively shorten as cells divide due to the linear conformation of chromosomes and directional replication machinery, a phenomenon that is critically important for diseases like cancer which are characterized by often rampant proliferation [88]. Telomeric DNA takes on a lasso conformation called the t-loop, and these loops are bound by the shelterin DNA binding protein complex. Together, these characteristics protect chromosome ends from being recognized as DNA double strand breaks and prevent inadvertent activation of DNA damage response pathways [89].

Dysfunctional telomeres are critically short and improperly protected telomeres lacking t-loops and shelterin complexes. They trigger the activation of DNA damage response pathways via ataxia telangiectasia mutated (*ATM*) and ataxia telangiectasia and Rad3-related (*ATR*) kinase, which are triggered by exposed and unprotected double stranded and single stranded DNA break ends, respectively. The exposed ends then fall victim to homology directed repair (HDR) and non-homologous end joining (NHEJ) repair processes, intended to repair accidental DNA breaks but lead to gross genomic instability when triggered by dysfunctional telomeres. When telomeres are unprotected, these repair processes prompt sister chromatids to fuse with one another, forming a dicentric chromosome. During the anaphase, the dicentric chromosome will form a bridge spanning the mitotic spindle and connecting the two daughter cells. Resolution of the chromatin bridge via cytoplasmic 3' nuclease *TREX1* results in breakage of the dicentric chromosome at a locus not necessarily at the site where the fusion had occurred, resulting in an unbalanced inheritance of genetic material between the two daughter cells. Because the resulting daughter cells also lack telomeres, this process of breakage-fusion-bridge (BFB) cycles (**Figure 7-3a**) will repeat itself every subsequent cell division until telomeres are restored [251] [407-410]. The detrimental genomic instability acquired via telomere dysfunction and BFB-cy-

cles endows these cells with powerful stochastic mutational mechanisms to acquire changes that provide a survival benefit under selective pressure.

The intensity of genomic instability endured during the stressed growth phase may depend on the severity of senescence pathway dysregulation incurred overcoming oncogene-induced senescence (**Figure 7-2**). In the case of H3-mutant, IDH-wildtype and IDH-mutant astrocytoma/glioblastoma this pathway is perturbed close to the source via the direct loss of RB, p53 or $p16^{INK4a}$ protein function. In IDH-mutant and 1p/19q-codeleted oligodendroglioma this pathway may be repressed indirectly, for example *via* the modulation of $p14^{ARF}$ activity through a partial deletion of *CHD5*. This may explain why the latter group of tumors show significantly less genomic instability compared to the former tumor types. Moreover, in pilocytic astrocytoma this pathway may not be affected at all and these tumors may not advance beyond oncogene-induced senescence. While genomic instability incurred during phase III demonstrate some patterns that are unique to a certain glioma subtype, these features are generally shared across all gliomas regardless of subtype. We therefore did not separate this section according to tumor type as we did for the other phases. More research is needed to carefully delineate the selective pressures at play to better understand differences and similarities between various glioma entities in this respect.

The advent of high-throughput sequencing has led to remarkable progress in understanding the complexity of genomic instability in cancer, including complex deletions, amplifications and translocations [322]. It is important to note that the genomic organization that can be reconstructed using sequencing at the time of analysis are those changes that resulted in viable cells and were selected for. Recent work has shown that telomere dysfunction directly leads to catastrophic genomic events, including genome shattering (chromothripsis), clustered regions of focal hypermutation (kataegis) and whole genome doubling (tetraploidization) [411-413]. Although the incidence of chromotripsis across the spectrum of gliomas is not known, a recent report suggests that chromotripsis is very common in glioblastoma [414]. Comparison of primary and recurrent tumors across various tumor types including gliomas demonstrated that recurrent tumors lack additional genomic instability, suggesting that these events occurred during a stressful period the initial development of the tumor that was later stabilized [412].



↑ **Figure 7-3. Genomic instability related to telomere stress.**

a. Schematic illustrating BFB-cycles. Following a single BFB-cycle, daughter cells are left with unequal DNA content, leading to a deletion in A and an amplification in B. BFB cycles may also involve fusion of non-sister telomeres (not shown). **b.** Repetitive BFB-cycles form palindromes demonstrating high intra-segmental homology. This can lead to intra-chromosomal fusions and formation of double minutes. **c.** Following chromothripsis segments can be rearranged, lost or circularized. TSG = tumor suppressor gene; OG = oncogene; DDR = DNA damage response; DM = double minutes; BFB = breakage-fusion-bridge

Recently there has been a renewed interest in circular extrachromosomal DNA elements called double-minute (DM) chromosomes in cancer, and it was shown that such DMs are common in gliomas [415, 416]. Although DM chromosomes have been long recognized as a cytogenetic feature of cancer, relatively little was known about its biological relevance. DM chromosomes have a predisposition to involve cancer oncogenes such as MYC Proto-Oncogene Protein (*MYC*), MDM2 Proto-Oncogene (*MDM2*) or Cyclin Dependent Kinase 4 (*CDK4*) [130, 417-420]. A unique feature of DM is that they lack centromeres and telomeres to dictate the organization of the mitotic spindle during mitosis, and are therefore randomly distributed across daughter cells [421]. Interestingly, this feature hypothetically provides DM chromosomes with an impressive fitness advantage over linear chromosomes as they do not need telomeres to protect them from inadvertent DNA damage response pathways and are not subjected to detrimental BFB-cycles. It has been proposed that DM chromosomes are the result of the fusion and circular assembly of stretches of linear DNA consisting of highly homologous sequences of inverted duplications in tandem following sequential BFB-cycles (**Figure 7-3b**) [422]. Others have proposed that ineffective DNA repair following chromotripsis can lead to linear DNA fragments getting pieced together in a circular fashion (**Figure 7-3c**) [423]. A study by our group found evidence of chromotripsis in several glioblastoma samples localized to chromosome 12 involving *MDM2* and *CDK4* and suggested that these segments may be arranged in extrachromosomal DMs [130]. Because these are inherently random processes, it is possible that DM-chromosomes that promote survival are positively selected for during telomere dysfunction. Compared to IDH-mutant gliomas, DMs in IDH-wild-type tumors more often involve established glioma oncogenes, despite what appears to be a comparable frequency of DMs in both glioma categories [415]. More research is needed to precisely determine the frequency of DMs in glioma subtypes and to pinpoint the genetic origin of these structures.

A recent study of paired primary and recurrent IDH-mutant gliomas reported that in some cases allelic imbalances of the IDH-mutant allele occurred upon recurrence, which led to a change in mutant protein expression and consequently decreased 2HG production [424]. Furthermore, IDH-mutant tumors are very hard to culture, and when it succeeds, IDH mutations that were present initially have been reported missing, raising the possibility that losing an IDH mutation is advantageous for survival in culture [425]. Introduction of mutant IDH in cell cycle checkpoint deficient cells rapidly transforms these cells into competent tumor cells [426]. However, IDH inhibition in these cells after as little as four days

after its first introduction did little to slow tumor growth. These findings suggest that some IDH-mutant gliomas may rapidly evolve and acquire additional driver events to uphold the tumor cell population.

To summarize, telomere dysfunction and stressed growth may promote the context-dependent evolution of glioma cells, rendering glioma initiation events redundant and providing gliomas with new fuel that rapidly increase intratumoral heterogeneity and can deal with various toxic stresses and bottlenecks. While stochastic mutational mechanisms in the stressed growth phase provide ample selection pressure to acquire beneficial changes, the detrimental genomic instability under which cells must operate acts as a powerful tumor suppressive barrier. Unchecked growth will rapidly lead to another round of DNA-damage induced replicative senescence, or when those checkpoints fail completely, cell crisis.

Phase IV: Replicative senescence/crisis

Sustained stressed growth is not durable and will eventually lead tumor cells down to one of two possible roads. Tumor cells with a partially intact senescence response (i.e. functional p53 and RB) may undergo a second round of senescence called replicative senescence in response to dysfunctional telomeres. Tumor cells with a completely dysfunctional senescence response (i.e. loss-of-function mutation in *TP53* or *RB1*) instead continue proliferating in a state of cellular crisis leading to cell death in a vast majority of cells [427]. Replicative senescence and crisis both pose a second population bottleneck to further tumor formation. It is essential that tumor cells transition to a less stressful environment with proper telomere maintenance to prevent further BFB-cycles and other catastrophic events. Cell culture experiments have demonstrated that direct immortalization of cells prior to a stressed growth phase enables them to bypass genomic instability and immortalize lacking the wild karyotypes typically associated with malignant transformation [254, 428]. These observations suggest that genomic instability in cancer development generally precedes immortal growth and is required to generate errors enabling telomere maintenance [428].

Acquisition of a telomere maintenance mechanism endows cancer cells with immortal growth, meaning that they are bestowed a limitless replicative potential [429]. Telomere maintenance is established once a tumor cell has reactivated telomerase or activated alternative lengthening of telomeres. Moreover, restoration of telomere function may prevent further BFB-cycles and restore genome stability. The canonical pathway involves the reactivation of the ribonucleoprotein telomerase which is transcriptionally silent in differentiated adult cells [253].

The telomerase catalytic component telomerase reverse transcriptase (*TERT*) is expressed in over 80% of human cancers and is thought to be rate limiting for telomerase activity [255]. In the alternative pathway tumors become immortalized via a recombination driven mechanism called alternative lengthening of telomeres (ALT) [259].

IDH-mutant diffuse astrocytomas

IDH-mutant diffuse astrocytomas almost universally demonstrate ALT [285]. ALT cells present with several defining characteristics, including a heterogeneous distribution of telomere length across chromosomes, extrachromosomal telomeric DNA fragments in a circular configuration (c-circles), increased expression of telomeric repeat-containing RNA (*TERRA*) from telomeres, the formation of ALT-associated promyelocytic leukaemia bodies (APBs), frequent telomere sister chromatid exchanges (T-SCEs) and recombination between telomeres from different chromosomes [430]. Telomeres in ALT cells are heterogeneous in length and relatively long, demonstrating telomere lengths much longer than telomerase positive cells on average [431]. ALT provides cancer cells with stabilizing telomere maintenance in a telomerase negative setting. Although it remains unclear how ALT becomes activated, its presence has been tightly associated to loss-of-function events targeting the a-thalassemia/mental retardation syndrome X-linked (*ATRX*) or death-domain associated protein (*DAXX*) genes and these events are also a hallmark feature of IDH-mutant astrocytomas [52, 262, 432]. *ATRX* functions as an ATP-dependent helicase within the SWI/SNF family and combined these two genes form the ATRX-DAXX complex, which functions as a histone chaperone to deposit the histone variant H3.3 at telomeres [433]. Telomeric DNA has a tendency to form secondary quadruplex structures that challenge the replication machinery and need to be resolved for proper replication [434]. How exactly these two genes protect telomeres from recombination and ALT is still unknown. It has been suggested that the combined helicase activity of *ATRX* and the histone chaperone capabilities of the ATRX-DAXX complex can resolve the secondary quadruplex structure at telomeres, thereby enabling proper progression of the replication fork during S-phase and preventing the inadvertent activation of recombination (ALT) mechanisms [435, 436].

The presence or absence of inactivating *ATRX* and *DAXX* mutations present a strong correlation with ALT in many tumor types including gliomas [262, 283, 437]. Recent *in vitro* studies have shown that knockout of *ATRX* alone is insufficient to cause ALT, however, *ATRX* knockout combined with inactivation of p53 and RB

enzymatic activity led to an increased incidence of ALT after enduring several cycles of telomere induced crisis [263, 298, 438]. Furthermore, the reintroduction of *ATRX* expression in *ATRX* mutant ALT cells led to a repression of T-SCE, APBs and c-circle formation [298, 438].

IDH-mutant oligodendrogliomas, 1p/19q-codeleted

Oligodendrogliomas, IDH-mutant and 1p/19q-codeleted almost universally use telomerase to maintain telomeres and virtually all of these tumors carry an activating *TERT* promoter mutation [103]. These *TERT* promoter mutations are amongst the most common non-coding mutations in cancer [103, 104, 256, 267]. Recurrent hotspot point mutations substitute a cytosine at -228 or -250 relative to the promoter to a thymine (C228>T or C250>T) to create a de novo e-twentysix (*ETS*) transcription factor binding site that recruits the ETS family member GA-binding protein alpha chain (*GABPA*) to activate transcription [439]. Although the timing of *TERT* promoter mutations is still under debate, several lines of evidence suggest that *TERT* promoter mutations arise early in gliomagenesis and perhaps even occur prior to the glioma initiation event. *TERT* promoter mutations preferentially occur in tissues with a lower rate of self-renewal and there are numerous reports on the extra-telomeric functions of *TERT*, including effects on the NF-κB and WNT/β-catenin pathway promoting tumor growth and invasiveness [103, 440, 441]. Combined this raises the possibility that these mutations may contribute to tumorigenesis via other pathways than its effect on telomerase alone, providing a biological reason for these mutations to contribute to gliomagenesis early in phase I and before the onset of dysfunctional telomeres. In a glioma-specific analysis it was found that nearly all tumors with the 'phase I event' +7/-10 or 'phase II event' 1p/19q-codeletion have *TERT* promoter mutations, whereas not all *TERT* promoter mutant gliomas have +7/-10 or 1p/19q-codeletions, which may indicate that *TERT* promoter mutations even precede +7/-10 and 1p/19q-codeletions [15]. Another group studied the mutation fraction using multi-sector sequencing in 1p/19q-codeleted oligodendrogliomas and found that *TERT* promoter mutations indeed are early events and may occur before IDH mutations [99]. The idea that *TERT* promoter mutation occurs early is further corroborated by the finding that genetically engineered *TERT* promoter mutations in telomerase positive embryonic stem cells do not affect telomerase activity, while upon differentiation these engineered cells remain telomerase positive and acquire immortality [442]. More recently, it was found that *TERT* promoter mutations in melanoma initially do not support telomere maintenance

and telomeres shorten to critically short length despite harboring promoter mutations [443]. The effect on telomere length was not observed until later, which the authors attributed to a two-step immortalization process. One study even reported canonical (-228 and -250) somatic *TERT* promoter mutations in the blood of multiple non-cancer patients, indicating that these events could even occur before the onset of cancer and act to prime the tumor bed [444]. Although roles for *TERT* outside of telomere maintenance remain to be understood, these observations provide a sound argument that *TERT* promoter mutations can occur early in or even before gliomagenesis while providing a means towards immortalization at a later stage.

IDH-wildtype diffuse astrocytomas

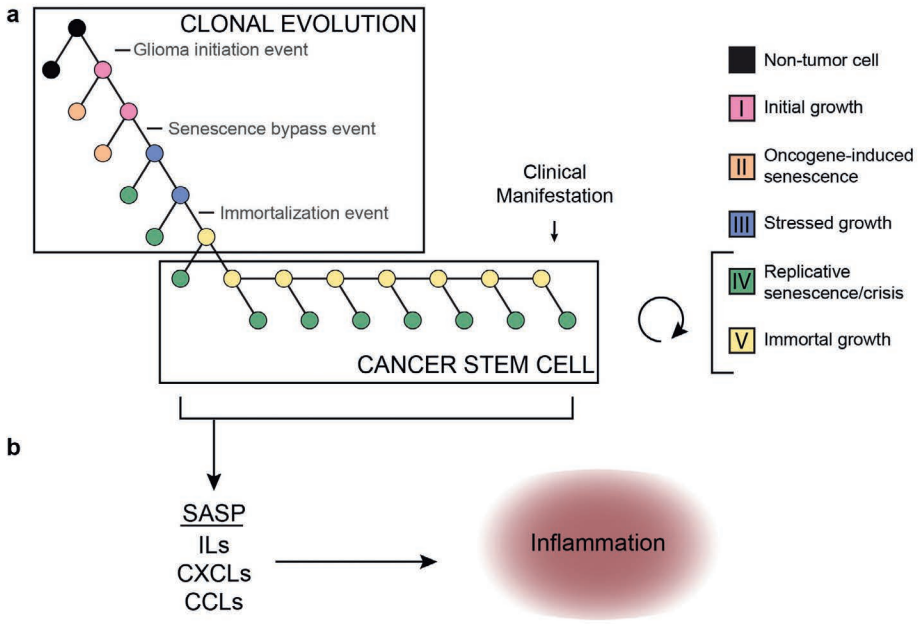
The majority of IDH-wildtype diffuse gliomas use telomerase for telomere maintenance [445]. Re-analysis of previously published samples reclassified according to WHO 2016 criteria demonstrated that approximately 75% of diffuse IDH-wildtype gliomas are *TERT* promoter mutant [446]. Thus, *TERT* promoter mutations are common in both the most and the least aggressive diffuse gliomas (IDH-wildtype diffuse astrocytomas and IDH-mutant and 1p19q-codeleted oligodendrogliomas, respectively), suggesting that *TERT* promoter mutations are not dictating their biological behavior. It was further found that *ATRX* mutations occur in approximately 5% of IDH-wildtype diffuse astrocytomas [446]. The prevalence of ALT in IDH-wildtype diffuse gliomas is higher than the frequency of *ATRX* mutations, suggesting that some of these tumors may use ATRX-independent mechanisms to activate ALT [285]. In a similar fashion, the prevalence of telomerase activity is higher than the prevalence of *TERT* promoter mutations in this tumor type, suggesting that these tumors may use *TERT* promoter independent mechanisms for the reactivation of telomerase [445]. Several candidate mechanisms have been previously described in glioma, including *TERT* promoter methylation or *TERT* amplifications [447]. Contrary to adult IDH-wildtype diffuse glioma, H3-mutant malignant pediatric glioma frequently demonstrate ALT [448], and several studies reported frequent co-occurrence of *ATRX* mutations in both H3 K27 and G34 mutant gliomas. However, reports of co-occurrence vary between 30% and 60% for K27 and 75% to 100% for G34, suggesting that there is a role for telomerase in many of these tumors as well [449].

Phase V: Immortalization & Dedifferentiation

The glioma stem cell theory states that amongst all cancerous cells in a tumor, a subset of cells act as progenitor or stem cells with reproductive capabilities and sustaining the cancer, much like normal bone marrow stem cells are responsible for replenishing the population of circulating leukocytes [450]. It has often been contrasted to the theory of clonal evolution, which suggests that cancers evolve through an iterative process of clonal expansion of from a single cell [451]. Recent advances in single-cell sequencing and lineage tracing have unveiled multiple populations of tumor cells in bulk tumor samples, providing fuel for the cancer stem cell hypothesis [132, 133, 452, 453]. One study used single cell RNA sequencing on IDH-mutant and 1p/19q-codeleted oligodendroglioma patient samples and uncovered distinct cell populations of undifferentiated tumor stem cells and cells that are more differentiated along various glial lineages [452]. In a similar study of IDH-mutant astrocytoma the authors were able to detect the same cellular populations but with a higher ratio of stem-like to differentiated cells that increased with increasing WHO grades [132]. Another study used DNA barcoding of repeatedly *in vivo* transplanted glioblastoma cells to trace the lineage during their engraftment and found a population of progenitor cells that sustained the tumor and gave rise to differentiated non-proliferative cells [453].

