

# **Mutational signatures of genotoxic cancer therapies**

Dissertation  
for  
the doctoral degree of  
Dr. rer. nat.

From the Faculty of Biology  
University of Duisburg-Essen  
Germany

Submitted by  
Dr. med. Emre Kocakavuk  
Born in Gelsenkirchen  
December 2022

## **Mutational signatures of genotoxic cancer therapies**

The computational experiments and analyses underlying the present work were conducted at the Department of Hematology and Stem Cell Transplantation, West German Cancer Center, University Hospital Essen & University of Duisburg-Essen, Essen, Germany and at The Jackson Laboratory for Genomic Medicine, CT, USA.

1. Examiner: Prof. Dr. Christian Reinhardt

2. Examiner: Prof. Dr. Ralf Küppers

3. Examiner: Prof. Dr. Ulrich Herrlinger

Chair of the Board of Examiners: Prof. Dr. Frank Kaiser

Date of the oral examination:

20.06.2023



# DuEPublico

Duisburg-Essen Publications online

UNIVERSITÄT  
DUISBURG  
ESSEN

*Offen im Denken*

ub | universitäts  
bibliothek

Diese Dissertation wird via DuEPublico, dem Dokumenten- und Publikationsserver der Universität Duisburg-Essen, zur Verfügung gestellt und liegt auch als Print-Version vor.

**DOI:** 10.17185/duepublico/78901

**URN:** urn:nbn:de:hbz:465-20230825-074845-5



Dieses Werk kann unter einer Creative Commons Namensnennung 4.0 Lizenz (CC BY 4.0) genutzt werden.

# Mutational signatures of genotoxic cancer therapies

## Table of contents

<b>Chapter 1.</b>	Summary	<b>4</b>
<b>Chapter 2.</b>	Introduction	<b>6</b>
<b>Chapter 3.</b>	Longitudinal molecular trajectories of gliomas	<b>13</b>
<b>Chapter 4.</b>	Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer	<b>42</b>
<b>Chapter 5.</b>	Glioma progression is shaped by genetic evolution and microenvironment interactions	<b>63</b>
<b>Chapter 6.</b>	Comparative molecular life history of spontaneous canine and human gliomas	<b>97</b>
<b>Chapter 7.</b>	Discussion	<b>121</b>
<b>Chapter 8.</b>	Appendices	<b>129</b>
	List of Abbreviations	<b>127</b>
	List of Figures	<b>130</b>
	List of Tables	<b>130</b>
	References	<b>131</b>
	Avoidance of self-plagiarism	<b>137</b>
	Statutory declarations	<b>138</b>

# **Chapter 1**

## **Summary (English)**

Cancer is a leading cause of death globally, and the incidence is expected to increase to 28.4 million cases by 2040. Precision oncology seeks to address this challenge through the identification and application of biomarkers to personalize treatment in a genomics-based, biomarker-driven approach. In this thesis, I aimed to comprehensively analyze the genomic impact of cancer chemotherapy and radiotherapy, both of which are genotoxic and expected to leave a mutational footprint in the cancer genome. By integration with comprehensive clinical data, I assessed the prognostic and predictive value of the signatures identified. My thesis has a strong focus on diffuse gliomas, the most common and aggressive type of brain tumors in adults, which have a poor prognosis despite intensive multimodal treatment. Standard therapy for diffuse gliomas includes surgery, radiotherapy, and chemotherapy. I started by examining the genomic impact of the alkylating agent temozolomide in gliomas, and subsequently analyzed the genomic effects of radiotherapy in a pan-cancer setting. Following the genomic analyses, I investigated the longitudinal non-genetic changes in gliomas associated with TMZ-associated hypermutation and RT-associated deletion signatures. To potentially overcome translational barriers in cancer, I propose the use of spontaneous canine gliomas as a potential model for preclinical and translational science to facilitate the knowledge transfer into clinical practice. I conclude by discussing these findings, focusing on the potential to integrate them into a precision oncology framework and identify novel therapeutic avenues. The use of next-generation sequencing technology and comprehensive molecular characterization have facilitated diagnostic advances in the field of gliomas, including the use of molecular markers and the updated WHO Classification of Tumors of the Central Nervous System. These advances have yet to be translated into successful treatment strategies. My analyses provide valuable insights into the mechanisms of treatment resistance and open future opportunities for biomarker-guided treatment strategies as part of a precision oncology approach.