These studies all provide support for a cancer stem cell hypothesis and raise the question how these findings fit with our model, which leans towards a model of clonal evolution. In fact, current evidence may suggest that both mechanisms are acting together (**Figure 7-4a**). Whereas clonal evolution is important to establish the initial cancer stem cell population, neutral evolution (in line with the cancer-stem cell hypothesis) may fit better once the initial cancer core has been established, especially so when further evolutionary stimuli (e.g. senescence barriers, hypoxia, treatment) are lacking. The concept of neutral evolution holds that most molecular changes are not caused by natural selection but rather by the stochastic allelic variation that are neutral and do not affect cellular fitness [454]. A recent study analyzed cancer genomes from TCGA and found neutral evolution in approximately one-third of a wide spectrum of over 900 tumors, including 35 gliomas of which 8 (23%) suggested evidence in support of a neutral evolution process [151]. The authors conclude that all clonal selection must have occurred before the onset of cancer growth and not in later arising subclones. Several groups have since challenged these findings and suggested that their analysis does not univocally prove neutral evolution starting from the first malignant cell [455-457]. While it may be possible that tumors evolve neutrally beyond the most

recent selective sweep we do not agree with the conclusion that these findings suggest that all clonal selection must occur before the onset of tumor growth. The authors do not take into account that at the time the tumor presents itself and is surgically removed, all remnants of a selection process that have been outcompeted or died will have completely disappeared, in contrast to "neutral" variants which do not affect fitness and will remain. According to our simplified model, tumorigenesis sequentially follows phase I to V. Once the cancer stem cell population has been established, tumor cells will follow neutral evolution, as long as new evolutionary or other stimuli for opportunistic growth are lacking (**Figure 7-4a**).



↑ **Figure 7-4. Integration of a clonal evolution and cancer stem cell model for gliomagenesis.**

a. This model assumes that sequential mutations and selection pressure drive the evolution of cancer stem-like cells. At the same time these stem-like cells may give rise to more differentiated (ie. phase IV) offspring that may divide further but rapidly become growth arrested. **b.** According to this model these cells may be senescent and contribute to the cancer phenotype by eliciting a micro-environment response via SASP.

SASP = senescence-associated secretory phenotype; ILs = interleukines; CXCLs = chemokines (C-X-C motif); CCLs = chemokines (C-C motif)

Aforementioned studies supporting the cancer stem cell hypothesis suggest that multiple cellular populations exist within a tumor, including a self-renewing cancer stem cell population and a less-proliferative differentiated population. A recent paper demonstrated that cells derived from glioma stem cells may differentiate and subsequently undergo senescence [458]. We speculate that phase V tumor cells represent cancer stem cells that may give rise to more differentiated, phase IV cells (**Figure 7-4a**). The transition from phase IV to V and vice versa is likely volatile in nature owing to transcriptional reprogramming including the activation of the stemness factors Oligodendrocyte Transcription Factor 2 (*OLIG2*), Sex Determining Region Y – Box 2 (*SOX2)* and the reactivation of telomerase [112]. On the other hand, the transition from phase I to phase IV is more rigid in nature, involving various changes on a genomic level, including somatic mutations and copy number changes as described. This important distinction in flexibility led us to believe that phase V cells may re-differentiate and assume a phase IV state.

Although they are growth arrested, senescent cells are metabolically active and release a plethora of signaling molecules to the microenvironment. The senescence associated secretory phenotype (SASP) is a feature of senescent cells that curtails these cells with the release of proinflammatory cytokines [459]. SASP does not depend on $p16^{INK4a}$ or p21 activity and senescence with intact $p16^{INKa}$ function actually suppresses SASP [460]. Similarly, activated p53 signaling also suppresses SASP while *TP53* loss induces SASP [461-463]. These findings suggest that the SASP response is stronger when senescence pathway genes are lost. Indeed, IDH-wildtype astrocytomas often harbor homozygous deletions in *CDKN2A/B* and are known to have a highly active microenvironment [74]. Genes associated with SASP were shown to be overexpressed in higher grades of glioma and older patients, the latter group more likely to be affected by high-grade IDH-wildtype astrocytoma [464]. Moreover, it was found that primary glioblastoma cells retain a functional senescence program despite mutations in the *TERT* promoter and *CDKN2A/B* locus [465]. These findings imply that senescent and differentiated phase IV cells may be crucial for shaping the immune microenvironment in gliomas (**Figure 7-4b**).

A Broader Perspective

While there are not many known environmental risk factors that predispose to glioma, large cohort genome wide association studies over the past two decades have identified multiple heritable polymorphisms conferring glioma risk

[320, 466-470]. Notably, several of these risk loci are localized to genes involved in telomere maintenance, including the telomerase reverse transcriptase *TERT*, telomerase RNA component *TERC*, and other telomere maintenance associated genes STN1, CST Complex Subunit (*OBFC1*), Protection Of Telomeres 1 (*POT1*) and Regulator of Telomere Elongation 1 (*RTEL1*) [466]. Moreover, there appears to be a significant increased glioma risk in people with increased leukocyte telomere lengths [471]. Unsurprisingly, glioma risk alleles at aforementioned genes are also associated with increased leukocyte telomere length [472-474]. Telomeres thus play an important role in not only the development of gliomas, but also in glioma risk [475]. In fact, the positive association between leukocyte telomere length and cancer risk is not specific to glioma and shared across many cancers. A recent Mendelian randomization study found that longer leukocyte telomere length was associated with an increased risk to cancer but a reduced risk to non-neoplastic disease such as cardiovascular disease [476]. It has been suggested that this difference is due to individuals with longer telomeres being more likely to acquire driver mutations due to an increased proliferative potential whereas the inverse relationship observed for non-neoplastic disease may be due to the impact of telomere shortening on tissue degeneration [477, 478].

Several hereditary disorders are associated with an increased risk of glioma development, including neurofibromatosis type 1 and type 2 (NF1, NF2) and the *TP53* germline mutation/Li-Fraumeni syndrome. NF1 and NF2 are autosomal dominant hereditary disorders with germline mutations in *NF1* and *NF2* and clinically characterized by multiple benign nerve sheath tumors (especially neurofibromas in NF1, schwannomas in NF2), but also by a markedly increased risk on particular gliomas (especially pilocytic astrocytoma in NF1 and ependymomas in NF2) *[479-481]*. Both genes are well-known tumor suppressor genes and key components in the MAPK-pathway [482]. It has been demonstrated that senescence commonly occurs in benign nerve sheath tumors and that prolonged *NF1* disruption leads to oncogene-induced senescence in a model system, providing a rationale as to why these germline disorders present with tumors that are often relatively indolent [483]. A germline perturbation affecting *NF1* or *NF2* can be considered a tumor initiation event, explaining why this germline disorder guarantees the formation of multiple benign nerve sheath tumors over one's lifetime.

Li-Fraumeni syndrome is a rare autosomal dominant hereditary disorder that is caused by the germline perturbation of *TP53* or *CHK2*, which regulates p53 activity [484-486]. Whereas NF1 and NF2 guarantee the formation of especially multiple benign tumors (including non-diffuse gliomas) in a lifetime, Li-Fraumeni

patients pose a greater risk to developing a malignant tumor, including a diffuse glioma [487]. This risk increases with age and is over 50% at age 30, with a lifetime cancer risk of over 70% in men and almost 100% in women [485]. Moreover, 15% and 4% of affected individuals were found to develop a second and third cancer [488]. Li-Fraumeni syndrome germline mutations affect phase II and prevent onset of oncogene-induced senescence following the acquisition of a glioma initiation event, thus increasing the risk of developing cancer over a lifetime.

The germline mutations underlying NF1/NF2 and Li-Fraumeni syndrome represent pathways that both need to be disrupted for a malignant tumor to form. The fact that nearly all patients with NF1/NF2 develop one or more benign tumors can be understood by acknowledging that in these disorders a germline tumor initiation event is involved. Unless this pathway is supplemented by a senescence bypass event these tumors do not readily proceed to malignancy. In contrast, Li-Fraumeni syndrome is characterized by an increased risk for malignant tumors in many (but not all) patients. In this syndrome, the germline senescence bypass event needs to be supplemented by a tumor initiation event for a tumor to develop, and such tumors may be more aggressive/malignant due to the defective senescence barrier, allowing the tumor (precursor) cell to instantly progress to phase III and instigate genomic instability.

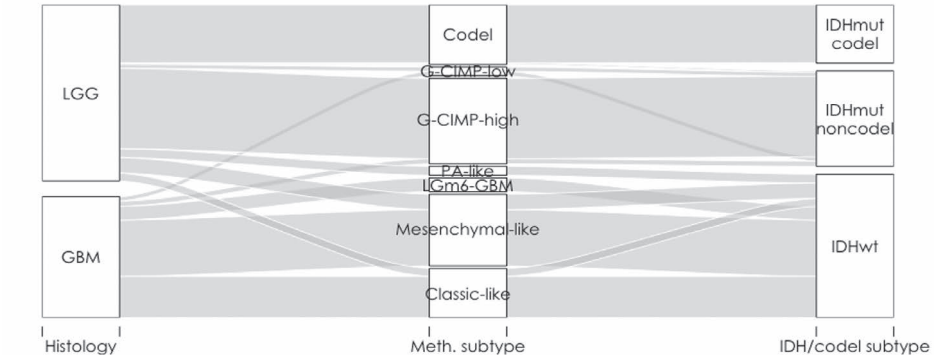
Conclusion

Our knowledge on the molecular events driving cancer has grown exponentially over the years. This review has aimed to put this new knowledge into the perspective of the temporal molecular pathogenesis of glioma, starting from the first aberrant cell all the way to a symptom-causing glioma. To this end we have combined what is known on gene (mal)function, tumor evolution, genomic instability and telomere maintenance to develop a model of gliomagenesis. This model describes five sequential phases cancer cells undergo during their gliomagenesis. We speculate that transitions from one phase to the next can be characterized by acquisition of tumor driving events that sequentially contribute to the hallmarks of cancer as previously proposed, including proliferation, evasion of apoptosis and limitless replicative potential [318, 319]. Our model is a simplified abstraction of what may be the truth and new insights will refine and improve our understanding. Meanwhile, we hope that our model will help foster hypotheses leading to new insights on the molecular life history of glioma, that will help identify convincing therapeutic vulnerabilities.

CHAPTER 8

Summary and Future Perspectives

This final chapter provides a brief summary of the preceding chapters, some of my thoughts on molecular neuropathology in clinical practice and perspectives on translating these new insights in tumor biology into new avenues for precision medicine. In **Chapter 1** I discussed the traditional histopathological distinction between lower grade glioma (grade II-III) and glioblastoma (grade IV) based on a tumors' histology including the presence or absence of various grading criteria.



↑ **Figure 8-1. Relationship between histology and molecular diagnosis.**

Sankey plot indicating the relationship between histology (left), DNA methylation subtype (middle) and IDH/codel subtype (right). LGG = lower grade glioma, GBM = glioblastoma.

In **Chapter 2** I presented a comprehensive molecular analysis of primary diffuse gliomas across histological types. This work revealed that molecular subtypes are not always faithful to a specific histological type (**Figure 8-1**). Using methylation analysis, we discovered two previously undescribed tumor subtypes. One subtype specific to IDH mutant tumors that we called "G-CIMP-low" harbored low levels of DNA methylation and was associated with poor outcomes. Another subtype we coined "PA-like" that was specific to IDH wild-type tumors demonstrated few baseline abnormalities in DNA methylation and frequently showed driver mutations rarely seen in diffuse gliomas such as BRAF V600E. These patients also had a remarkably favorable outcome when compared to other IDH wild-type tumors. Overall, molecularly defined subtypes provided a much more refined and accurate prognostication of patients when compared to a purely histological classifier. This research, combined with the work of many others in the field promoting a molecular take on neuropathology, influenced recent updates to the WHO diagnostic standards of CNS tumors, which are increasingly dependent on molecular testing for making a conclusive diagnosis.

Unfortunately, despite our best efforts to treat diffuse glioma patients with radio- and chemotherapy, tumor recurrence is inescapable, and all patients eventually succumb to the disease. In **Chapter 3**, I set out to use genomics to determine why these tumors are so refractory to the prevailing treatment regiments. To this end we collected a set of over two hundred temporally separated initial and recurrent (posttreatment) diffuse glioma samples spanning the full range of disease subtypes. Recurrent tumors demonstrated higher mutation rates compared to initial tumors and in some cases showed hypermutation after systemic treatment with alkylating agents. Recurrent IDH mutant tumors were enriched in *de novo* Cyclin Dependent Kinase Inhibitor 2A/B (*CDKN2A/B*) deletions, a finding that was further associated with impaired survival and increased aneuploidy. Contrary to our initial expectations, we observed a striking overall similarity of recurrent tumors when compared to their primary counterparts. Analyses into selection and tumor evolution indicated little to no selection secondary to treatment. Rather than an orderly pattern of resistance-specific changes we found that drivers from the initial disease were mostly retained, indicating that the same features driving tumor growth in the initial tumor were likely driving its growth at recurrence. In some cases, we could observe common recurrence-associated changes, including the aforementioned deletion of the tumor suppressor *CDKN2A/B*, suggesting that the acquisition of this feature could play a role in tumor progression. Taken together these results argued that the strongest selective pressures occur early in gliomagenesis and that the effects of current therapies are largely stochastic. Therefore, while treatment effectively slows tumor growth and slightly prolongs survival, the inevitable tumor recurrence shows that more effective treatments are needed to provide a sufficient evolutionary bottleneck to tumor cells.

A key question that had remained unanswered, is whether the DNA methylation subtypes proposed in **Chapter 2** were subject to the same extent of spatial heterogeneity and uncertainty that is seen with transcriptomic subtypes. **Chapter 4** set out to answer this question and results from this analysis indicate that DNA methylation subtypes are relatively robust and show little heterogeneity. This result is important and has clinical implications because it suggests that such an unbiased molecular analysis may not be subject to the intra- and interobserver disagreement that is quite common in a purely histological diagnosis.

One of the more surprising results to come from the work presented in **Chapter 3** was the lack of a clear association between systemic therapy and selection, suggesting that existing treatments do little to drive tumor evolution in a particular direction. Because previous studies had found a distinct genomic

signature associated with alkylating agent treatment secondary to thymine mispairing with alkylated guanine, we wanted to assess if radiotherapy treatment could be associated with a similarly distinct genomic signature. In **Chapter 5** I discuss our work using pre- and posttreatment data from diffuse glioma patients to study radiotherapy treatment failure. We found that radiotherapy treatment was associated with the acquisition of small deletions 5-15 bp in size without evidence for microhomology at breakpoint junctions. These characteristics were reminiscent of DNA repair by the classical pathway of non-homologous end joining, suggesting it is the predominant pathway for the repair of radiation-induced DNA damage. These deletions were randomly spread throughout the genome and were not biased to fragile regions of the genome, suggesting that radiotherapy pervasively invades tissue. Large deletions >100 bp were similarly affected and we surprisingly found convergence at the *CDKN2A/B* locus in post-treatment IDH mutant tumors, raising the possibility that radiotherapy-induced DNA damage could be responsible for the acquisition of this feature of tumor progression. Together these results suggest that radiotherapy may be driving tumor evolution in ways beyond what we had previously recognized and reported in **Chapter 3.**

Results from **Chapter 2** suggested that distinct glioma subtypes might harbor unique mechanisms to maintain their telomeres, the ends of chromosomes described in more detail in **Chapter 1**. Following these findings, I pursued an interest in mapping telomere maintenance mechanisms across different cancer types beyond the brain in **Chapter 6**. There we found frequent overexpression of the telomerase component gene telomerase reverse transcriptase (*TERT*), suggestive of telomere maintenance utilizing the reverse transcriptase telomerase for telomere extension. We proposed several mechanisms for its transcriptional reactivation including rearrangements involving enhancer elements and methylation of a potential repressor element upstream of the *TERT* promoter. Some tumors lacked *TERT* expression but demonstrated genomic alterations affecting a-thalassemia/mental retardation syndrome X-linked (*ATRX*) or Death Domain Associated Protein (*DAXX*) and showed significantly longer telomeres in tumors compared to matched control samples, suggesting telomerase-independent telomere maintenance via alternative lengthening of telomeres. Finally, a number of relatively indolent tumors in this large cohort did not show evidence of any telomere maintenance mechanism, suggesting that the clinical detection of these lesions could precede telomere maintenance (**Table 8-1**). Collectively, my analyses demonstrated the universal nature of telomere maintenance in cancer and emphasized the role of this pathway as a promising therapeutic target.

Table 8-1. Molecular subgroup and suspected telomere maintenance mechanism (TMM).

Molecular subgroup	TMM	Cancer type enrichment
TERT expressing (73%) Tumor TL > control TL	Likely telomerase-driven	Colorectal adenocarcinoma Esophageal adenocarcinoma Gastric adenocarcinoma Lung adenocarcinoma Pancreatic adenocarcinoma Prostate adenocarcinoma Cervical squamous cell carcinoma Esophageal squamous cell carcinoma Lung squamous cell carcinoma Oral squamous cell carcinoma Hepatocellular carcinoma Serous ovarian carcinoma Urothelial cell carcinoma Melanoma
ATRX/DAXX altered (5%) Tumor TL < control TL	Likely ALT-driven	Diffuse glioma* Sarcoma**
TERT non-expressing, ATRX/DAXX wild-type (22%) Tumor TL = control TL	Possibly ALT-driven Possibly telomerase-driven Possibly lacking TMM Possibly novel TMM	Pheochromocytoma Kidney papillary carcinoma Kidney chromophobe carcinoma Thyroid papillary carcinoma Thyroid follicular carcinoma

Table indicating the relationship between molecular subtype and suspected TMM. The total proportion of the entire cohort (n=6,835) for each molecular subgroup is indicated in parenthesis. Cancer types enriched in a specific molecular subgroup are indicated in the rightmost column. * IDH mutant astrocytoma is likely ALT-driven, whereas IDH mutant oligodendroglioma and IDH wild-type diffuse glioma tend to be telomerase-driven. ** Similar subtype specificities are noted for other tumor types, including sarcoma, but such level of detail goes beyond the scope of this table. ALT = alternative lengthening of telomeres, TL = telomere length.

Combining the results from all the main chapters, we began to realize that diffuse gliomas across the major subtypes can be grossly characterized by a combination of hits directly or indirectly influencing the major pathways of proliferation, cell cycle and immortalization (**Table 8-2**). We condensed this observation into

a model for the life history of glioma, that I describe in detail in **Chapter 7**. In summary, the collection of hits across all three pathways provides the molecular ingredients for gliomagenesis.

Table 8-2. Essential pathway activation across major diffuse glioma subtypes.

	Proliferation	Cell Cycle	Telomere Maintenance
IDH wild-type	*EGFR* *PTEN* +7/-10	*CDKN2A/B* *TP53*	*TERT* promoter
IDH mut	IDH	*TP53*	*ATRX*
IDH mut and 1p/19q-codel	IDH	1p/19q (?)	*TERT* promoter

Indicates major genes hit in a predominant share of tumors across the three subtypes determined by IDH and 1p/19q status.

The role of molecular neuropathology in the diagnostic workup of glioma patients

Since the start of this PhD trajectory, molecular testing is increasingly being used in a clinical setting for diagnostic purposes. The discovery of a highly sensitive and specific IDH1 R132H antibody for immunohistochemistry (IHC) use has been instrumental in making molecular neuropathology much more widely available for clinical practice [489]. The R132H *IDH1* variant makes up the majority of IDH mutations in diffuse glioma. Moreover, the IHC test is highly sensitive and specific and requires no specialized equipment. In fact, as I will argue in the following paragraphs, IHC testing alone at a minimum provides an excellent starting point and, in many cases, can provide a definitive diagnosis according to the revised 4th edition WHO classification (**Table 8-3**).

Especially in children and younger adults, histone 3 (H3) variants affecting a lysine at position 27 (K27) or glycine at position 34 (G34) are prevalent. Fortunately, much like IDH R132H, good IHC antibodies exist for all three H3 variants (H3G34R/V, H3K27M), enabling the detection of subtypes characterized by this variant without resorting to more comprehensive testing.

As discussed elsewhere in this thesis, patients with an IDH mutant astrocytoma frequently demonstrate loss-of-function variants of *ATRX*, and these variants are rarely seen in IDH wild-type glioma. Because *ATRX* mutations are not seen in IDH mutant oligodendrogliomas, negative ATRX protein expression on IHC (indicative of an *ATRX* mutation) can help steer the diagnosis towards IDH mutant astrocytoma [21].

Further IHC testing can provide additional evidence to assist the diagnosis. Evidence for elevated Epidermal Growth Factor Receptor (EGFR) protein expression can support a glioblastoma diagnosis. Positive staining for BRAF V600E can support a non-diffuse glioma diagnosis or support the diagnosis of an IDH wild-type diffuse astrocytic tumor when backed by other evidence of this tumor type. Strong nuclear positivity of p53 is often observed alongside *ATRX* mutations in IDH mutant astrocytoma and can help support this diagnosis.

Taken together, with a comprehensive combination of surrogate markers, IHC alone can provide an impressively complete molecular diagnosis fitting neatly into categories defined by the WHO criteria (**Table 8-3**). Moreover, IHC is both relatively affordable and widely available and applicable. Therefore, IHC alone can make a partial molecular neuropathology workup possible even in non-academic lower-resource clinical settings. While DNA sequencing can provide even more complete information, its high cost and limited availability is prohibitive in many clinical settings.

Table 8-3. Comparison of key biomarkers for molecular neuropathology across assays.

	IHC	FISH	DNA methylation array	DNA sequencing
Cost	low	medium	medium	high
Access/availability	high	medium	low	medium
IDH mutation				
IDH1 R132H	X		X*	X
others			X*	X
1p/19q codeletion				
partial 1p36/19q13		X	X	X
complete codeletion			X	X
Combined +7/-10				
partial		X	X	X
complete			X	X
MGMT				
promoter methylation			X	
EGFR				
amplification	X	X	X	
mutation				X
vIII variant		X		X
CDKN2A/B				
deletion		X	X	X

Table 8-3. Comparison of key biomarkers for molecular neuropathology across assays. *Continued*

	IHC	FISH	DNA methylation array	DNA sequencing
Cost	low	medium	medium	high
Access/availability	high	medium	low	medium
***TP53* (p53)**				
mutation	X**			X
ATRX				
mutation	X			X
BRAF				
V600E mutation	X		X*	X
fusion		X	X*	
H3F3A*, *HIST1H3B/C				
K27 mutation	X		X*	X
G34 mutation	X		X*	X
TERT				
promoter mutation				X
Other RTK alteration* **				
fusion		X	X*	

Relative cost of each test is indicated in the second row. Note that other techniques for the assessment of molecular markers exist, but the discussion of these techniques is outside the scope of this thesis [490]. * DNA methylation array is not able to directly test for the presence or absence of this marker, however it is able to test for the resulting phenotype with extremely high accuracy. ** p53 mutations are notoriously difficult to detect from IHC because missense mutants may lead to nuclear accumulation and deleterious mutants lead to loss of protein expression. *** Includes fusion involving *MYB*, *MYBL*, *FGFR1*, *FGFR3*, *NTRK*, *MET*, *RELA*.

As this thesis emphasizes the value of DNA methylation in patient prognostication, I would not do this work justice if I did not argue in favor of its clinical application. As discussed in several chapters of this thesis, DNA methylation profiling using the Illumina 450k or EPIC (850k probe) microarray platform can predict tumor subtypes with extremely high accuracy [94]. While these tests require specialized equipment and expertise, they provide a substantial amount of clinically relevant information with a single test. Several centers worldwide have started to incorporate this test as a part of the workup for most brain tumor patients. In one center, analysis of patient material using the methylation microarray platform led to a change or refinement of the diagnosis in 73% of cases [491].

DNA methylation arrays can also replace other tests on several fronts. Methylation arrays can supersede fluorescence in situ hybridization (FISH) in the detection of copy number events including broad changes such as the 1p/19q codeletion as well as focal events including an *EGFR* amplification and *CDKN2A/B* loss. Unlike FISH, array-based tests provide a picture of the entire genome and are not prone to reporting false positive partial copy number events as changes affecting entire chromosome arms. While DNA sequencing can also provide comprehensive copy number information, because this requires broad coverage across the genome, array-based tests (including Illumina EPIC methylation profiling) for copy number are several orders of magnitude more affordable.

A shortcoming of the DNA methylation array is that it cannot test directly for specific mutations or gene fusions, which can be accurately detected by DNA sequencing or FISH, respectively. However, we and others have found that clinically relevant biomarkers such as IDH, H3 and MAPK/RTK fusion positive tumor subtypes can be predicted from the DNA methylation data with high confidence, overriding the need for direct testing to detect these variants. Thus, in my opinion, in combination with traditional histology and IHC, the DNA methylation microarray test should be considered the "gold standard" for the diagnostic workup of CNS tumors.

Tumor genomics paving the road for precision medicine in glioma

A critical next step is translating breakthroughs in the molecular taxonomy of tumors into targeted treatment and effective therapies. One group of variants that has drawn particular translational interest over the past decade are variants affecting the *EGFR* gene. *EGFR* is often amplified and/or mutated in IDH wild-type glioma. Moreover, it is commonly affected by the EGFR variant III (EGFRvIII) variant. In this variant, exons 2-7 of *EGFR* are lost, affecting the extracellular receptor domains of this transmembrane glycoprotein and leading to its constituent activity. Unfortunately, trials using the EGFR inhibitor erlotinib, the EGFRvIII-specific peptide vaccine rindopepimut or the EGFR antibody-drug conjugate depatuxizumab mafodotin have been unsuccessful [492-494]. It remains to be seen whether new strategies to target *EGFR* variants will arise, or whether responders to existing drugs can be accurately identified, but so far this area of investigation is looking grim.

Protein kinases such as EGFR are attractive therapeutic targets because of the strong dependency of protein kinase driven signaling pathways in oncology. While drugs targeting EGFR have not led to benefits for glioma patients, trials directed at other kinases have been somewhat more successful. Drugs target-

ing the BRAF V600E hotspot variant, *BRAF* fusion, *NTRK* fusion and *FGFR3* fusion have shown promise in preclinical studies and phase II/III clinical trials [495-498]. Nevertheless, these latter alterations, although frequent in certain subtypes, are rare events among all diffuse glioma and collectively make up about 5% of glioma patients [499, 500]. Thus, optimistically assuming that these new drugs lead to a breakthrough in disease management for these patients, at least 95% of glioma patients would remain without good treatment options.

One possible explanation for EGFR inhibition treatment failure in glioma is that these changes are rarely clonal, meaning that they arise late during tumor development and are not a characteristic feature of all tumor cells [501]. Thus, therapies targeting EGFR may only affect a variable subset of the tumor cell population. It is now broadly recognized that effective therapies need to preferably target a tumor at its trunk, defined by the set of changes that are shared across all tumor cells [502]. An obvious target to consider is the IDH mutation, which is the characteristic feature of all canonical oligodendrogliomas and IDH-mutant astrocytomas. While drugs directly repressing mutant IDH protein activity have been approved for use in leukemia, they are still in trial for use in glioma [503]. One potential pitfall of this strategy is that pre-clinical studies using these drugs have shown that, while critical for tumor development, the effects of the mutant IDH protein are dispensable for tumor maintenance [426]. Therefore, suppressing mutant IDH activity in a symptomatic tumor may not have the desired effect of tumor suppression. A second strategy utilizing truncal IDH mutations as a therapeutic target is the cytotoxic targeting of the IDH1 R132H neoantigen, which appears to be strongly expressed in all tumor cells [503, 504]. Exciting results from a very recent phase 1 trial of a vaccine targeting IDH1 R132H show that this therapeutic strategy is safe and leads to treatment response in a large number of patients [505].

Closing thoughts

Genomic sequencing has progressed with leaps and bounds over the past two decades. Interestingly, somewhat similar to Moore's Law (born from the observation that computing power doubled every two years), data from the National Human Genome Research Institute indicated that the cost to sequence a human genome approximately halved every year for the first decade following the millennium. However, since the start of 2008 the decline of sequencing costs rapidly outpaced Moore's Law. In other words, at the turn of the millennium the costs for sequencing an entire human genome were anywhere between $100 and $300

million, while these costs were reduced to roughly $10 million in 2008, and today are approximately $1000.

Such advancements have brought along with them unprecedented changes to our view of human health and disease. New insights gained from sequencing are already driving disease diagnosis. However, history has taught us that translation of such information into improved patient outcomes is years behind biological discovery. The Philadelphia chromosome was first discovered in 1960 in patients with chronic myelogenous leukemia [506, 507]. It took over forty years for the first phase I and II trials on imatinib to complete, the inhibitor of the BCR-ABL tyrosine kinase expressed on the Philadelphia chromosome that provided a curative treatment for a previously uncurable disease [508, 509]. For comparison, IDH mutations were discovered in 2008 and *TERT* promoter mutations were first described in 2013 [48, 256]. So, if historical data can be extrapolated to today, then we are not too far away from a life changing therapy for glioma and the question is not "if", but "when".

CHAPTER 9

Appendices

References

1. Ostrom, Q.T., et al., *CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2013-2017.* Neuro Oncol, 2020. **22**(12 Suppl 2): p. iv1-iv96.
2. Perry, A. and P. Wesseling, *Histologic classification of gliomas.* Handb Clin Neurol, 2016. **134**: p. 71-95.
3. Wesseling, P. and D. Capper, *WHO 2016 Classification of gliomas.* Neuropathol Appl Neurobiol, 2018. **44**(2): p. 139-150.
4. van den Bent, M.J., *Interobserver variation of the histopathological diagnosis in clinical trials on glioma: a clinician's perspective.* Acta Neuropathol, 2010. **120**(3): p. 297-304.
5. Coons, S.W., et al., *Improving diagnostic accuracy and interobserver concordance in the classification and grading of primary gliomas.* Cancer, 1997. **79**(7): p. 1381-93.
6. Aldape, K., et al., *Discrepancies in diagnoses of neuroepithelial neoplasms: the San Francisco Bay Area Adult Glioma Study.* Cancer, 2000. **88**(10): p. 2342-9.
7. Vogelstein, B., et al., *Cancer genome landscapes.* Science, 2013. **339**(6127): p. 1546-58.
8. Tomasetti, C., et al., *Only three driver gene mutations are required for the development of lung and colorectal cancers.* Proc Natl Acad Sci U S A, 2015. **112**(1): p. 118-23.
9. Vogelstein, B. and K.W. Kinzler, *The Path to Cancer --Three Strikes and You're Out.* N Engl J Med, 2015. **373**(20): p. 1895-8.
10. Cairncross, G., et al., *Phase III trial of chemoradiotherapy for anaplastic oligodendroglioma: long-term results of RTOG 9402.* J Clin Oncol, 2013. **31**(3): p. 337-43.
11. Cairncross, J.G., et al., *Specific genetic predictors of chemotherapeutic response and survival in patients with anaplastic oligodendrogliomas.* J Natl Cancer Inst, 1998. **90**(19): p. 1473-9.

12. Eckel-Passow, J.E., et al., *Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors.* N Engl J Med, 2015. **372**(26): p. 2499-508.
13. Cancer Genome Atlas Research, N., et al., *Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas.* N Engl J Med, 2015. **372**(26): p. 2481-98.
14. Wiestler, B., et al., *Assessing CpG island methylator phenotype, 1p/19q codeletion, and MGMT promoter methylation from epigenome-wide data in the biomarker cohort of the NOA-04 trial.* Neuro Oncol, 2014. **16**(12): p. 1630-8.
15. Ceccarelli, M., et al., *Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma.* Cell, 2016. **164**(3): p. 550-63.
16. Yan, H., et al., *IDH1 and IDH2 mutations in gliomas.* N Engl J Med, 2009. **360**(8): p. 765-73.
17. Schindler, G., et al., *Analysis of BRAF V600E mutation in 1,320 nervous system tumors reveals high mutation frequencies in pleomorphic xanthoastrocytoma, ganglioglioma and extra-cerebellar pilocytic astrocytoma.* Acta Neuropathol, 2011. **121**(3): p. 397-405.
18. Kepes, J.J., L.J. Rubinstein, and L.F. Eng, *Pleomorphic xanthoastrocytoma: a distinctive meningocerebral glioma of young subjects with relatively favorable prognosis. A study of 12 cases.* Cancer, 1979. **44**(5): p. 1839-52.
19. Collins, V.P., D.T. Jones, and C. Giannini, *Pilocytic astrocytoma: pathology, molecular mechanisms and markers.* Acta Neuropathol, 2015. **129**(6): p. 775-88.
20. Olar, A. and K.D. Aldape, *Using the molecular classification of glioblastoma to inform personalized treatment.* J Pathol, 2014. **232**(2): p. 165-77.
21. Reuss, D.E., et al., *ATRX and IDH1-R132H immunohistochemistry with subsequent copy number analysis and IDH sequencing as a basis for an "integrated" diagnostic approach for adult astrocytoma, oligodendroglioma and glioblastoma.* Acta Neuropathol, 2015. **129**(1): p. 133-46.
22. Verhaak, R.G., et al., *Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1.* Cancer Cell, 2010. **17**(1): p. 98-110.
23. Noushmehr, H., et al., *Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma.* Cancer Cell, 2010. **17**(5): p. 510-22.
24. Brennan, C.W., et al., *The somatic genomic landscape of glioblastoma.* Cell, 2013. **155**(2): p. 462-77.
25. Virchow, R., *Die krankhaften Geschwülste. Dreissig Vorlesungen, gehalten während des Wintersemesters 1862-1863 an Der Universität Zu Berlin.* 1863, Berlin, Germany: A. Hirschwald.
26. Bailey, P. and H. Cushing, *Microchemical Color Reactions as an Aid to the Identification and Classification of Brain Tumors.* Proc Natl Acad Sci U S A, 1925. **11**(1): p. 82-4.
27. Broders, A.C., *Carcinoma: Grading the malignancy of carcinoma, grading and practical application.* Arch. Path. (Lab. Med.), 1926. **2**: p. 376.
28. Scherer, H.J., *A Critical Review: The Pathology of Cerebral Gliomas.* Journal of Neurology and Psychiatry, 1940: p. 31.
29. Kernohan, J.W., R.F. Mabon, and et al., *A simplified classification of the gliomas.* Proc Staff Meet Mayo Clin, 1949. **24**(3): p. 71-5.
30. Ringertz, N., *Grading of gliomas.* Acta Pathol Microbiol Scand, 1950. **27**(1): p. 51-64.
31. Louis, D.N. and A. von Deimling, *Grading of diffuse astrocytic gliomas: Broders, Kernohan, Zulch, the WHO... and Shakespeare.* Acta Neuropathol, 2017.
32. Zülch, K., *Histologic typing of tumours of the central nervous system.* World Health Organization Classification of Tumours. 1979, Geneva, Switzerland: World Health Organization.

33. Kleihues, P., P. Burger, and B. Scheithauer, *Histological typing of tumours of the central nervous system. World Health Organization international histological classification of tumours, 2nd Ed.* World Health Organization Classification of Tumours. 1993, Berlin, Heidelberg: Springer Verlag.
34. Kleihues, P., P.C. Burger, and B.W. Scheithauer, *The new WHO classification of brain tumours.* Brain pathology (Zurich, Switzerland), 1993. **3**(3): p. 255-68.
35. Daumas-Duport, C., et al., *Grading of astrocytomas. A simple and reproducible method.* Cancer, 1988. **62**(10): p. 2152-65.
36. Kleihues, P. and W. Cavenee, *World Health Organisation classification of tumours: Pathology and genetics of tumours of the nervous system.* World Health Organization Classification of Tumours. 2000, Lyon, France: IARC Press.
37. Louis, D.N., et al., *WHO Classification of Tumours of the Central Nervous System.* 4 ed. World Health Organization Classification of Tumours. 2007, Lyon, France: IARC Press.
38. Kleihues, P. and L.H. Sobin, *World Health Organization classification of tumors.* Cancer, 2000. **88**(12): p. 2887.
39. Louis, D.N., et al., *The 2007 WHO classification of tumours of the central nervous system.* Acta Neuropathol, 2007. **114**(2): p. 97-109.
40. Scheithauer, B.W., *Development of the WHO classification of tumors of the central nervous system: a historical perspective.* Brain Pathol, 2009. **19**(4): p. 551-64.
41. Kros, J.M., et al., *Oligodendroglioma. A comparison of two grading systems.* Cancer, 1988. **61**(11): p. 2251-9.
42. Kros, J.M., et al., *Panel review of anaplastic oligodendroglioma from European Organization For Research and Treatment of Cancer Trial 26951: assessment of consensus in diagnosis, influence of 1p/19q loss, and correlations with outcome.* J Neuropathol Exp Neurol, 2007. **66**(6): p. 545-51.
43. Kros, J.M., *Grading of gliomas: the road from eminence to evidence.* J Neuropathol Exp Neurol, 2011. **70**(2): p. 101-9.
44. Reifenberger, J., et al., *Molecular genetic analysis of oligodendroglial tumors shows preferential allelic deletions on 19q and 1p.* Am J Pathol, 1994. **145**(5): p. 1175-90.
45. Kraus, J.A., et al., *Shared allelic losses on chromosomes 1p and 19q suggest a common origin of oligodendroglioma and oligoastrocytoma.* J Neuropathol Exp Neurol, 1995. **54**(1): p. 91-5.
46. von Deimling, A., et al., *Evidence for a tumor suppressor gene on chromosome 19q associated with human astrocytomas, oligodendrogliomas, and mixed gliomas.* Cancer Res, 1992. **52**(15): p. 4277-9.
47. Cairncross, J.G. and D.R. Macdonald, *Successful chemotherapy for recurrent malignant oligodendroglioma.* Ann Neurol, 1988. **23**(4): p. 360-4.
48. Parsons, D.W., et al., *An integrated genomic analysis of human glioblastoma multiforme.* Science, 2008. **321**(5897): p. 1807-12.
49. Watanabe, T., et al., *IDH1 mutations are early events in the development of astrocytomas and oligodendrogliomas.* Am J Pathol, 2009. **174**(4): p. 1149-53.
50. Flavahan, W.A., et al., *Insulator dysfunction and oncogene activation in IDH mutant gliomas.* Nature, 2016. **529**(7584): p. 110-4.
51. Flavahan, W.A., E. Gaskell, and B.E. Bernstein, *Epigenetic plasticity and the hallmarks of cancer.* Science, 2017. **357**(6348): p. eaal2380.
52. Jiao, Y., et al., *Frequent ATRX, CIC, FUBP1 and IDH1 mutations refine the classification of malignant gliomas.* Oncotarget, 2012. **3**(7): p. 709-22.

53. Esteller, M., et al., *Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents.* N Engl J Med, 2000. **343**(19): p. 1350-4.