# Summary (Deutsch)

Krebs ist weltweit eine der häufigsten Todesursachen. Es wird erwartet, dass bis 2040 die Inzidenz auf 28,4 Millionen ansteigen wird. Die „Präzisionsonkologie“ versucht, die hiermit verbundenen Herausforderung durch die Identifizierung und Anwendung von Biomarkern zur Personalisierung der Behandlung in einem Genom-basierten, Biomarker-gesteuerten Ansatz zu bewältigen. In dieser Arbeit habe die genomischen Auswirkungen von Chemo- und Strahlentherapie umfassend analysiert. Bei diesen genotoxischen Therapien ist es zu erwarten, dass sie Mutationen im Krebsgenom hinterlassen. Durch die Integration mit klinischen Daten habe ich den prognostischen und prädiktiven Wert der identifizierten Signaturen erfasst. Der Schwerpunkt meiner Arbeit lag auf diffusen Gliomen, der häufigsten und aggressivsten Form von Hirntumoren bei Erwachsenen, die trotz intensiver multimodaler Behandlung eine schlechte Prognose haben. Die Standardtherapie für diffuse Gliome umfasst eine Operation, Strahlentherapie und Chemotherapie. Zunächst habe ich die genomischen Auswirkungen der alkylierenden Substanz Temozolomid bei Gliomen untersucht und anschließend analysierte ich die genomischen Auswirkungen der Strahlentherapie in einem Krebsentitäts-übergreifenden Datensatz. Im Anschluss an die genomischen Analysen untersuchte ich die longitudinalen nicht-genetischen Veränderungen in Gliomen, die mit zuvor identifizierten genomischen Biomarkern einhergehen. Um translationale Barrieren in der Krebsforschung zu überwinden, schlage ich die potentielle Verwendung von spontanen Hundegliomen als Modell für die präklinische und translationale Wissenschaft vor, um den Wissenstransfer in die klinische Praxis zu erleichtern. Abschließend erörtere ich diese Ergebnisse und setze dabei einen besonderen Fokus auf das Potenzial, sie in einen präzisionsonkologischen Ansatz zu integrieren und neue therapeutische Wege zu erörtern. Der Einsatz der NGS-Technologie und eine umfassende molekulare Charakterisierung haben diagnostische Fortschritte auf dem Gebiet der Gliome ermöglicht, einschließlich der Verwendung molekularer Marker und der aktualisierten WHO-Klassifikation von ZNS-Tumoren. Diese Fortschritte sind bisher noch nicht in erfolgreiche Behandlungsstrategien umgesetzt worden. Meine Analysen liefern wertvolle Einblicke in die molekularen Mechanismen der Therapieresistenz und eröffnen künftige Möglichkeiten für Biomarker-gesteuerte Behandlungsstrategien als Teil eines präzisionsonkologischen Ansatzes.

# **Chapter 2**

## **Introduction**

Cancer is one of the leading causes of death worldwide, and its incidence is increasing due to the aging population<sup>2,3</sup>. In particular, cancer incidence is expected to increase to 28.4 million cases by 2040<sup>2</sup>, posing a major medical, societal, and economic challenge. In recent decades, an increasing number of cancer treatment modalities have evolved, currently comprising surgical resection, chemo-, radio-, targeted-, immuno-, and cellular therapy. These treatment options require careful selection of the right treatment for the right patient at the right time. The dynamically evolving field of “precision oncology” seeks to address this challenge through the identification and application of biomarkers<sup>4</sup>. The overall goal is to apply personalized treatment in a genomics-based, biomarker-driven approach. However, a comprehensive picture of the mutational patterns associated with cancer therapy is needed to understand the precise genomic effects and outcomes of these treatments.

In this thesis, I focused on mutational signatures of commonly used and established cancer therapies, i.e. chemotherapy and radiotherapy, which are known to be genotoxic and expected to leave a mutational footprint in the cancer genome<sup>5</sup>. To enable a potential translation of these findings into the clinic, I linked this knowledge to comprehensive clinical data and assessed the prognostic and predictive value of the identified signatures.