54. Esteller, M., et al., *Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia.* Cancer Res, 1999. **59**(4): p. 793-7.

55. Costello, J.F., et al., *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, 1994. **269**(25): p. 17228-37.

56. Hegi, M.E., et al., *MGMT gene silencing and benefit from temozolomide in glioblastoma.* The New England journal of medicine, 2005. **352**(10): p. 997-1003.

57. Wick, W., et al., *Temozolomide chemotherapy alone versus radiotherapy alone for malignant astrocytoma in the elderly: the NOA-08 randomised, phase 3 trial.* Lancet Oncol, 2012. **13**(7): p. 707-15.

58. Wick, W., et al., *MGMT testing-the challenges for biomarker-based glioma treatment.* Nat Rev Neurol, 2014. **10**(7): p. 372-85.

59. Louis, D.N., et al., *International Society Of Neuropathology--Haarlem consensus guidelines for nervous system tumor classification and grading.* Brain Pathol, 2014. **24**(5): p. 429-35.

60. Louis, D.N., et al., *WHO Classification of Tumours of the Central Nervous System.* World Health Organization Classification of Tumours. 2016, Lyon, France: IARC Press.

61. Louis, D.N., et al., *The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary.* Acta Neuropathol, 2016. **131**(6): p. 803-20.

62. Barresi, V., et al., *Dual-Genotype Diffuse Low-Grade Glioma: Is It Really Time to Abandon Oligoastrocytoma As a Distinct Entity?* J Neuropathol Exp Neurol, 2017. **76**(5): p. 342-346.

63. Jakola, A.S., et al., *COmparison of a strategy favoring early surgical resection vs a strategy favoring watchful waiting in low-grade gliomas.* JAMA, 2012. **308**(18): p. 1881-1888.

64. Jakola, A.S., et al., *Surgical resection versus watchful waiting in low-grade gliomas.* Ann Oncol, 2017. **28**(8): p. 1942-1948.

65. Stupp, R., et al., *Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial.* Lancet Oncology, 2009. **10**(5): p. 459-466.

66. Stupp, R., et al., *Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma.* The New England journal of medicine, 2005. **352**(10): p. 987-996.

67. van den Bent, M.J., et al., *Long-term efficacy of early versus delayed radiotherapy for low-grade astrocytoma and oligodendroglioma in adults: the EORTC 22845 randomised trial.* Lancet, 2005. **366**(9490): p. 985-90.

68. Ryken, T.C., et al., *The role of radiotherapy in the management of patients with diffuse low grade glioma: A systematic review and evidence-based clinical practice guideline.* J Neurooncol, 2015. **125**(3): p. 551-83.

69. Dhawan, S., et al., *Early versus delayed postoperative radiotherapy for treatment of low-grade gliomas.* Cochrane Database Syst Rev, 2020. **1**: p. CD009229.

70. Phillips, H.S., et al., *Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis.* Cancer Cell, 2006. **9**(3): p. 157-73.

71. Sottoriva, A., et al., *Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics.* Proc Natl Acad Sci U S A, 2013. **110**(10): p. 4009-14.

72. Gill, B.J., et al., *MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma.* Proc Natl Acad Sci U S A, 2014. **111**(34): p. 12550-5.

73. Kim, H., et al., *Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution.* Genome Res, 2015. **25**(3): p. 316-27.

74. Wang, Q., et al., *Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment.* Cancer Cell, 2017. **32**(1): p. 42-56 e6.

75. Van Loo, P. and P.J. Campbell, *ABSOLUTE cancer genomics.* Nat Biotech, 2012. **30**(7): p. 620-621.

76. Carter, S.L., et al., *Absolute quantification of somatic DNA alterations in human cancer.* Nat Biotechnol, 2012. **30**(5): p. 413-21.

77. Aran, D., M. Sirota, and A.J. Butte, *Systematic pan-cancer analysis of tumour purity.* Nat Commun, 2015. **6**: p. 8971.

78. Deaton, A.M. and A. Bird, *CpG islands and the regulation of transcription.* Genes Dev, 2011. **25**(10): p. 1010-22.

79. Martinez, R., et al., *A microarray-based DNA methylation study of glioblastoma multiforme.* Epigenetics, 2009. **4**(4): p. 255-64.

80. Christensen, B.C., et al., *DNA methylation, isocitrate dehydrogenase mutation, and survival in glioma.* J Natl Cancer Inst, 2011. **103**(2): p. 143-53.

81. Xu, W., et al., *Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases.* Cancer Cell, 2011. **19**(1): p. 17-30.

82. Lu, C., et al., *IDH mutation impairs histone demethylation and results in a block to cell differentiation.* Nature, 2012. **483**(7390): p. 474-8.

83. McCarthy, N., *Metabolism: unmasking an oncometabolite.* Nat Rev Cancer, 2012. **12**(4): p. 229.

84. Turcan, S., et al., *IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype.* Nature, 2012. **483**(7390): p. 479-83.

85. Sandoval, J., et al., *Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome.* Epigenetics, 2011. **6**(6): p. 692-702.

86. Sturm, D., et al., *Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma.* Cancer Cell, 2012. **22**(4): p. 425-37.

87. O'Sullivan, R.J. and J. Karlseder, *Telomeres: protecting chromosomes against genome instability.* Nat Rev Mol Cell Biol, 2010. **11**(3): p. 171-81.

88. Olovnikov, A.M., *A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon.* J Theor Biol, 1973. **41**(1): p. 181-90.

89. de Lange, T., *How telomeres solve the end-protection problem.* Science, 2009. **326**(5955): p. 948-52.

90. Hackett, J.A. and C.W. Greider, *Balancing instability: dual roles for telomerase and telomere dysfunction in tumorigenesis.* Oncogene, 2002. **21**(4): p. 619-26.

91. Shay, J.W., *Role of Telomeres and Telomerase in Aging and Cancer.* Cancer Discov, 2016.

92. Sobinoff, A.P. and H.A. Pickett, *Alternative Lengthening of Telomeres: DNA Repair Pathways Converge.* Trends Genet, 2017. **33**(12): p. 921-932.

93. Louis, D.N., et al., *Announcing cIMPACT-NOW: the Consortium to Inform Molecular and Practical Approaches to CNS Tumor Taxonomy.* Acta Neuropathol, 2017. **133**(1): p. 1-3.

94. Capper, D., et al., *DNA methylation-based classification of central nervous system tumours.* Nature, 2018. **555**(7697): p. 469-474.

95. Shirahata, M., et al., *Novel, improved grading system(s) for IDH-mutant astrocytic gliomas.* Acta Neuropathol, 2018. **136**(1): p. 153-166.

96. Gonzalez Castro, L.N. and P. Wesseling, *The cIMPACT-NOW updates and their significance to current neuro-oncology practice.* Neurooncol Pract, 2021. **8**(1): p. 4-10.

97. Schwartzbaum, J.A., et al., *Epidemiology and molecular pathology of glioma.* Nat Clin Pract Neurol, 2006. **2**(9): p. 494-503; quiz 1 p following 516.

98. Frattini, V., et al., *The integrated landscape of driver genomic alterations in glioblastoma.* Nat Genet, 2013. **45**(10): p. 1141-9.

99. Suzuki, H., et al., *Mutational landscape and clonal architecture in grade II and III gliomas.* Nat Genet, 2015. **47**(5): p. 458-68.

100. TCGA_Network, et al., *Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas.* N Engl J Med, 2015. **372**(26): p. 2481-98.

101. Peters, J.M. and T. Nishiyama, *Sister chromatid cohesion.* Cold Spring Harb Perspect Biol, 2012. **4**(11).

102. Kon, A., et al., *Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms.* Nat Genet, 2013. **45**(10): p. 1232-7.

103. Killela, P.J., et al., *TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal.* Proc Natl Acad Sci U S A, 2013. **110**(15): p. 6021-6.

104. Horn, S., et al., *TERT promoter mutations in familial and sporadic melanoma.* Science, 2013. **339**(6122): p. 959-61.

105. Ozawa, T., et al., *Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma.* Cancer Cell, 2014. **26**(2): p. 288-300.

106. Sturm, D., et al., *Paediatric and adult glioblastoma: multiform (epi)genomic culprits emerge.* Nat Rev Cancer, 2014. **14**(2): p. 92-107.

107. Pineda, C.T., et al., *Degradation of AMPK by a cancer-specific ubiquitin ligase.* Cell, 2015. **160**(4): p. 715-28.

108. Kurscheid, S., et al., *Chromosome 7 gain and DNA hypermethylation at the HOXA10 locus are associated with expression of a stem cell related HOX-signature in glioblastoma.* Genome Biol, 2015. **16**: p. 16.

109. Berman, B.P., et al., *Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains.* Nat Genet, 2012. **44**(1): p. 40-6.

110. Lodato, M.A., et al., *SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state.* PLoS Genet, 2013. **9**(2): p. e1003288.

111. Graham, V., et al., *SOX2 functions to maintain neural progenitor identity.* Neuron, 2003. **39**(5): p. 749-65.

112. Suva, M.L., et al., *Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells.* Cell, 2014. **157**(3): p. 580-94.

113. Mur, P., et al., *Codeletion of 1p and 19q determines distinct gene methylation and expression profiles in IDH-mutated oligodendroglial tumors.* Acta Neuropathol, 2013. **126**(2): p. 277-89.

114. Lambert, S.R., et al., *Differential expression and methylation of brain developmental genes define location-specific subsets of pilocytic astrocytoma.* Acta Neuropathol, 2013. **126**(2): p. 291-301.

115. Guintivano, J., M.J. Aryee, and Z.A. Kaminsky, *A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression.* Epigenetics, 2013. **8**(3): p. 290-302.

116. Jones, D.T., et al., *Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma.* Nat Genet, 2013. **45**(8): p. 927-32.

117. Chi, A.S., et al., *BRAF V600E mutation identifies a subset of low-grade diffusely infiltrating gliomas in adults.* J Clin Oncol, 2013. **31**(14): p. e233-6.

118. Roggendorf, W., S. Strupp, and W. Paulus, *Distribution and characterization of microglia/macrophages in human brain tumors.* Acta Neuropathol, 1996. **92**(3): p. 288-93.

119. Yoshihara, K., et al., *Inferring tumour purity and stromal and immune cell admixture from expression data.* Nat Commun, 2013. **4**: p. 2612.

120. Olar, A., et al., *IDH mutation status and role of WHO grade and mitotic index in overall survival in grade II-III diffuse gliomas.* Acta Neuropathol, 2015. **129**(4): p. 585-96.

121. Reuss, D.E., et al., *IDH mutant diffuse and anaplastic astrocytomas have similar age at presentation and little difference in survival: a grading problem for WHO.* Acta Neuropathol, 2015. **129**(6): p. 867-73.

122. Mazor, T., et al., *DNA Methylation and Somatic Mutations Converge on the Cell Cycle and Define Similar Evolutionary Histories in Brain Tumors.* Cancer Cell, 2015. **28**(3): p. 307-17.

123. Holland, E.C., et al., *Combined activation of Ras and Akt in neural progenitors induces glioblastoma formation in mice.* Nat Genet, 2000. **25**(1): p. 55-7.

124. Solomon, D.A., et al., *Mutational inactivation of STAG2 causes aneuploidy in human cancer.* Science, 2011. **333**(6045): p. 1039-43.

125. Bailey, M.L., et al., *Glioblastoma cells containing mutations in the cohesin component STAG2 are sensitive to PARP inhibition.* Mol Cancer Ther, 2014. **13**(3): p. 724-32.

126. Borah, S., et al., *Cancer. TERT promoter mutations and telomerase reactivation in urothelial cancer.* Science, 2015. **347**(6225): p. 1006-10.

127. Flynn, R.L., et al., *Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors.* Science, 2015. **347**(6219): p. 273-7.

128. Barthel, F.P., P. Wesseling, and R.G.W. Verhaak, *Reconstructing the molecular life history of gliomas.* Acta Neuropathol, 2018. **135**(5): p. 649-670.

129. Bettegowda, C., et al., *Mutations in CIC and FUBP1 contribute to human oligodendroglioma.* Science, 2011. **333**(6048): p. 1453-5.

130. Zheng, S., et al., *A survey of intragenic breakpoints in glioblastoma identifies a distinct subset associated with poor survival.* Genes Dev, 2013. **27**(13): p. 1462-72.

131. Cancer Genome Atlas Research, N., *Comprehensive genomic characterization defines human glioblastoma genes and core pathways.* Nature, 2008. **455**(7216): p. 1061-8.

132. Venteicher, A.S., et al., *Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq.* Science, 2017. **355**(6332).

133. Patel, A.P., et al., *Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma.* Science, 2014. **344**(6190): p. 1396-401.

134. Snuderl, M., et al., *Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma.* Cancer Cell, 2011. **20**(6): p. 810-7.

135. Nejo, T., et al., *Reduced Neoantigen Expression Revealed by Longitudinal Multiomics as a Possible Immune Evasion Mechanism in Glioma.* Cancer Immunol Res, 2019.

136. Williams, M.J., et al., *Quantification of subclonal selection in cancer from bulk sequencing data.* Nat Genet, 2018. **50**(6): p. 895-903.

137. Consortium, G., *Glioma through the looking GLASS: molecular evolution of diffuse gliomas and the Glioma Longitudinal Analysis Consortium.* Neuro Oncol, 2018. **20**(7): p. 873-884.

138. Hu, H., et al., *Mutational Landscape of Secondary Glioblastoma Guides MET-Targeted Trial in Brain Tumor.* Cell, 2018. **175**(6): p. 1665-1678 e18.

139. Wang, J., et al., *Clonal evolution of glioblastoma under therapy.* Nat Genet, 2016. **48**(7): p. 768-76.

140. Droop, A., et al., *How to analyse the spatiotemporal tumour samples needed to investigate cancer evolution: A case study using paired primary and recurrent glioblastoma.* Int J Cancer, 2018. **142**(8): p. 1620-1626.

141. Johnson, B.E., et al., *Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma.* Science, 2014. **343**(6167): p. 189-93.

142. Kim, J., et al., *Spatiotemporal Evolution of the Primary Glioblastoma Genome.* Cancer Cell, 2015. **28**(3): p. 318-28.

143. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer.* Nature, 2013. **500**(7463): p. 415-21.

144. Hunter, C., et al., *A hypermutation phenotype and somatic MSH6 mutations in recurrent human malignant gliomas after alkylator chemotherapy.* Cancer Res, 2006. **66**(8): p. 3987-91.

145. Jolly, C. and P. Van Loo, *Timing somatic events in the evolution of cancer.* Genome Biol, 2018. **19**(1): p. 95.

146. Turajlic, S., et al., *Resolving genetic heterogeneity in cancer.* Nat Rev Genet, 2019.

147. Choi, S., et al., *Temozolomide-associated hypermutation in gliomas.* Neuro Oncol, 2018. **20**(10): p. 1300-1309.

148. Baumert, B.G., et al., *Temozolomide chemotherapy versus radiotherapy in high-risk low-grade glioma (EORTC 22033-26033): a randomised, open-label, phase 3 intergroup study.* Lancet Oncol, 2016. **17**(11): p. 1521-1532.

149. Buckner, J.C., et al., *Radiation plus Procarbazine, CCNU, and Vincristine in Low-Grade Glioma.* N Engl J Med, 2016. **374**(14): p. 1344-55.

150. Yap, T.A., et al., *Intratumor heterogeneity: seeing the wood for the trees.* Sci Transl Med, 2012. **4**(127): p. 127ps10.

151. Williams, M.J., et al., *Identification of neutral tumor evolution across cancer types.* Nat Genet, 2016. **48**(3): p. 238-244.

152. Korber, V., et al., *Evolutionary Trajectories of IDH(WT) Glioblastomas Reveal a Common Path of Early Tumorigenesis Instigated Years ahead of Initial Diagnosis.* Cancer Cell, 2019. **35**(4): p. 692-704 e12.

153. deCarvalho, A.C., et al., *Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma.* Nat Genet, 2018. **50**(5): p. 708-717.

154. Giam, M. and G. Rancati, *Aneuploidy and chromosomal instability in cancer: a jackpot to chaos.* Cell Div, 2015. **10**: p. 3.

155. Marty, R., et al., *Evolutionary Pressure against MHC Class II Binding Cancer Mutations.* Cell, 2018. **175**(2): p. 416-428 e13.

156. McGranahan, N., et al., *Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution.* Cell, 2017. **171**(6): p. 1259-1271 e11.

157. Rooney, M.S., et al., *Molecular and genetic properties of tumors associated with local immune cytolytic activity.* Cell, 2015. **160**(1-2): p. 48-61.

158. Dunn, G.P., et al., *Cancer immunoediting: from immunosurveillance to tumor escape.* Nat Immunol, 2002. **3**(11): p. 991-8.

159. Hundal, J., et al., *pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens.* Genome Med, 2016. **8**(1): p. 11.

160. Newman, A.M., et al., *Robust enumeration of cell subsets from tissue expression profiles.* Nat Methods, 2015. **12**(5): p. 453-7.

161. Rosenthal, R., et al., *Neoantigen-directed immune escape in lung cancer evolution.* Nature, 2019. **567**(7749): p. 479-485.

162. Raub, T.J., et al., *Brain Exposure of Two Selective Dual CDK4 and CDK6 Inhibitors and the Antitumor Activity of CDK4 and CDK6 Inhibition in Combination with Temozolomide in an Intracranial Glioblastoma Xenograft.* Drug Metab Dispos, 2015. **43**(9): p. 1360-71.

163. van den Bent, M., et al., *Efficacy of depatuxizumab mafodotin (ABT-414) monotherapy in patients with EGFR-amplified, recurrent glioblastoma: results from a multi-center, international study.* Cancer Chemother Pharmacol, 2017. **80**(6): p. 1209-1217.

164. Keskin, D.B., et al., *Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial.* Nature, 2019. **565**(7738): p. 234-239.

165. Schumacher, T., et al., *A vaccine targeting mutant IDH1 induces antitumour immunity.* Nature, 2014. **512**(7514): p. 324-7.

166. Koster, J. and S. Rahmann, *Snakemake-a scalable bioinformatics workflow engine.* Bioinformatics, 2018. **34**(20): p. 3600.

167. Van der Auwera, G.A., et al., *From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline.* Curr Protoc Bioinformatics, 2013. **43**: p. 11 10 1-33.

168. Ewels, P., et al., *MultiQC: summarize analysis results for multiple tools and samples in a single report.* Bioinformatics, 2016. **32**(19): p. 3047-8.

169. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.

170. Mermel, C.H., et al., *GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers.* Genome Biol, 2011. **12**(4): p. R41.

171. Beroukhim, R., et al., *Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma.* Proc Natl Acad Sci U S A, 2007. **104**(50): p. 20007-12.

172. Martincorena, I., et al., *Universal Patterns of Selection in Cancer and Somatic Tissues.* Cell, 2017. **171**(5): p. 1029-1041 e21.

173. Taylor, A.M., et al., *Genomic and Functional Approaches to Understanding Cancer Aneuploidy.* Cancer Cell, 2018. **33**(4): p. 676-689 e3.

174. Roth, A., et al., *PyClone: statistical inference of clonal population structure in cancer.* Nat Methods, 2014. **11**(4): p. 396-8.

175. Turajlic, S., et al., *Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal.* Cell, 2018. **173**(3): p. 581-594 e12.

176. Ha, G., et al., *TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data.* Genome Res, 2014. **24**(11): p. 1881-93.

177. Szolek, A., et al., *OptiType: precision HLA typing from next-generation sequencing data.* Bioinformatics, 2014. **30**(23): p. 3310-6.

178. Hoof, I., et al., *NetMHCpan, a method for MHC class I binding prediction beyond humans.* Immunogenetics, 2009. **61**(1): p. 1-13.

179. Ho, V.K., et al., *Changing incidence and improved survival of gliomas.* Eur J Cancer, 2014. **50**(13): p. 2309-18.

180. Lapointe, S., A. Perry, and N.A. Butowski, *Primary brain tumours in adults.* Lancet, 2018. **392**(10145): p. 432-446.

181. Ellingson, B.M., et al., *Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials.* Neuro Oncol, 2015. **17**(9): p. 1188-98.

182. Pallud, J., et al., *Diffuse low-grade oligodendrogliomas extend beyond MRI-defined abnormalities.* Neurology, 2010. **74**(21): p. 1724-1731.

183. Kelly, P.J., et al., *Imaging-based stereotaxic serial biopsies in untreated intracranial glial neoplasms.* J Neurosurg, 1987. **66**(6): p. 865-74.

184. Petrecca, K., et al., *Failure pattern following complete resection plus radiotherapy and temozolomide is at the resection margin in patients with glioblastoma.* J Neurooncol, 2013. **111**(1): p. 19-23.

185. Barthel, F.P., et al., *Longitudinal molecular trajectories of diffuse glioma in adults.* Nature, 2019. **576**(7785): p. 112-120.

186. Neftel, C., et al., *An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma.* Cell, 2019. **178**(4): p. 835-849 e21.

187. Szerlip, N.J., 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, 2012. **109**(8): p. 3041-6.

188. Benelli, M., D. Romagnoli, and F. Demichelis, *Tumor purity quantification by clonal DNA methylation signatures.* Bioinformatics, 2018. **34**(10): p. 1642-1649.

189. Verburg, N., et al., *Accurate Delineation of Glioma Infiltration by Advanced PET/MR Neuro-Imaging (FRONTIER Study): A Diagnostic Study Protocol.* Neurosurgery, 2016. **79**(4): p. 535-40.

190. Nathanson, V., *Revising the Declaration of Helsinki.* BMJ, 2013. **346**: p. f2837.

191. Darmanis, S., et al., *Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma.* Cell Rep, 2017. **21**(5): p. 1399-1410.

192. Mazor, T., et al., *DNA Methylation and Somatic Mutations Converge on the Cell Cycle and Define Similar Evolutionary Histories in Brain Tumors.* Cancer Cell, 2015. **28**(3): p. 307-317.

193. Morrissy, A.S., et al., *Spatial heterogeneity in medulloblastoma.* Nat Genet, 2017. **49**(5): p. 780-788.

194. Ziller, M.J., et al., *Charting a dynamic DNA methylation landscape of the human genome.* Nature, 2013. **500**(7463): p. 477-81.

195. Wenger, A., et al., *Intratumor DNA methylation heterogeneity in glioblastoma: implications for DNA methylation-based classification.* Neuro Oncol, 2019. **21**(5): p. 616-627.

196. Kevin C. Johnson, K.J.A., Elise T. Courtois, Floris P. Barthel, Frederick S. Varn, Diane Luo, Martine Seignon, Eunhee Yi, Hoon Kim, Marcos RH Estecio, Ming Tang, Nicholas E. Navin, Rahul Maurya, Chew Yee Ngan, Niels Verburg, Philip C De Witt Hamer, Ketan Bulsara, Michael L. Samuels, Sunit Das, Paul Robson, Roel GW Verhaak, *Single-cell multimodal glioma analyses reveal epigenetic regulators of cellular plasticity and environmental stress response* bioRxiv 2020. **2020.07.22.215335**.

197. Jonsson, P., et al., *Genomic Correlates of Disease Progression and Treatment Response in Prospectively Characterized Gliomas.* Clin Cancer Res, 2019. **25**(18): p. 5537-5547.

198. Verburg, N., et al., *Improved detection of diffuse glioma infiltration with imaging combinations: a diagnostic accuracy study.* Neuro Oncol, 2020. **22**(3): p. 412-422.

199. Barton, M.B., et al., *Estimating the demand for radiotherapy from the evidence: a review of changes from 2003 to 2012.* Radiother Oncol, 2014. **112**(1): p. 140-4.

200. Tyldesley, S., et al., *Estimating the need for radiotherapy for patients with prostate, breast, and lung cancers: verification of model estimates of need with radiotherapy utilization data from British Columbia.* Int J Radiat Oncol Biol Phys, 2011. **79**(5): p. 1507-15.

201. Chang, H.H.Y., et al., *Non-homologous DNA end joining and alternative pathways to double-strand break repair.* Nat Rev Mol Cell Biol, 2017. **18**(8): p. 495-506.

202. Campbell, B.B., et al., *Comprehensive Analysis of Hypermutation in Human Cancer.* Cell, 2017. **171**(5): p. 1042-1056 e10.

203. Touat, M., et al., *Mechanisms and therapeutic implications of hypermutation in gliomas.* Nature, 2020. **580**(7804): p. 517-523.

204. Behjati, S., et al., *Mutational signatures of ionizing radiation in second malignancies.* Nat Commun, 2016. **7**: p. 12605.

205. Priestley, P., et al., *Pan-cancer whole-genome analyses of metastatic solid tumours.* Nature, 2019. **575**(7781): p. 210-216.

206. Pich, O., et al., *The mutational footprints of cancer therapies.* Nat Genet, 2019. **51**(12): p. 1732-1740.

207. Kucab, J.E., et al., *A Compendium of Mutational Signatures of Environmental Agents.* Cell, 2019. **177**(4): p. 821-836 e16.

208. Stupp, R., et al., *Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma.* N Engl J Med, 2005. **352**(10): p. 987-96.

209. Lutz, S.T., J. Jones, and E. Chow, *Role of radiation therapy in palliative care of the patient with cancer.* J Clin Oncol, 2014. **32**(26): p. 2913-9.

210. Nguyen, L., et al., *Pan-cancer landscape of homologous recombination deficiency.* Nat Commun, 2020. **11**(1): p. 5584.

211. Georgakopoulos-Soares, I., et al., *Noncanonical secondary structures arising from non-B DNA motifs are determinants of mutagenesis.* Genome research, 2018. **28**(9): p. 1264-1271.

212. Martincorena, I., et al., *Universal Patterns of Selection in Cancer and Somatic Tissues.* Cell, 2017. **171**(5): p. 1029-1041.e21.

213. Alexandrov, L.B., et al., *The repertoire of mutational signatures in human cancer.* Nature, 2020. **578**(7793): p. 94-101.

214. Davies, H., et al., *HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures.* Nat Med, 2017. **23**(4): p. 517-525.

215. Touat, M., et al., *Mechanisms and therapeutic implications of hypermutation in gliomas.* Nature, 2020. **580**(7804): p. 517-523.

216. Nik-Zainal, S., et al., *Landscape of somatic mutations in 560 breast cancer whole-genome sequences.* Nature, 2016. **534**(7605): p. 47-54.

217. Roberts, S.A., et al., *An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers.* Nat Genet, 2013. **45**(9): p. 970-6.

218. Lei, L., et al., *APOBEC3 induces mutations during repair of CRISPR-Cas9-generated DNA breaks.* Nat Struct Mol Biol, 2018. **25**(1): p. 45-52.

219. Nowarski, R. and M. Kotler, *APOBEC3 cytidine deaminases in double-strand DNA break repair and cancer promotion.* Cancer Res, 2013. **73**(12): p. 3494-8.

220. Nowarski, R., et al., *APOBEC3G enhances lymphoma cell radioresistance by promoting cytidine deaminase-dependent DNA repair.* Blood, 2012. **120**(2): p. 366-75.

221. Volkova, N.V., et al., *Mutational signatures are jointly shaped by DNA damage and repair.* Nat Commun, 2020. **11**(1): p. 2169.

222. Ceccarelli, M., et al., *Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma.* Cell, 2016. **164**(3): p. 550-563.

223. Adewoye, A.B., et al., *The genome-wide effects of ionizing radiation on mutation induction in the mammalian germline.* Nature communications, 2015. **6**: p. 6684-6684.

224. Bakhoum, S.F., et al., *Numerical chromosomal instability mediates susceptibility to radiation treatment.* Nature communications, 2015. **6**: p. 5990-5990.

225. Rose Li, Y., et al., *Mutational signatures in tumours induced by high and low energy radiation in Trp53 deficient mice.* Nat Commun, 2020. **11**(1): p. 394.

226. Touil, N., et al., *Analysis of chromosome loss and chromosome segregation in cytokinesis-blocked human lymphocytes: non-disjunction is the prevalent mistake in chromosome segregation produced by low dose exposure to ionizing radiation.* Mutagenesis, 2000. **15**(1): p. 1-7.

227. Behjati, S., et al., *Mutational signatures of ionizing radiation in second malignancies.* Nat Commun, 2016. **7**: p. 12605.

228. Davidson, P.R., et al., *A pooled mutational analysis identifies ionizing radiation-associated mutational signatures conserved between mouse and human malignancies.* Sci Rep, 2017. **7**(1): p. 7645.

229. Lopez, G.Y., et al., *The genetic landscape of gliomas arising after therapeutic radiation.* Acta Neuropathol, 2019. **137**(1): p. 139-150.

230. Phi, J.H., et al., *Genomic analysis reveals secondary glioblastoma after radiotherapy in a subset of recurrent medulloblastomas.* Acta Neuropathol, 2018. **135**(6): p. 939-953.

231. Hu, Z., et al., *Quantitative evidence for early metastatic seeding in colorectal cancer.* Nat Genet, 2019. **51**(7): p. 1113-1122.

232. Reiter, J.G., et al., *Lymph node metastases develop through a wider evolutionary bottleneck than distant metastases.* Nat Genet, 2020. **52**(7): p. 692-700.

233. Wiggans, A.J., et al., *Poly(ADP-ribose) polymerase (PARP) inhibitors for the treatment of ovarian cancer.* Cochrane Database Syst Rev, 2015(5): p. CD007929.

234. Su, J.M., et al., *A phase I trial of veliparib (ABT-888) and temozolomide in children with recurrent CNS tumors: a pediatric brain tumor consortium report.* Neuro Oncol, 2014. **16**(12): p. 1661-8.

235. O'Neil, N.J., M.L. Bailey, and P. Hieter, *Synthetic lethality and cancer.* Nature reviews. Genetics, 2017. **18**(10): p. 613-623.

236. Munster, P., et al., *First-In-Human Phase I Study Of A Dual mTOR Kinase And DNA-PK Inhibitor (CC-115) In Advanced Malignancy.* Cancer Manag Res, 2019. **11**: p. 10463-10476.

237. Goldberg, F.W., et al., *The Discovery of 7-Methyl-2-[(7-methyl[1,2,4]triazolo[1,5-a]pyridin-6-yl)amino]-9-(tetrahydro-2H-p yran-4-yl)-7,9-dihydro-8H-purin-8-one (AZD7648), a Potent and Selective DNA-Dependent Protein Kinase (DNA-PK) Inhibitor.* J Med Chem, 2020.

238. Thijssen, R., et al., *Dual TORK/DNA-PK inhibition blocks critical signaling pathways in chronic lymphocytic leukemia.* Blood, 2016. **128**(4): p. 574-83.

239. Timme, C.R., et al., *The DNA-PK Inhibitor VX-984 Enhances the Radiosensitivity of Glioblastoma Cells Grown In Vitro and as Orthotopic Xenografts.* Mol Cancer Ther, 2018. **17**(6): p. 1207-1216.

240. Li, M., et al., *First-in-class small molecule inhibitors of the single-strand DNA cytosine deaminase APOBEC3G.* ACS Chem Biol, 2012. **7**(3): p. 506-17.

241. Bins, S., et al., *Implementation of a Multicenter Biobanking Collaboration for Next-Generation Sequencing-Based Biomarker Discovery Based on Fresh Frozen Pretreatment Tumor Tissue Biopsies.* Oncologist, 2017. **22**(1): p. 33-40.

242. Cer, R.Z., et al., *Non-B DB v2.0: a database of predicted non-B DNA-forming motifs and its associated tools.* Nucleic acids research, 2013. **41**(Database issue): p. D94-D100.

243. Blokzijl, F., et al., *MutationalPatterns: comprehensive genome-wide analysis of mutational processes.* Genome medicine, 2018. **10**(1): p. 33-33.

244. Bergstrom, E.N., et al., *SigProfilerMatrixGenerator: a tool for visualizing and exploring patterns of small mutational events.* BMC Genomics, 2019. **20**(1): p. 685.

245. Layer, R.M., et al., *LUMPY: a probabilistic framework for structural variant discovery.* Genome Biol, 2014. **15**(6): p. R84.

246. Abyzov, A., et al., *CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing.* Genome Res, 2011. **21**(6): p. 974-84.

247. Chiang, C., et al., *SpeedSeq: ultra-fast personal genome analysis and interpretation.* Nat Methods, 2015. **12**(10): p. 966-8.

248. Van der Auwera, G.A., et al., *From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline.* Curr Protoc Bioinformatics, 2013. **43**(1110): p. 11 10 1-11 10 33.

249. Shay, J.W., O.M. Pereira-Smith, and W.E. Wright, *A role for both RB and p53 in the regulation of human cellular senescence.* Exp Cell Res, 1991. **196**(1): p. 33-9.

250. Stewart, S.A. and R.A. Weinberg, *Telomeres: cancer to human aging.* Annu Rev Cell Dev Biol, 2006. **22**: p. 531-57.

251. Maser, R.S. and R.A. DePinho, *Connecting chromosomes, crisis, and cancer.* Science, 2002. **297**(5581): p. 565-9.

252. Sahin, E. and R.A. DePinho, *Axis of ageing: telomeres, p53 and mitochondria.* Nat Rev Mol Cell Biol, 2012. **13**(6): p. 397-404.

253. Greider, C.W. and E.H. Blackburn, *Identification of a specific telomere terminal transferase activity in Tetrahymena extracts.* Cell, 1985. **43**(2 Pt 1): p. 405-13.

254. Morales, C.P., et al., *Absence of cancer-associated changes in human fibroblasts immortalized with telomerase.* Nat Genet, 1999. **21**(1): p. 115-8.

255. Shay, J.W. and S. Bacchetti, *A survey of telomerase activity in human cancer.* Eur J Cancer, 1997. **33**(5): p. 787-91.

256. Huang, F.W., et al., *Highly recurrent TERT promoter mutations in human melanoma.* Science, 2013. **339**(6122): p. 957-9.

257. Zhang, A., et al., *Frequent amplification of the telomerase reverse transcriptase gene in human tumors.* Cancer Res, 2000. **60**(22): p. 6230-5.

258. Peifer, M., et al., *Telomerase activation by genomic rearrangements in high-risk neuroblastoma.* Nature, 2015. **526**(7575): p. 700-704.

259. Bryan, T.M., et al., *Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines.* Nat Med, 1997. **3**(11): p. 1271-4.

260. Dilley, R.L. and R.A. Greenberg, *ALTernative Telomere Maintenance and Cancer.* Trends Cancer, 2015. **1**(2): p. 145-156.

261. Jiao, Y., et al., *DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors.* Science, 2011. **331**(6021): p. 1199-203.

262. Heaphy, C.M., et al., *Altered telomeres in tumors with ATRX and DAXX mutations.* Science, 2011. **333**(6041): p. 425.

263. Ramamoorthy, M. and S. Smith, *Loss of ATRX Suppresses Resolution of Telomere Cohesion to Control Recombination in ALT Cancer Cells.* Cancer Cell, 2015. **28**(3): p. 357-69.

264. Ding, Z., et al., *Estimating telomere length from whole genome sequence data.* Nucleic Acids Res, 2014. **42**(9): p. e75.

265. Dlouha, D., et al., *Comparison of the relative telomere length measured in leukocytes and eleven different human tissues.* Physiol Res, 2014. **63 Suppl 3**: p. S343-50.

266. Albanell, J., et al., *Telomerase activity in germ cell cancers and mature teratomas.* J Natl Cancer Inst, 1999. **91**(15): p. 1321-6.

267. Vinagre, J., et al., *Frequency of TERT promoter mutations in human cancers.* Nat Commun, 2013. **4**: p. 2185.

268. Khan, A. and X. Zhang, *dbSUPER: a database of super-enhancers in mouse and human genome.* Nucleic Acids Res, 2016. **44**(D1): p. D164-71.

269. Roadmap Epigenomics, C., et al., *Integrative analysis of 111 reference human epigenomes.* Nature, 2015. **518**(7539): p. 317-30.

270. Calo, E. and J. Wysocka, *Modification of enhancer chromatin: what, how, and why?* Mol Cell, 2013. **49**(5): p. 825-37.

271. Torres-Garcia, W., et al., *PRADA: pipeline for RNA sequencing data analysis.* Bioinformatics, 2014. **30**(15): p. 2224-6.

272. Yoshihara, K., et al., *The landscape and therapeutic relevance of cancer-associated transcript fusions.* Oncogene, 2015. **34**(37): p. 4845-54.

273. Yates, A., et al., *Ensembl 2016.* Nucleic Acids Research, 2016. **44**(D1): p. D710-D716.

274. Hrdlicková, R., J. Nehyba, and H.R. Bose, *Alternatively spliced telomerase reverse transcriptase variants lacking telomerase activity stimulate cell proliferation.* Mol Cell Biol, 2012. **32**(21): p. 4283-96.

275. Castelo-Branco, P., et al., *Methylation of the TERT promoter and risk stratification of childhood brain tumours: an integrative genomic and molecular study.* Lancet Oncol, 2013. **14**(6): p. 534-42.

276. Blasco, M.A., *Telomeres and human disease: ageing, cancer and beyond.* Nat Rev Genet, 2005. **6**(8): p. 611-22.

277. Counter, C.M., et al., *Telomerase activity is restored in human cells by ectopic expression of hTERT (hEST2), the catalytic subunit of telomerase.* Oncogene, 1998. **16**(9): p. 1217-22.

278. Rohde, V., et al., *Expression of the human telomerase reverse transcriptase is not related to telomerase activity in normal and malignant renal tissue.* Clin Cancer Res, 2000. **6**(12): p. 4803-9.

279. Hrdlickova, R., J. Nehyba, and H.R. Bose, Jr., *Alternatively spliced telomerase reverse transcriptase variants lacking telomerase activity stimulate cell proliferation.* Mol Cell Biol, 2012. **32**(21): p. 4283-96.

280. Kilian, A., et al., *Isolation of a candidate human telomerase catalytic subunit gene, which reveals complex splicing patterns in different cell types.* Hum Mol Genet, 1997. **6**(12): p. 2011-9.

281. Wong, M.S., W.E. Wright, and J.W. Shay, *Alternative splicing regulation of telomerase: a new paradigm?* Trends Genet, 2014. **30**(10): p. 430-8.

282. Katz, Y., et al., *Analysis and design of RNA sequencing experiments for identifying isoform regulation.* Nat Methods, 2010. **7**(12): p. 1009-15.

283. Schwartzentruber, J., et al., *Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma.* Nature, 2012. **482**(7384): p. 226-31.

284. Reddel, R.R., *The role of senescence and immortalization in carcinogenesis.* Carcinogenesis, 2000. **21**(3): p. 477-84.