### **Diffuse gliomas in adults**

Cancer can affect many different organs and tissues. Some cancer types are notoriously aggressive and difficult to treat, including particularly devastating tumors in the brain. Brain tumors constitute the leading cause of cancer-related mortality among males under the age of 40 in the United States<sup>6,7</sup>. Diffuse glioma is the most common and most aggressive type of brain tumors in adults, placing an immense burden on patients, their families, and caregivers<sup>6</sup>. Particularly devastating is the fact that despite intensive multimodal treatment with surgery, radiotherapy, and chemotherapy, nearly every patient suffers a tumor relapse with dismal prognosis. The

## Mutational signatures of genotoxic cancer therapies

most aggressive glioma subtype, known as glioblastoma (GBM) is characterized by a median overall survival of 15 months<sup>8</sup> and a 5-year relative survival rate of only 7%<sup>7</sup>.

Over the past two decades, many large-scale randomized and controlled clinical trials (RCT) have been conducted to improve survival outcomes<sup>8-14</sup>, but only a very limited number of breakthrough treatment modalities have been reported since the gold standard “Stupp regimen” from 2005<sup>8,15</sup>. The standard therapy comprises maximal extent of resection<sup>16-18</sup>, followed by postsurgical, involved-field, fractionated radiotherapy (RT) and concomitant chemotherapy with the alkylating agent temozolomide (TMZ), which is followed by 6 cycles of adjuvant TMZ<sup>8</sup>. A cumulative radiation dose of 60 Gy (2Gy per fraction, 5 days per week over 6 weeks) is applied. During RT, 75mg/m<sup>2</sup> of body surface area per day (7 days per week over 6 weeks) of TMZ is given concurrently. In the adjuvant phase, 150-200mg/m<sup>2</sup> of body surface area per day (d1-5, 28-day cycle, 6 cycles) of TMZ is applied.

The utility of next-generation sequencing technology and comprehensive molecular characterization have provided an excellent foundation for diagnostic advances in the field of gliomas<sup>19-25</sup>. Diagnostic criteria have shifted from traditional histological criteria to the use of molecular markers and have been updated with the fourth edition of the WHO Classification of Tumors of the Central Nervous System in 2016<sup>26</sup>. Clinically important biomarkers include mutations of the isocitrate dehydrogenase (*IDH*) 1/2 genes, loss of heterozygosity with specific deletions of the chromosome arms 1p and 19q (1p/19q-codeletion) and promoter methylation of the DNA repair enzyme O(6)-methylguanine-DNA methyltransferase (*MGMT*). The classification distinguishes three types of adult diffuse gliomas: *Oligodendroglioma, IDH-mutant and 1p/19q-codeleted* (IDHmut-codel, better prognosis); *Astrocytoma, IDH-mutant* (IDHmut-noncodel, intermediate prognosis); and *Glioblastoma, IDH-wildtype* (IDHwt, worst prognosis)<sup>26</sup>.

Despite progress, tumor relapse occurs inevitably, and recurrent tumors clinically display a more aggressive phenotype that is less responsive to further treatment. To improve survival outcomes of glioma patients, the Glioma Longitudinal Analysis (GLASS) Consortium set out to improve our understanding of mechanisms of resistance to RT and TMZ<sup>27</sup>. During the time of this thesis work, I joined the analysis working group of GLASS located at the Jackson Laboratory for Genomic Medicine. I

## Mutational signatures of genotoxic cancer therapies

performed longitudinal studies comparing glioma samples before and after treatment, which provided a unique opportunity to thoroughly characterize the genomic, molecular, and cellular evolutionary trajectories following treatment<sup>26</sup>.