285. Heaphy, C.M., et al., *Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes.* Am J Pathol, 2011. **179**(4): p. 1608-15.

286. Gonzalo, S., et al., *Role of the RB1 family in stabilizing histone methylation at constitutive heterochromatin.* Nat Cell Biol, 2005. **7**(4): p. 420-8.

287. García-Cao, M., et al., *A role for the Rb family of proteins in controlling telomere length.* Nat Genet, 2002. **32**(3): p. 415-9.

288. Robin, J.D., et al., *Telomere position effect: regulation of gene expression with progressive telomere shortening over long distances.* Genes Dev, 2014. **28**(22): p. 2464-76.

289. Martinez, E., et al., *Comparison of gene expression patterns across 12 tumor types identifies a cancer supercluster characterized by TP53 mutations and cell cycle defects.* Oncogene, 2015. **34**(21): p. 2732-40.

290. Renaud, S., et al., *Dual role of DNA methylation inside and outside of CTCF-binding regions in the transcriptional regulation of the telomerase hTERT gene.* Nucleic Acids Res, 2007. **35**(4): p. 1245-56.

291. Valentijn, L.J., et al., *TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors.* Nat Genet, 2015. **47**(12): p. 1411-1414.

292. Davis, C.F., et al., *The somatic genomic landscape of chromophobe renal cell carcinoma.* Cancer Cell, 2014. **26**(3): p. 319-30.

293. Khoury, J.D., et al., *Landscape of DNA virus associations across human malignant cancers: analysis of 3,775 cases using RNA-Seq.* J Virol, 2013. **87**(16): p. 8916-26.

294. Xi, L. and T.R. Cech, *Inventory of telomerase components in human cells reveals multiple subpopulations of hTR and hTERT.* Nucleic Acids Res, 2014. **42**(13): p. 8565-77.

295. Chapon, C., T.R. Cech, and A.J. Zaug, *Polyadenylation of telomerase RNA in budding yeast.* RNA, 1997. **3**(11): p. 1337-51.

296. Porro, A., et al., *Molecular dissection of telomeric repeat-containing RNA biogenesis unveils the presence of distinct and multiple regulatory pathways.* Mol Cell Biol, 2010. **30**(20): p. 4808-17.

297. Feuerhahn, S., et al., *TERRA biogenesis, turnover and implications for function.* FEBS Lett, 2010. **584**(17): p. 3812-8.

298. Clynes, D., et al., *Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX.* Nat Commun, 2015. **6**: p. 7538.

299. Przybycin, C.G., et al., *Chromophobe renal cell carcinoma: a clinicopathologic study of 203 tumors in 200 patients with primary resection at a single institution.* Am J Surg Pathol, 2011. **35**(7): p. 962-70.

300. Guo, Z. and R.V. Lloyd, *Pheochromocytomas and Paragangliomas: An Update on Recent Molecular Genetic Advances and Criteria for Malignancy.* Adv Anat Pathol, 2015. **22**(5): p. 283-93.

301. Davies, L. and H.G. Welch, *Increasing incidence of thyroid cancer in the United States, 1973-2002.* JAMA, 2006. **295**(18): p. 2164-7.

302. Lenders, J.W.M., et al., *Phaeochromocytoma.* Lancet, 2005. **366**(9486): p. 665-75.

303. Seth, S., et al., *Flowr: Robust and efficient pipelines using a simple language-agnostic approach.* bioRxiv, 2015.

304. Wilks, C., et al., *The Cancer Genomics Hub (CGHub): overcoming cancer through the power of torrential data.* Database (Oxford), 2014. **2014**.

305. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.

306. Li, H., *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.* 2013.

307. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.

308. Consortium, R.E., et al., *Integrative analysis of 111 reference human epigenomes.* Nature, 2015. **518**(7539): p. 317-30.

309. Ben-Porath, I., et al., *An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors.* Nat Genet, 2008. **40**(5): p. 499-507.

310. Lawrence, M.S., et al., *Mutational heterogeneity in cancer and the search for new cancer-associated genes.* Nature, 2013. **499**(7457): p. 214-8.

311. Wright, W.E., O.M. Pereira-Smith, and J.W. Shay, *Reversible cellular senescence: implications for immortalization of normal human diploid fibroblasts.* Mol Cell Biol, 1989. **9**(7): p. 3088-92.

312. Romanov, S.R., et al., *Normal human mammary epithelial cells spontaneously escape senescence and acquire genomic changes.* Nature, 2001. **409**(6820): p. 633-7.

313. Hahn, W.C., et al., *Creation of human tumour cells with defined genetic elements.* Nature, 1999. **400**(6743): p. 464-8.

314. Ashley, D.J., *The two "hit" and multiple "hit" theories of carcinogenesis.* Br J Cancer, 1969. **23**(2): p. 313-28.

315. Nordling, C.O., *A new theory on cancer-inducing mechanism.* Br J Cancer, 1953. **7**(1): p. 68-72.

316. Nowell, P.C., *The clonal evolution of tumor cell populations.* Science, 1976. **194**(4260): p. 23-8.

317. Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma.* Proc Natl Acad Sci U S A, 1971. **68**(4): p. 820-3.

318. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.

319. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.

320. Shete, S., et al., *Genome-wide association study identifies five susceptibility loci for glioma.* Nat Genet, 2009. **41**(8): p. 899-904.

321. McGranahan, N. and C. Swanton, *Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future.* Cell, 2017. **168**(4): p. 613-628.

322. Yates, L.R. and P.J. Campbell, *Evolution of the cancer genome.* Nat Rev Genet, 2012. **13**(11): p. 795-806.

323. Johnson, B.E., et al., *Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma.* Science, 2014. **343**(6167): p. 189-193.

324. Bai, H., et al., *Integrated genomic characterization of IDH1-mutant glioma malignant progression.* Nat Genet, 2016. **48**(1): p. 59-66.

325. Lee, J.K., et al., *Spatiotemporal genomic architecture informs precision oncology in glioblastoma.* Nat Genet, 2017. **49**(4): p. 594-599.

326. Prensner, J.R. and A.M. Chinnaiyan, *Metabolism unhinged: IDH mutations in cancer.* Nat Med, 2011. **17**(3): p. 291-3.

327. Koivunen, P., et al., *Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation.* Nature, 2012. **483**(7390): p. 484-8.

328. Ward, P.S., 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, 2010. **17**(3): p. 225-34.

329. Dang, L., et al., *Cancer-associated IDH1 mutations produce 2-hydroxyglutarate.* Nature, 2009. **462**(7274): p. 739-44.

330. Figueroa, M.E., et al., *Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation.* Cancer Cell, 2010. **18**(6): p. 553-67.

331. Lenting, K., et al., *Glioma: experimental models and reality.* Acta Neuropathol, 2017. **133**(2): p. 263-282.

332. Cairns, R.A. and T.W. Mak, *Oncogenic isocitrate dehydrogenase mutations: mechanisms, models, and clinical opportunities.* Cancer Discov, 2013. **3**(7): p. 730-41.

333. Sasaki, M., et al., *D-2-hydroxyglutarate produced by mutant IDH1 perturbs collagen maturation and basement membrane function.* Genes Dev, 2012. **26**(18): p. 2038-49.

334. Bardella, C., et al., *Expression of Idh1(R132H) in the Murine Subventricular Zone Stem Cell Niche Recapitulates Features of Early Gliomagenesis.* Cancer Cell, 2016. **30**(4): p. 578-594.

335. Bigner, S.H., et al., *Specific chromosomal abnormalities in malignant human gliomas.* Cancer Res, 1988. **48**(2): p. 405-11.

336. Ichimura, K., et al., *Distinct patterns of deletion on 10p and 10q suggest involvement of multiple tumor suppressor genes in the development of astrocytic gliomas of different malignancy grades.* Genes Chromosomes Cancer, 1998. **22**(1): p. 9-15.

337. Gerstung, M., et al., *The evolutionary history of 2,658 cancers.* bioRxiv, 2017.

338. Tang, Y.C. and A. Amon, *Gene copy-number alterations: a cost-benefit analysis.* Cell, 2013. **152**(3): p. 394-405.

339. Martini, M., et al., *PI3K/AKT signaling pathway and cancer: an updated review.* Ann Med, 2014. **46**(6): p. 372-83.

340. Li, J., et al., *PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer.* Science, 1997. **275**(5308): p. 1943-7.

341. Steck, P.A., et al., *Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers.* Nat Genet, 1997. **15**(4): p. 356-62.

342. Pennisi, E., *New tumor suppressor found--twice.* Science, 1997. **275**(5308): p. 1876-8.

343. Wu, G., et al., *Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas.* Nat Genet, 2012. **44**(3): p. 251-3.

344. Korshunov, A., et al., *Histologically distinct neuroepithelial tumors with histone 3 G34 mutation are molecularly similar and comprise a single nosologic entity.* Acta Neuropathol, 2016. **131**(1): p. 137-46.

345. Bender, S., et al., *Reduced H3K27me3 and DNA hypomethylation are major drivers of gene expression in K27M mutant pediatric high-grade gliomas.* Cancer Cell, 2013. **24**(5): p. 660-72.

346. Lewis, P.W., et al., *Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma.* Science, 2013. **340**(6134): p. 857-61.

347. Castel, D., et al., *Histone H3F3A and HIST1H3B K27M mutations define two subgroups of diffuse intrinsic pontine gliomas with different prognosis and phenotypes.* Acta Neuropathol, 2015. **130**(6): p. 815-27.

348. Cantwell-Dorris, E.R., J.J. O'Leary, and O.M. Sheils, *BRAFV600E: implications for carcinogenesis and molecular therapy.* Mol Cancer Ther, 2011. **10**(3): p. 385-94.

349. Davies, H., et al., *Mutations of the BRAF gene in human cancer.* Nature, 2002. **417**(6892): p. 949-54.

350. Jones, D.T., et al., *Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas.* Cancer Res, 2008. **68**(21): p. 8673-7.

351. Forshew, T., et al., *Activation of the ERK/MAPK pathway: a signature genetic defect in posterior fossa pilocytic astrocytomas.* J Pathol, 2009. **218**(2): p. 172-81.

352. Sanoudou, D., et al., *Analysis of pilocytic astrocytoma by comparative genomic hybridization.* Br J Cancer, 2000. **82**(6): p. 1218-22.

353. Collado, M. and M. Serrano, *Senescence in tumours: evidence from mice and humans.* Nat Rev Cancer, 2010. **10**(1): p. 51-7.

354. Hayflick, L. and P.S. Moorhead, *The serial cultivation of human diploid cell strains.* Exp Cell Res, 1961. **25**: p. 585-621.

355. He, S. and N.E. Sharpless, *Senescence in Health and Disease.* Cell, 2017. **169**(6): p. 1000-1011.

356. Halazonetis, T.D., V.G. Gorgoulis, and J. Bartek, *An oncogene-induced DNA damage model for cancer development.* Science, 2008. **319**(5868): p. 1352-5.

357. Collado, M., M.A. Blasco, and M. Serrano, *Cellular senescence in cancer and aging.* Cell, 2007. **130**(2): p. 223-33.

358. Sharpless, N.E. and C.J. Sherr, *Forging a signature of in vivo senescence.* Nat Rev Cancer, 2015. **15**(7): p. 397-408.

359. Campisi, J. and F. d'Adda di Fagagna, *Cellular senescence: when bad things happen to good cells.* Nat Rev Mol Cell Biol, 2007. **8**(9): p. 729-40.

360. Perez-Mancera, P.A., A.R. Young, and M. Narita, *Inside and out: the activities of senescence in cancer.* Nat Rev Cancer, 2014. **14**(8): p. 547-58.

361. Shain, A.H., et al., *The Genetic Evolution of Melanoma from Precursor Lesions.* N Engl J Med, 2015. **373**(20): p. 1926-36.

362. Meeker, A.K., et al., *Telomere length abnormalities occur early in the initiation of epithelial carcinogenesis.* Clin Cancer Res, 2004. **10**(10): p. 3317-26.

363. Kuilman, T., et al., *Oncogene-induced senescence relayed by an interleukin-dependent inflammatory network.* Cell, 2008. **133**(6): p. 1019-31.

364. Gray-Schopfer, V.C., et al., *Cellular senescence in naevi and immortalisation in melanoma: a role for p16?* Br J Cancer, 2006. **95**(4): p. 496-505.

365. Michaloglou, C., et al., *BRAFE600-associated senescence-like cell cycle arrest of human naevi.* Nature, 2005. **436**(7051): p. 720-4.

366. Wright, W.E. and J.W. Shay, *The two-stage mechanism controlling cellular senescence and immortalization.* Exp Gerontol, 1992. **27**(4): p. 383-9.

367. Ciriello, G., et al., *Mutual exclusivity analysis identifies oncogenic network modules.* Genome Res, 2012. **22**(2): p. 398-406.

368. Ohgaki, H., et al., *Genetic pathways to glioblastoma: a population-based study.* Cancer Res, 2004. **64**(19): p. 6892-9.

369. Harris, S.L. and A.J. Levine, *The p53 pathway: positive and negative feedback loops.* Oncogene, 2005. **24**(17): p. 2899-908.

370. Bromberg, J.E. and M.J. van den Bent, *Oligodendrogliomas: molecular biology and treatment.* Oncologist, 2009. **14**(2): p. 155-63.

371. Jenkins, R.B., et al., *A t(1;19)(q10;p10) mediates the combined deletions of 1p and 19q and predicts a better prognosis of patients with oligodendroglioma.* Cancer Res, 2006. **66**(20): p. 9852-61.

372. Jeuken, J.W., et al., *The nature and timing of specific copy number changes in the course of molecular progression in diffuse gliomas: further elucidation of their genetic "life story".* Brain Pathol, 2011. **21**(3): p. 308-20.

373. van Thuijl, H.F., et al., *Spatial and temporal evolution of distal 10q deletion, a prognostically unfavorable event in diffuse low-grade gliomas.* Genome Biol, 2014. **15**(9): p. 471.

374. Aihara, K., et al., *Genetic and epigenetic stability of oligodendrogliomas at recurrence.* Acta Neuropathol Commun, 2017. **5**(1): p. 18.

375. Azar, S., et al., *Cellular and molecular characterization of IDH1-mutated diffuse low grade gliomas reveals tumor heterogeneity and absence of EGFR/PDGFRalpha activation.* Glia, 2018. **66**(2): p. 239-255.

376. Wilcox, P., et al., *Oligoastrocytomas: throwing the baby out with the bathwater?* Acta Neuropathol, 2015. **129**(1): p. 147-9.

377. Huse, J.T., et al., *Mixed glioma with molecular features of composite oligodendroglioma and astrocytoma: a true "oligoastrocytoma"?* Acta Neuropathol, 2015. **129**(1): p. 151-3.

378. Aldape, K., et al., *Glioma Through the Looking GLASS: Molecular Evolution of Diffuse Gliomas and the Glioma Longitudinal AnalySiS Consortium.* Neuro Oncol, 2018.

379. Hu, X., et al., *Multigene signature for predicting prognosis of patients with 1p19q co-deletion diffuse glioma.* Neuro Oncol, 2017. **19**(6): p. 786-795.

380. Chan, A.K., et al., *Loss of CIC and FUBP1 expressions are potential markers of shorter time to recurrence in oligodendroglial tumors.* Mod Pathol, 2014. **27**(3): p. 332-42.

381. Solimini, N.L., et al., *Recurrent hemizygous deletions in cancers may optimize proliferative potential.* Science, 2012. **337**(6090): p. 104-9.

382. Bagchi, A. and A.A. Mills, *The quest for the 1p36 tumor suppressor.* Cancer Res, 2008. **68**(8): p. 2551-6.

383. Vestin, A. and A.A. Mills, *The tumor suppressor Chd5 is induced during neuronal differentiation in the developing mouse brain.* Gene Expr Patterns, 2013. **13**(8): p. 482-9.

384. Bagchi, A., et al., *CHD5 is a tumor suppressor at human 1p36.* Cell, 2007. **128**(3): p. 459-75.

385. Liggett, W.H., Jr. and D. Sidransky, *Role of the p16 tumor suppressor gene in cancer.* J Clin Oncol, 1998. **16**(3): p. 1197-206.

386. Uhrbom, L., et al., *Ink4a-Arf loss cooperates with KRas activation in astrocytes and neural progenitors to generate glioblastomas of various morphologies depending on activated Akt.* Cancer Res, 2002. **62**(19): p. 5551-8.

387. Uhrbom, L., M. Nister, and B. Westermark, *Induction of senescence in human malignant glioma cells by p16INK4A.* Oncogene, 1997. **15**(5): p. 505-14.

388. Holland, E.C., et al., *Modeling mutations in the G1 arrest pathway in human gliomas: overexpression of CDK4 but not loss of INK4a-ARF induces hyperploidy in cultured mouse astrocytes.* Genes Dev, 1998. **12**(23): p. 3644-9.

389. Ohgaki, H. and P. Kleihues, *Genetic pathways to primary and secondary glioblastoma.* Am J Pathol, 2007. **170**(5): p. 1445-53.

390. Buczkowicz, P., et al., *Genomic analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent activating ACVR1 mutations.* Nat Genet, 2014. **46**(5): p. 451-6.

391. Taylor, K.R., et al., *Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma.* Nat Genet, 2014. **46**(5): p. 457-461.

392. Fontebasso, A.M., et al., *Recurrent somatic mutations in ACVR1 in pediatric midline high-grade astrocytoma.* Nat Genet, 2014. **46**(5): p. 462-6.

393. Wu, G., et al., *The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma.* Nat Genet, 2014. **46**(5): p. 444-450.

394. Fiscella, M., et al., *Wip1, a novel human protein phosphatase that is induced in response to ionizing radiation in a p53-dependent manner.* Proc Natl Acad Sci U S A, 1997. **94**(12): p. 6048-53.

395. Raabe, E.H., et al., *BRAF activation induces transformation and then senescence in human neural stem cells: a pilocytic astrocytoma model.* Clin Cancer Res, 2011. **17**(11): p. 3590-9.

396. Horbinski, C., *To BRAF or not to BRAF: is that even a question anymore?* J Neuropathol Exp Neurol, 2013. **72**(1): p. 2-7.

397. Koelsche, C., et al., *Mutant BRAF V600E protein in ganglioglioma is predominantly expressed by neuronal tumor cells.* Acta Neuropathol, 2013. **125**(6): p. 891-900.

398. Zhang, J., et al., *Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas.* Nat Genet, 2013. **45**(6): p. 602-12.

399. Turcan, S. and T.A. Chan, *MAPping the genomic landscape of low-grade pediatric gliomas.* Nat Genet, 2013. **45**(8): p. 847-9.

400. Rozen, W.M., S. Joseph, and P.A. Lo, *Spontaneous regression of low-grade gliomas in pediatric patients without neurofibromatosis.* Pediatr Neurosurg, 2008. **44**(4): p. 324-8.

401. Burkhard, C., et al., *A population-based study of the incidence and survival rates in patients with pilocytic astrocytoma.* J Neurosurg, 2003. **98**(6): p. 1170-4.

402. Gunny, R.S., et al., *Spontaneous regression of residual low-grade cerebellar pilocytic astrocytomas in children.* Pediatr Radiol, 2005. **35**(11): p. 1086-91.

403. Buder, T., et al., *Model-Based Evaluation of Spontaneous Tumor Regression in Pilocytic Astrocytoma.* PLoS Comput Biol, 2015. **11**(12): p. e1004662.