### Genotoxic cancer therapies

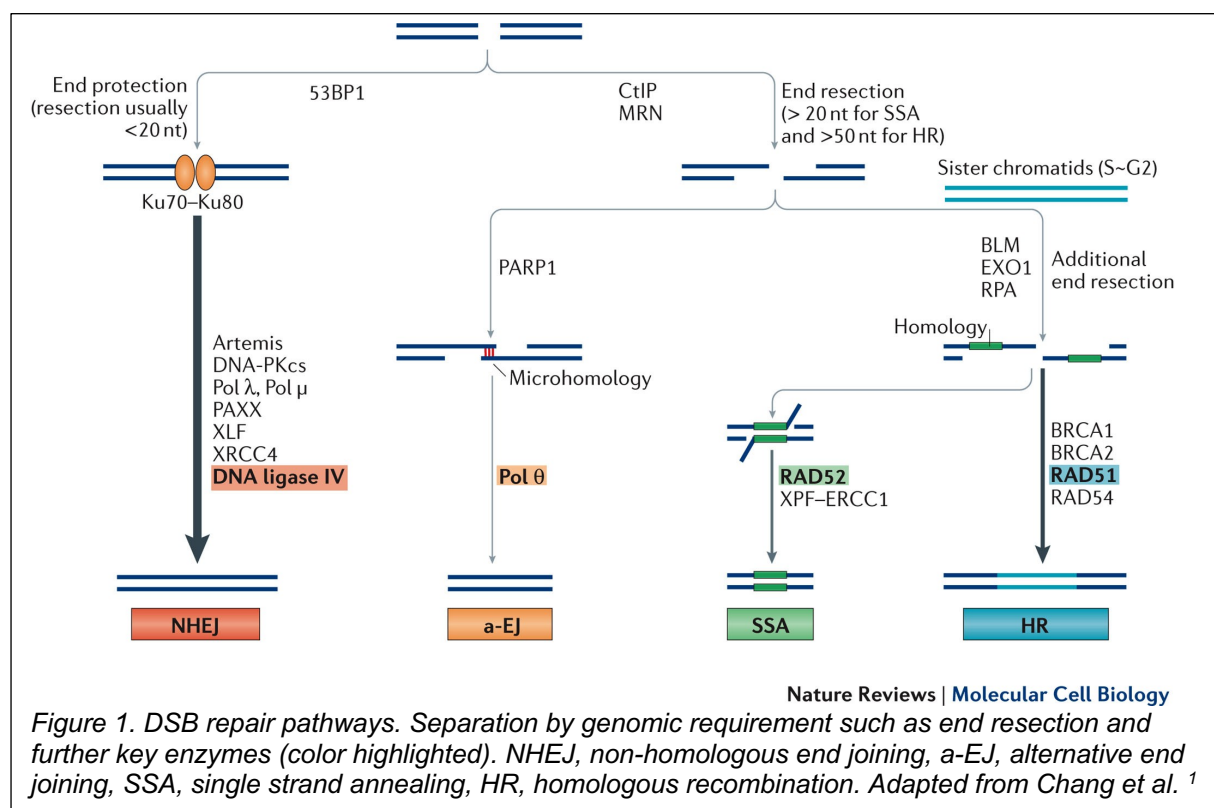
Cancer treatment strategies have evolved rapidly in recent decades, contributing to improved cancer survival outcomes<sup>3</sup>. One of the best-known mechanisms of action of cancer therapies is their ability to cause DNA damage. Radiotherapy and alkylating agent-based chemotherapeutic agents can have a wide range of direct and indirect genotoxic properties that target cancer cells<sup>28,29</sup>. Exposure towards these genotoxic therapies may shape the cancer genome throughout their clonal evolution. However, cell-intrinsic repair mechanisms may limit therapeutic efficacy<sup>30</sup>, and when unintentionally targeting healthy cells, these treatments risk causing second malignancies<sup>31,32</sup>.

RT is a mainstay of treatment for cancer patients. Ionizing radiation is genotoxic and interacts with DNA in a largely stochastic manner. It can cause DNA damage indirectly via reactive oxygen species and directly by inducing single strand breaks (SSBs) and double strand breaks (DSBs)<sup>33</sup>. Other effects include the activation of intracellular cytotoxic signaling pathways<sup>34</sup> and modulation of the tumor- and immune microenvironment<sup>35</sup>.

Although DSBs are less abundant than SSBs, they are considered to be more lethal to cancer cells. If DSBs are not repaired or are repaired in a way that inhibits subsequent replication, cancer cells usually undergo cell death. Importantly, DNA repair mechanisms can be activated in response to DSBs, which can broadly be divided into error-free and error-prone pathways<sup>1</sup>. The homologous recombination (HR) pathway uses a sister chromatid as a template for error-free repair of DSBs<sup>36</sup>. However, due to this template requirement, HR can only be active during cell cycle phases when the sister chromatid is available. i.e., during the S and G2 phases of the mitotic cell cycle<sup>36</sup>. The three pathways, canonical/classical non-homologous end joining (c-NHEJ), alternative end joining (a-EJ) and single-strand annealing (SSA) typically result in mutations due to imperfect repair of the nuclear sequence, rendering them “error-prone”<sup>1</sup>. Each of these pathways has distinct genomic and molecular