404. Robinson, J.P., et al., *Activated BRAF induces gliomas in mice when combined with Ink4a/Arf loss or Akt activation.* Oncogene, 2010. **29**(3): p. 335-44.

405. Huillard, E., et al., *Cooperative interactions of BRAFV600E kinase and CDKN2A locus deficiency in pediatric malignant astrocytoma as a basis for rational therapy.* Proc Natl Acad Sci U S A, 2012. **109**(22): p. 8710-5.

406. Schiffman, J.D., et al., *Oncogenic BRAF mutation with CDKN2A inactivation is characteristic of a subset of pediatric malignant astrocytomas.* Cancer Res, 2010. **70**(2): p. 512-9.

407. McClintock, B., *The Production of Homozygous Deficient Tissues with Mutant Characteristics by Means of the Aberrant Mitotic Behavior of Ring-Shaped Chromosomes.* Genetics, 1938. **23**(4): p. 315-76.

408. McClintock, B., *The Stability of Broken Ends of Chromosomes in Zea Mays.* Genetics, 1941. **26**(2): p. 234-82.

409. Gisselsson, D., et al., *Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity.* Proc Natl Acad Sci U S A, 2000. **97**(10): p. 5357-62.

410. Artandi, S.E., et al., *Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice.* Nature, 2000. **406**(6796): p. 641-5.

411. Maciejowski, J., et al., *Chromothripsis and Kataegis Induced by Telomere Crisis.* Cell, 2015. **163**(7): p. 1641-54.

412. Ernst, A., et al., *Telomere dysfunction and chromothripsis.* Int J Cancer, 2016. **138**(12): p. 2905-14.

413. Maciejowski, J. and T. de Lange, *Telomeres in cancer: tumour suppression and genome instability.* Nat Rev Mol Cell Biol, 2017. **18**(3): p. 175-186.

414. Malhotra, A., et al., *Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms.* Genome Res, 2013. **23**(5): p. 762-76.

415. deCarvalho, A.C., et al., *Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma.* bioRxiv, 2017.

416. Turner, K.M., et al., *Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity.* Nature, 2017. **543**(7643): p. 122-125.

417. Cox, D., C. Yuncken, and A.I. Spriggs, *Minute Chromatin Bodies in Malignant Tumours of Childhood.* Lancet, 1965. **1**(7402): p. 55-8.

418. Kohl, N.E., et al., *Transposition and amplification of oncogene-related sequences in human neuroblastomas.* Cell, 1983. **35**(2 Pt 1): p. 359-67.

419. Sanborn, J.Z., et al., *Double minute chromosomes in glioblastoma multiforme are revealed by precise reconstruction of oncogenic amplicons.* Cancer Res, 2013. **73**(19): p. 6036-45.

420. Nikolaev, S., et al., *Extrachromosomal driver mutations in glioblastoma and low-grade glioma.* Nat Commun, 2014. **5**: p. 5690.

421. Kanda, T., K.F. Sullivan, and G.M. Wahl, *Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells.* Curr Biol, 1998. **8**(7): p. 377-85.

422. Tanaka, H. and M.C. Yao, *Palindromic gene amplification--an evolutionarily conserved role for DNA inverted repeats in the genome.* Nat Rev Cancer, 2009. **9**(3): p. 216-24.

423. Forment, J.V., A. Kaidi, and S.P. Jackson, *Chromothripsis and cancer: causes and consequences of chromosome shattering.* Nat Rev Cancer, 2012. **12**(10): p. 663-70.

424. Mazor, T., et al., *Clonal expansion and epigenetic reprogramming following deletion or amplification of mutant IDH1.* Proc Natl Acad Sci U S A, 2017. **114**(40): p. 10743-10748.

425. Luchman, H.A., et al., *Spontaneous loss of heterozygosity leading to homozygous R132H in a patient-derived IDH1 mutant cell line.* Neuro Oncol, 2013. **15**(8): p. 979-80.

426. Johannessen, T.A., et al., *Rapid Conversion of Mutant IDH1 from Driver to Passenger in a Model of Human Gliomagenesis.* Mol Cancer Res, 2016. **14**(10): p. 976-983.

427. Counter, C.M., et al., *Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity.* EMBO J, 1992. **11**(5): p. 1921-9.

428. Garbe, J.C., et al., *Immortalization of normal human mammary epithelial cells in two steps by direct targeting of senescence barriers does not require gross genomic alterations.* Cell Cycle, 2014. **13**(21): p. 3423-35.

429. Kim, N.W., et al., *Specific association of human telomerase activity with immortal cells and cancer.* Science, 1994. **266**(5193): p. 2011-5.

430. Dunham, M.A., et al., *Telomere maintenance by recombination in human cells.* Nat Genet, 2000. **26**(4): p. 447-50.

431. Bryan, T.M., et al., *Telomere elongation in immortal human cells without detectable telomerase activity.* EMBO J, 1995. **14**(17): p. 4240-8.

432. Kannan, K., et al., *Whole-exome sequencing identifies ATRX mutation as a key molecular determinant in lower-grade glioma.* Oncotarget, 2012. **3**(10): p. 1194-203.

433. Goldberg, A.D., et al., *Distinct factors control histone variant H3.3 localization at specific genomic regions.* Cell, 2010. **140**(5): p. 678-91.

434. Parkinson, G.N., M.P. Lee, and S. Neidle, *Crystal structure of parallel quadruplexes from human telomeric DNA.* Nature, 2002. **417**(6891): p. 876-80.

435. Law, M.J., et al., *ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner.* Cell, 2010. **143**(3): p. 367-78.

436. Lazzerini-Denchi, E. and A. Sfeir, *Stop pulling my strings - what telomeres taught us about the DNA damage response.* Nat Rev Mol Cell Biol, 2016. **17**(6): p. 364-78.

437. Lovejoy, C.A., et al., *Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway.* PLoS Genet, 2012. **8**(7): p. e1002772.

438. Napier, C.E., et al., *ATRX represses alternative lengthening of telomeres.* Oncotarget, 2015. **6**(18): p. 16543-58.

439. Bell, R.J., et al., *Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer.* Science, 2015. **348**(6238): p. 1036-9.

440. Martinez, P. and M.A. Blasco, *Telomeric and extra-telomeric roles for telomerase and the telomere-binding proteins.* Nat Rev Cancer, 2011. **11**(3): p. 161-76.

441. Park, J.I., et al., *Telomerase modulates Wnt signalling by association with target gene chromatin.* Nature, 2009. **460**(7251): p. 66-72.

442. Chiba, K., et al., *Cancer-associated TERT promoter mutations abrogate telomerase silencing.* Elife, 2015. **4**.

443. Chiba, K., et al., *Mutations in the promoter of the telomerase gene TERT contribute to tumorigenesis by a two-step mechanism.* Science, 2017. **357**(6358): p. 1416-1420.

444. Maryoung, L., et al., *Somatic mutations in telomerase promoter counterbalance germline loss-of-function mutations.* J Clin Invest, 2017. **127**(3): p. 982-986.

445. Langford, L.A., et al., *Telomerase activity in human brain tumours.* Lancet, 1995. **346**(8985): p. 1267-8.

446. Pekmezci, M., et al., *Adult infiltrating gliomas with WHO 2016 integrated diagnosis: additional prognostic roles of ATRX and TERT.* Acta Neuropathol, 2017. **133**(6): p. 1001-1016.

447. Barthel, F.P., et al., *Systematic analysis of telomere length and somatic alterations in 31 cancer types.* Nat Genet, 2017. **49**(3): p. 349-357.

448. Mangerel, J., et al., *Alternative lengthening of telomeres is enriched in, and impacts survival of TP53 mutant pediatric malignant brain tumors.* Acta Neuropathol, 2014. **128**(6): p. 853-62.

449. Lee, J., D.A. Solomon, and T. Tihan, *The role of histone modifications and telomere alterations in the pathogenesis of diffuse gliomas in adults and children.* J Neurooncol, 2017. **132**(1): p. 1-11.

450. Vescovi, A.L., R. Galli, and B.A. Reynolds, *Brain tumour stem cells.* Nat Rev Cancer, 2006. **6**(6): p. 425-36.

451. Greaves, M. and C.C. Maley, *Clonal evolution in cancer.* Nature, 2012. **481**(7381): p. 306-13.

452. Tirosh, I., et al., *Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma.* Nature, 2016. **539**(7628): p. 309-313.

453. Lan, X., et al., *Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy.* Nature, 2017. **549**(7671): p. 227-232.

454. Kimura, M., *The neutral theory of molecular evolution: a review of recent evidence.* Jpn J Genet, 1991. **66**(4): p. 367-86.

455. Noorbakhsh, J. and J.H. Chuang, *Uncertainties in tumor allele frequencies limit power to infer evolutionary pressures.* Nat Genet, 2017. **49**(9): p. 1288-1289.

456. Tarabichi, M., et al., *Neutral tumor evolution?* bioRxiv, 2017.

457. Williams, M.J., et al., *Reply: Uncertainties in tumor allele frequencies limit power to infer evolutionary pressures.* Nat Genet, 2017. **49**(9): p. 1289-1291.

458. Ouchi, R., et al., *Senescence from glioma stem cell differentiation promotes tumor growth.* Biochem Biophys Res Commun, 2016. **470**(2): p. 275-281.

459. Coppe, J.P., et al., *The senescence-associated secretory phenotype: the dark side of tumor suppression.* Annu Rev Pathol, 2010. **5**: p. 99-118.

460. Coppe, J.P., et al., *Tumor suppressor and aging biomarker p16(INK4a) induces cellular senescence without the associated inflammatory secretory phenotype.* J Biol Chem, 2011. **286**(42): p. 36396-403.

461. Coppe, J.P., et al., *Senescence-associated secretory phenotypes reveal cell-nonautonomous functions of oncogenic RAS and the p53 tumor suppressor.* PLoS Biol, 2008. **6**(12): p. 2853-68.

462. Raj, N. and L.D. Attardi, *Tumor suppression: p53 alters immune surveillance to restrain liver cancer.* Curr Biol, 2013. **23**(12): p. R527-30.

463. Lujambio, A., et al., *Non-cell-autonomous tumor suppression by p53.* Cell, 2013. **153**(2): p. 449-60.

464. Coppola, D., et al., *Senescence-associated-gene signature identifies genes linked to age, prognosis, and progression of human gliomas.* J Geriatr Oncol, 2014. **5**(4): p. 389-99.

465. Kumar, R., et al., *Induction of senescence in primary glioblastoma cells by serum and TGFbeta.* Sci Rep, 2017. **7**(1): p. 2156.

466. Melin, B.S., et al., *Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors.* Nat Genet, 2017. **49**(5): p. 789-794.

467. Wrensch, M., et al., *Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility.* Nat Genet, 2009. **41**(8): p. 905-8.

468. Rajaraman, P., et al., *Genome-wide association study of glioma and meta-analysis.* Hum Genet, 2012. **131**(12): p. 1877-88.

469. Jenkins, R.B., et al., *A low-frequency variant at 8q24.21 is strongly associated with risk of oligodendroglial tumors and astrocytomas with IDH1 or IDH2 mutation.* Nat Genet, 2012. **44**(10): p. 1122-5.

470. Kinnersley, B., et al., *Genome-wide association study identifies multiple susceptibility loci for glioma.* Nat Commun, 2015. **6**: p. 8559.

471. Connelly, J.M. and M.G. Malkin, *Environmental risk factors for brain tumors.* Curr Neurol Neurosci Rep, 2007. **7**(3): p. 208-14.

472. Codd, V., et al., *Identification of seven loci affecting mean telomere length and their association with disease.* Nat Genet, 2013. **45**(4): p. 422-7, 427e1-2.

473. Walsh, K.M., et al., *Longer genotypically-estimated leukocyte telomere length is associated with increased adult glioma risk.* Oncotarget, 2015. **6**(40): p. 42468-77.

474. Walsh, K.M., et al., *Telomere maintenance and the etiology of adult glioma.* Neuro Oncol, 2015. **17**(11): p. 1445-52.

475. Walsh, K.M., et al., *Variants near TERT and TERC influencing telomere length are associated with high-grade glioma risk.* Nat Genet, 2014. **46**(7): p. 731-5.

476. Telomeres Mendelian Randomization, C., et al., *Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study.* JAMA Oncol, 2017. **3**(5): p. 636-651.

477. Blasco, M.A., *Telomere length, stem cells and aging.* Nat Chem Biol, 2007. **3**(10): p. 640-9.

478. Stone, R.C., et al., *Telomere Length and the Cancer-Atherosclerosis Trade-Off.* PLoS Genet, 2016. **12**(7): p. e1006144.

479. Gutmann, D.H., et al., *The diagnostic evaluation and multidisciplinary management of neurofibromatosis 1 and neurofibromatosis 2.* JAMA, 1997. **278**(1): p. 51-7.

480. Cawthon, R.M., et al., *A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations.* Cell, 1990. **62**(1): p. 193-201.

481. Rouleau, G.A., et al., *Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2.* Nature, 1993. **363**(6429): p. 515-21.

482. Cichowski, K., et al., *Mouse models of tumor development in neurofibromatosis type 1.* Science, 1999. **286**(5447): p. 2172-6.

483. Courtois-Cox, S., et al., *A negative feedback signaling network underlies oncogene-induced senescence.* Cancer Cell, 2006. **10**(6): p. 459-72.

484. Bell, D.W., et al., *Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome.* Science, 1999. **286**(5449): p. 2528-31.

485. McBride, K.A., et al., *Li-Fraumeni syndrome: cancer risk assessment and clinical management.* Nat Rev Clin Oncol, 2014. **11**(5): p. 260-71.

486. Birch, J.M., et al., *Prevalence and diversity of constitutional mutations in the p53 gene among 21 Li-Fraumeni families.* Cancer Res, 1994. **54**(5): p. 1298-304.

487. Birch, J.M., et al., *Relative frequency and morphology of cancers in carriers of germline TP53 mutations.* Oncogene, 2001. **20**(34): p. 4621-8.

488. Hisada, M., et al., *Multiple primary cancers in families with Li-Fraumeni syndrome.* J Natl Cancer Inst, 1998. **90**(8): p. 606-11.

489. Capper, D., et al., *Monoclonal antibody specific for IDH1 R132H mutation.* Acta Neuropathol, 2009. **118**(5): p. 599-601.

490. Priesterbach-Ackley, L.P., et al., *Molecular tools for the pathologic diagnosis of central nervous system tumors.* Neurooncol Pract, 2019. **6**(1): p. 4-16.

491. Jaunmuktane, Z., et al., *Methylation array profiling of adult brain tumours: diagnostic outcomes in a large, single centre.* Acta Neuropathol Commun, 2019. **7**(1): p. 24.

492. van den Bent, M.J., et al., *Randomized phase II trial of erlotinib versus temozolomide or carmustine in recurrent glioblastoma: EORTC brain tumor group study 26034.* J Clin Oncol, 2009. **27**(8): p. 1268-74.

493. Weller, M., et al., *Rindopepimut with temozolomide for patients with newly diagnosed, EGFRvIII-expressing glioblastoma (ACT IV): a randomised, double-blind, international phase 3 trial.* Lancet Oncol, 2017. **18**(10): p. 1373-1385.

494. Lassman, A., et al., *ACTR-21. A Randomized, Double-Blind, Placebo-Controlled Phase 3 Trial of Depatuxizumab Mafodotin (Abt-414) in Epidermal Growth Factor Receptor (EGFR) Amplified (amp) Newly Diagnosed Glioblastoma (NGBM).* Neuro-Oncology, 2019. **Volume 21**(Issue Supplement_6).

495. Drilon, A., *TRK inhibitors in TRK fusion-positive cancers.* Ann Oncol, 2019. **30 Suppl 8**: p. viii23-viii30.

496. Kaley, T., et al., *BRAF Inhibition in BRAF(V600)-Mutant Gliomas: Results From the VE-BASKET Study.* J Clin Oncol, 2018. **36**(35): p. 3477-3484.

497. Lasorella, A., M. Sanson, and A. Iavarone, *FGFR-TACC gene fusions in human glioma.* Neuro Oncol, 2017. **19**(4): p. 475-483.

498. Wen, P., et al., *ACTR-30. Updated Efficacy and Safety of Dabrafenib Plus Trametinib in Patients with Recurrent/Refractory BRAF V600E–mutated High-Grade Glioma (HGG) And Low-Grade Glioma (LGG).* Neuro-Oncology, 2019. **Volume 21**(Issue Supplement_6): p. Pages vi19–vi20.

499. Behling, F. and J. Schittenhelm, *Oncogenic BRAF Alterations and Their Role in Brain Tumors.* Cancers (Basel), 2019. **11**(6).

500. Xu, T., et al., *Gene Fusion in Malignant Glioma: An Emerging Target for Next-Generation Personalized Treatment.* Transl Oncol, 2018. **11**(3): p. 609-618.

501. Francis, J.M., et al., *EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing.* Cancer Discov, 2014. **4**(8): p. 956-71.

502. Amirouchene-Angelozzi, N., C. Swanton, and A. Bardelli, *Tumor Evolution as a Therapeutic Target.* Cancer Discov, 2017.

503. Karpel-Massler, G., et al., *Novel IDH1-Targeted Glioma Therapies.* CNS Drugs, 2019. **33**(12): p. 1155-1166.

504. Schumacher, T., et al., *Mutant IDH1: An immunotherapeutic target in tumors.* Oncoimmunology, 2014. **3**(12): p. e974392.

505. Platten, M., et al., *A vaccine targeting mutant IDH1 in newly diagnosed glioma.* Nature, 2021.

506. PC, N. and H. DA, *A minute chromosome in human chronic granulocytic leukemia.* Science, 1960(132).

507. Wong, S. and O.N. Witte, *The BCR-ABL story: bench to bedside and back.* Annu Rev Immunol, 2004. **22**: p. 247-306.

508. Druker, B.J., et al., *Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia.* N Engl J Med, 2001. **344**(14): p. 1031-7.

509. Kantarjian, H., et al., *Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia.* N Engl J Med, 2002. **346**(9): p. 645-52.

Nederlandse samenvatting

Diffuse gliomen betreffen een diverse groep tumoren van het centraal zenuwstelsel, met name van de hersenen. Ondanks behandeling met chirurgie, radio- en chemotherapie is de prognose slecht en overlijdt de patiënt meestal binnen een paar maanden tot enkele jaren. Al langer dan een eeuw is de geaccepteerde gouden standaard voor de klinische diagnose en bepaling van de vervolgbehandeling gebaseerd op histologische evaluatie van de tumor. Dit proefschrift omvat een aantal studies waarin de moleculaire analyse van diffuse gliomen centraal staat. Deze maken deel uit van een grote reeks publicaties in het laatste decennium waaruit blijkt dat moleculaire diagnostiek ons in staat stelt tot een veel meer eenduidige bepaling van het behandelingstraject van patienten met een diffuus glioom.

Ter inleiding biedt **hoofdstuk 1** een uitgebreide samenvatting van de historische aspecten van met name de histopathologische diagnostiek van gliomen. In **hoofdstuk 2** bespreek ik een moleculaire indeling van gliomen. We weten door middel van de analyse van epigenetische aanpassingen van het DNA (zogeheten DNA methylatie aanpassingen) zeven verschillende tumor subtypen te onder-

scheiden in een set van zeker 1.100 primaire gliomen. Uit onze analyse bleek dat deze subtypen niet zomaar overlappen met de histologische subtypen. Belangrijk daarbij is dat de via DNA methylatie bepaalde subtypen nauwer geassocieerd bleken te zijn met het biologische gedrag van de tumor en zo dus een betere voorspelling van de prognose van de patiënt mogelijk maakten.

In **hoofdstuk 3** bestudeer ik recidief groei van diffuse gliomen. Ondanks (huidige) optimale behandeling keert deze tumor in nagenoeg alle gevallen terug en veroorzaakt het uiteindelijk het overlijden van de patiënt. Om hier meer grip op te krijgen hebben wij weefsel van zowel primaire als recidief tumoren verzameld van ruim 200 patiënten. In recidief tumoren konden wij over het algemeen meer DNA-mutaties opsporen dan in primaire tumoren, en soms was er sprake van hypermutatie ten gevolge van behandeling met alkylerende middelen. In recidiverende IDH-gemuteerde tumoren troffen wij in een aantal gevallen nieuw ontstane deleties aan van het gen *CDKN2A*, wat inmiddels een criterium is voor een hogere graad tumor. Ondanks dat er sprake was van een hogere mutatie last en zeldzame *CDKN2A* deleties, kwamen recidief tumoren moleculair bekeken nauw overeen met de bijbehorende primaire tumor. Dit suggereert dat, ondanks dat de huidige behandelingen de groei van de tumor vertragen, deze therapieën de tumor onvoldoende genetische uitdaging bieden. Immers, het blijkt dat onder druk van de huidige behandelingen het voor gecontinueerde groei van een diffuus glioom niet noodzakelijk is om zich significant genetisch verder te ontwikkelen.

In **hoofdstuk 4** ga ik in op de vraag of DNA methylatie subtypen (eerder in **hoofdstuk 2** bepaald) consistent zijn over het gehele volume van de tumor. In een set van 16 patiënten bekeken we 8-12 verschillende sectoren per tumor en constateerden we dat DNA methylatie subtypen over het algemeen homogeen zijn. Dit is een belangrijk resultaat omdat dit aangeeft dat moleculaire DNA methylatie diagnostiek van diffuse gliomen robuust is. Dit is in tegenstelling tot histologische diagnose waarbij er regelmatig sprake is van onderwaardering van de maligniteitsgraad vanwege 'sampling effect' alsmede van inter-observer variatie bij beoordeling door verschillende (neuro)pathologen.

Hoofdstuk 5 vormt een logisch vervolg op **hoofdstuk 3** en gaat dieper in op de vraag hoe DNA-mutaties verschillen tussen primaire en recidief-tumoren en wat de invloed van radiotherapie hierop is. Hiervoor hebben we de gliomen dataset uit **hoofdstuk 3** aangevuld met ruim 3.500 patiënten met metastatische tumoren van de Hartwig Medical Foundation, waarbij bekend was of de patiënt voorbehandeld was met radiotherapie of niet. We stelden vast dat behandeling met

radiotherapie geassocieerd was met een toename van kleine en grote deleties. Op basis van de omringende nucleotidesequentie bepaalden we dat DNA-reparatie door middel van de klassieke niet-homologe DNA-reparatie waarschijnlijk ten grondslag ligt van de reparatie van radiatie-geïnduceerde DNA-schade. In gliomen was de acquisitie van nieuwe *CDKN2A* deleties in recidief tumoren zeer sterk geassocieerd met status na radiotherapie, wat een mogelijk oorzakelijk verband suggereert.

Volgend op resultaten uit **hoofdstuk 2** bespreek ik in **hoofdstuk 6** verschillen in onderhoudsmechanismen van de uiteinden van het DNA, zogeheten telomeren, in een dataset van bijna 7.000 patiënten en 31 verschillende tumor typen. Door middel van de analyse van DNA-sequentie gegevens was het mogelijk om de gemiddelde telomeer-lengte van tumoren in te schatten. Vervolgens konden we kijken naar de expressie van het gen *TERT*, dat voor het enzymatische onderdeel van het eiwit telomerase codeert dat verlenging van telomeren bewerkstelligt. We constateerden een aantal genetische en epigenetische afwijkingen die sterk correleerden met een toename van *TERT*. In een klein aantal tumoren leek er ook sprake van *TERT*-onafhankelijke onderhoud van telomeren. Bij deze patiënten was waarschijnlijk een alternatief, telomerase-onafhankelijk mechanisme verantwoordelijk voor telomeer-onderhoud. Bij sommige tumoren leek er geen sprake van telomeer-onderhoud, deze tumoren werden vaak gekenmerkt door (relatief) indolent gedrag. Samengevat gaven mijn analyses de universele rol van telomeer-onderhoud in kanker aan en benadrukte het dit mechanisme als een veelbelovend therapeutisch aangrijpingspunt.

Ter afsluiting van dit proefschrift destilleer ik in **hoofdstuk 7** de kennis uit de voorafgaande hoofdstukken en verwerk ik die in een model voor de moleculaire ontwikkeling van gliomen. Hierbij identificeerden we drie moleculaire eigenschappen die elk glioom kenmerkt, te weten een toename van proliferatie, een ononderbroken celcyclus en telomeer-onderhoud. Tot slot vat ik in **hoofdstuk 8** de voorgaande hoofdstukken samen en bediscussieer ik de rol van moleculaire diagnostiek in de kliniek.

List of Publications

** denotes first authorship, † denotes corresponding authorship*

*Varn FS, Johnson KC, Martinek J, Huse JT, Nasrallah MP, Wesseling P, Cooper LAD, Malta TM, Wade TE, Sabedot TS, Brat D, Gould PV, Wöehrer A, Aldape K, Ismail A, Sivajothi SK, **Barthel FP**, et al., Verhaak RGW; GLASS Consortium. Glioma progression is shaped by genetic evolution and microenvironment interactions. Cell. 2022 Jun 9;185(12):2184-2199.e16. doi: 10.1016/j.cell.2022.04.038. PMID: 35649412.*

*Johnson KC, Anderson KJ, Courtois ET, Gujar AD, **Barthel FP**, Varn FS, et al., Verhaak RGW. Single-cell multimodal glioma analyses identify epigenetic regulators of cellular plasticity and environmental stress response. **Nat Genet**. 2021 Oct;53(10):1456-1468. doi: 10.1038/s41588-021-00926-8.*

*Verburg N, **Barthel FP***, Anderson KJ, Johnson KC, et al., Wesseling P, De Witt Hamer PC, Verhaak RGW. Spatial concordance of DNA methylation classification in diffuse glioma. Neuro Oncol. 2021 Dec 1;23(12):2054-2065. doi: 10.1093/neuonc/noab134. PMID: 34049406; PMCID: PMC8643482.*

*Kocakavuk E, Anderson KJ, Varn FS, Johnson KC, Amin SB, Sulman EP, Lolkema M, **Barthel FP**†, Verhaak RGW. Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer.* ***Nat Genet****. 2021. doi 10.1038/s41588-021-00874-3*

*Wen PY, Weller M, Lee EQ, Alexander BA, Barnholtz-Sloan JS, **Barthel FP**, et al. Glioblastoma in Adults: A Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) Consensus Review on Current Management and Future Directions.* ***Neuro-Oncology****. 24 April 2020. doi: 10.1093/neuonc/noaa106*

*Amin SB, Anderson KJ, Boudreau CE, Martinez-Ledesma E, Kocakavuk E, Johnson KC, **Barthel FP**, et al. Comparative Molecular Life History of Spontaneous Canine and Human Gliomas.* ***Cancer Cell****. 2020 Feb 10;37(2):243-257.e7. doi: 10.1016/j.ccell.2020.01.004. PubMed PMID: 32049048.*

*Anderson KJ, Tan AC, Parkinson J, Back M, Kastelan M, Newey A, Brewer J, Wheeler H, Hudson AL, Amin SB, Johnson KC, **Barthel FP**, Verhaak RGW, Khasraw M. Molecular and clonal evolution in recurrent metastatic gliosarcoma.* ***Cold Spring Harb Mol Case Stud****. 2020 Jan 2. pii: mcs.a004671. doi: 10.1101/mcs.a004671. PubMed PMID: 31896544.*

Barthel FP**, Johnson KC*, Varn FS, Moskalik AD, Tanner G, The GLASS Consortium, Huse JT, Groot JF de, Stead LF, Verhaak RGW. Longitudinal Molecular Trajectories of Diffuse Glioma in Adults.* ***Nature.*** *2019 Dec;576(7785):112-120. doi: 10.1038/s41586-019-1775-1. PubMed PMID: 31748746.*

*Menghi F, **Barthel FP**, Yadav V, Tang M, Ji B, Tang Z, Carter GW, Ruan Y, Scully R, Verhaak RGW, Jonkers J, Liu ET. The Tandem Duplicator Phenotype Is a Prevalent Genome-Wide Cancer Configuration Driven by Distinct Gene Mutations.* ***Cancer Cell.*** *2018 Aug 13;34(2):197-210.e5. doi: 10.1016/j.ccell.2018.06.008. Epub 2018 Jul 12. PubMed PMID: 30017478.*

Barthel FP*, Johnson KC, Wesseling P, Verhaak RGW. Evolving Insights into the Molecular Neuropathology of Diffuse Gliomas in Adults.* ***Neurol Clin.*** *2018 Aug;36(3):421-437. doi: 10.1016/j.ncl.2018.04.002. PubMed PMID: 30072063.*

***Barthel FP**†, Wesseling P, Verhaak RGW. Reconstructing the molecular life history of gliomas. **Acta Neuropathol.** 2018 May;135(5):649-670. doi: 10.1007/s00401-018-1842-y. Epub 2018 Apr 3. PubMed PMID: 29616301.*

*Hu X, Wang Q, Tang M, **Barthel FP**, Amin S, Yoshihara K, Lang FM, Martinez-Ledesma E, Lee SH, Zheng S, Verhaak RGW. TumorFusions: an integrative resource for cancer-associated transcript fusions. **Nucleic Acids Res.** 2018 Jan 4;46(D1):D1144-D1149. doi: 10.1093/nar/gkx1018. PubMed PMID: 29099951.*

*Hu X, Martinez-Ledesma E, Zheng S, Kim H, **Barthel FP**, Jiang T, Hess KR, Verhaak RGW. Multigene signature for predicting prognosis of patients with 1p19q co-deletion diffuse glioma. **Neuro Oncol.** 2017 Jun 1;19(6):786-795. doi: 10.1093/neuonc/now285. PubMed PMID: 28340142.*

*Cancer Genome Atlas Research Network. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. **Cell.** 2017 Nov 2;171(4):950-965.e28. doi: 10.1016/j.cell.2017.10.014. PubMed PMID: 29100075.*

***Barthel FP**, Wei W, Tang M, Martinez-Ledesma E, Hu X, Amin SB, Akdemir KC, Seth S, Song X, Wang Q, Lichtenberg T, Hu J, Zhang J, Zheng S, Verhaak RG. Systematic analysis of telomere length and somatic alterations in 31 cancer types. **Nat Genet.** 2017 Mar;49(3):349-357. doi: 10.1038/ng.3781. Epub 2017 Jan 30. PubMed PMID: 28135248.*

*Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L, deCarvalho AC, Lyu S, Li P, Li Y, **Barthel FP**, et al., Heimberger AB, Sulman EP, Nam DH, Verhaak RGW. Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. **Cancer Cell**. 2018 Jan 8;33(1):152. doi: 10.1016/j.ccell.2017.12.012. PubMed PMID: 29316430.*

Ceccarelli M, **Barthel FP***, Malta TM*, Sabedot TS*, TCGA Research Network, Noushmehr H, Iavarone A, Verhaak RG. Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. **Cell**. 2016 Jan 28;164(3):550-63. doi: 10.1016/j.cell.2015.12.028. PubMed PMID: 26824661*

Epilogue

When I first came to the US in February 2014 to join Roel Verhaak's lab at MD Anderson, I had no idea that more than 8 years down the road I'd still be in this country. My naive initial plan was simple: stay for a short 6-month research internship, publish paper(s) to build my CV and return to the Netherlands to do my clinical residency. Easy, right?

A few months into my internship, Roel called me into his office and suggested that we reach out to Pieter Wesseling at the VU to ask if there was a possibility for me to join a PhD program in the Netherlands, while continuing my research in the US with him. I was just starting to realize that I needed more time to finish the work I had embarked on and importantly, the open-ended exploratory nature of science gave me a new creative outlet that I desperately needed.

I had met Pieter before while working on an imaging study. Pieter was excited about the idea and we met in the Netherlands during one of my visits to discuss these plans further. In the following months the graduate program office got involved to go over the relevant paperwork. Signing those documents effectively set in motion the trajectory leading up to this thesis over the past eight years, eventually concluding into the start of my own research group.

Roel and Pieter, I am grateful for believing in me and giving me the opportunity to learn from the both of you. Roel, you've trusted me to be the front man on several key lab projects, starting with the TCGA follow-up analysis and culminating in the GLASS consortium. Watching your career (and family) develop in the time that I was a part of your team inspired me to follow in your footsteps. Pieter, you've downplayed your role in this thesis but I strongly disagree: I have learned so much from you. Your comments on my writing, your interactions with others, were done all done so diligently and respectfully in a style were exemplary to me and that I now try to emulate.

To all the members of my thesis committee: prof. de Witt Hamer, dr. Brandsma, prof. Cuppen, dr. French, and dr. Niclou and prof. Holstege. Thank you for taking the time to review my tiny contribution to this field. Philip, we've known each other longer than the others. You were the one who first introduced me to programming in R, which turned out to be quite an important skillset when joining Roel's group. We've had some exciting projects together with Niels and I hope we'll have more chances to work together in the future.

Nearly eight years in the Verhaak lab meant I got to experience quite a few people joining and leaving. Starting that journey at MD Anderson I was so fortunate to have my first desk right next to Siyuan Zheng. Siyuan, you were an incredible mentor during those first years. Thanks for sustaining my daily interruptions with programming questions and random small talk. Kosuke, Qianghu, Kathy, Emmanuel and everyone else who was there during that time: we had a great time together.

About three years into my PhD the whole lab moved to The Jackson Laboratory in Connecticut. Some of us transitioned into new roles; others stayed at MD Anderson while a few dared to make the jump out east. Hoon and Samir: I'm glad we made that jump together. Both being computational heroes, you've taught me many tricks on making my work more efficient. Kevin Johnson, you were quite literally the first team member on-site at JAX, and I was lucky to get a desk right next to you. We juggled a few projects between the two of us, and somehow, we managed to pull things together irrespective of the slightly challenging, if not insane, timelines Roel gave us. Kevin Anderson, occasionally I would get lucky and beat you (or Shirley) at a board game, but I was no match for your encyclopedic knowledge in trivia. Fred, you really outdid yourself putting together a worthy GLASS successor. Eunhee and Amit, the two of you essentially ran the Verhaak wet lab: you've taught me a lot about molecular biology and you've had some beautiful papers. Moly and Dacheng, you helped bridge the generational gap that

was starting to form in the lab by introducing us all to TikTok. To everyone in the lab, we were such a fun group of people that would also occasionally hang out outside of our office. I was fortunate to get a chance to meet up with Fred again at a conference recently and I hope for more of those moments with many of you from now on.

To some it is puzzling that I was able to start my lab before defending my PhD. From the start, my position at MD Anderson and JAX was complicated because having graduated from medical school (which is considered a graduate program in the US), I was both a postdoctoral fellow (in the US) and a graduate student (in the Netherlands) in parallel. Therefore, there was minimal pressure to finish my thesis and moreover, COVID-19 prevented having an in-person defense for well over a year. I am thankful to Jeff Trent and Mike Berens at TGen for believing in me, and a special mention to Russ Rockne at City of Hope for making some of those critical first connections that made this happen. A very special mention to my team thus far (Yi-An, Noelle, Mimi) who are the real MVPs. Without reservation I believe that I have the best job in the planet.

I am thankful to many friends, both in the US and Netherlands, with whom I have shared food, drinks, arts, music and culture. Houston, New York, Phoenix, Netherlands friends: I look forward to many more chances to spend time together as we all grow old together. A very special mention goes out to my two paranymphs. Frank, we met through our roles as research exchange officers at our national chapter of the IFMSA and quickly became friends. It's somehow fitting that both of us have largely given up our clinical careers and are now working as full-time scientists. Jorn, we were roommates for several years and really bonded during that time, including dealing with a devastating loss in our circle of friends. We don't see each other as much these days living so far separated but I look forward to our next video calls and in-person dinners.

Then of course, my family. Mom, you've had a rough 2021 and it has been difficult for me to be so far away and not be able to help when you needed it most. Dad, although you've now retired you can't seem to stop working. I hope you'll find the time to finally build that train table you've been dreaming of. Josine, you've really picked things up in the last year and it's encouraging to see you shine. I'm so glad that you've found someone to confide in. Having the whole family visit us in Phoenix has been such a welcome treat. I already look forward to your next visit and making another trip together.

To all the rest of my family in the Netherlands: we don't see each other often enough, which I realize is mostly my fault for being so far away. I'm glad for all the chances that we do get to see each other. Perhaps we can all plan a big reunion sometime in the next few years? Know that our doors in Phoenix are always open so if you're ever planning a trip in the US, we're just a phone call or e-mail away. Last but not least: lieve Oma, wat ben ik blij dat u hier bij kunt zijn!

Then my in-laws: Becky, Lutz and the rest of the family in Manilla, Eastern Samar, Australia and New York City. You have all been so incredibly hospitable and kind to me from the first time that I came to visit the Philippines. Every trip I would gain at least a few pounds because of all the new foods you'd have me try. I'm thankful for introducing me to Filipino culture and for welcoming me into your family. I am looking forward to showing you everything the Netherlands has to offer.

Most importantly, my wife Nina: I could not have done this without you. You have been by my side since the start of this trajectory as we met right at the beginning in 2014. I'm grateful that you were one of the few who rarely asked me when I was finally going to "graduate", and trusted that it would all work out in the end. I can't wait to finally get all our immigration documents in order, so we can finally travel the world freely like we've always dreamed of. I'm so lucky to have you in my life and I cannot wait to find out what the future has in store for us.

Diffuse gliomas encompass a diverse group of tumors of the central nervous system that primarily affect the brain. Patients suffer from a poor prognosis despite radical treatment with surgery, radio- and chemotherapy. Microscopic evaluation by a neuropathologist has been the established method for the clinical diagnosis and determination of follow-up for over a century. Nevertheless, pathology assessment suffers from sampling biases and interobserver variability. This thesis details a number of studies that focus on the molecular analysis of this tumor type. They are a part of a larger series of papers in the past decade that demonstrate that molecular diagnostics can provide an unambiguous diagnosis and determination of follow-up treatment.

Floris Pieter Barthel was born on August 24, 1987 in the historical fortified town of Naarden, The Netherlands. In 2014 he obtained his MD degree at the VU University Medical Center in Amsterdam, The Netherlands. During his studies he developed an interest in neuro-oncology and joined the laboratory of Roel Verhaak at the MD Anderson Cancer Center (Houston, Texas, United States) and later The Jackson Laboratory (Farmington, Connecticut, United States). There he conducted a scientific traineeship studying the molecular makeup of glial brain tumors, leading to several scientific publications, including those described in this thesis. He currently leads his own research group at the Translational Genomics Research Institute (TGen) in Phoenix, Arizona.