## Mutational signatures of genotoxic cancer therapies

requirements and conversely leads to different genomic outcomes. c-NHEJ does not require microhomologous sequences present at exposed DNA ends, is initiated by 53BP1, and end-resection is limited due to the protected binding of Ku70-Ku80<sup>37</sup>. Recruitment of DNA-dependent protein kinase (DNA-PK) and Artemis to Ku binding sites is followed by further downstream ligation of the DNA ends by DNA ligase IV<sup>38</sup>. In contrast, a-EJ (also referred to as microhomology-mediated end joining) and SSA typically require end-resection involving the carboxy-terminal binding protein interacting protein (CtIP) and the MRE11-RAD50-NBS1 (MRN) complex<sup>39</sup>. In addition, a-EJ typically uses microhomology lengths of 2-20 base pairs (bp) and SSA uses (micro)homology lengths of >20bp<sup>1</sup>. The error-prone pathways can cause insertion-deletion mutations after repair of RT-induced DSBs<sup>1</sup>. I reasoned that these mutations and their features could be detected by genomic sequencing of post-treatment cancer samples, which I performed in a large-scale analysis using cancer genomics techniques.



TMZ is frequently used in the treatment of glioma patients. The mechanism of action of this alkylating agent prodrug is the addition of a methyl groups to DNA purine bases, particularly O6-guanine. Instead of the natural pairing of guanine (G) with



## Mutational signatures of genotoxic cancer therapies

cytosine (C), the resulting O6-methylguanine (O6-MeG) forms a mismatch with thymine (T), resulting in a C > T transition mutation<sup>40</sup>. However, the methyl group can be removed by MGMT through direct repair, resulting in TMZ-resistance. Therefore, it is not surprising that gene silencing via promotor methylation of *MGMT* causes a treatment response<sup>10</sup>. *MGMT* promoter methylation has therefore been described as a strong predictive marker for TMZ response in IDHwt GBM<sup>41</sup>.

Moreover, effective mismatch repair (MMR) is required for a successful TMZ response. The MMR machinery excises the mis-paired T, replaces it with the correct C, but in the following replication cycle, T is reinserted, resulting in a futile repair cycle associated with the generation of DSBs and cell death<sup>40</sup>. However, when the MMR machinery is defective due to underlying mutations in MMR genes (*MSH6*, *MSH2*, *MLH1*, *PMS2*), the C > T transition can be tolerated. In fact, the combination of TMZ treatment and acquired mutations in the MMR pathway can result in the hypermutator (HM) phenotype at recurrence<sup>42</sup>. The dominant mutations in HM samples occur in a very specific trinucleotide context, resulting in a dominant single base substitution mutation signature 11 (SBS11)<sup>43,44</sup>. In this thesis, I have analyzed the genomic properties associated with alkylating agent-induced hypermutation, quantified the distribution across the distinct glioma molecular subtypes and shed light on the potential clinical relevance of this signature.

## Cancer mutational signatures

Somatic mutations are an important hallmark of cancer<sup>45</sup>. Distinct mutational processes can cause unique patterns of genetic changes in the cancer genome, referred to as mutational signatures. The concept of mutational signatures was first introduced by Nik-Zainal et al. in 2012 through an analysis of whole genome sequencing (WGS) data of 21 breast cancer samples<sup>46</sup>. In addition to focusing on the single base substitution (SBS) itself, the trinucleotide context including 5'- and 3'-neighboring bases was also considered<sup>46</sup>. A matrix of 96-possible trinucleotide contexts was mapped, and signatures were extracted using a non-negative matrix factorization (NMF) approach<sup>47</sup>.

In a comprehensive, pan-cancer analysis, these mutational signatures were then linked to the mutational processes that may underly them<sup>44</sup>. Mutations can

## Mutational signatures of genotoxic cancer therapies

originate from endogenous and exogeneous processes. Examples of endogenous mechanisms include DNA repair deficiencies, APOBEC mutagenesis, or aging-associated deamination of 5-methyl-cytosine<sup>44,48</sup>. Well known exogenous mutational processes in certain cancer types include ultraviolet light in skin cancer, and tobacco smoking in lung cancer<sup>44</sup>. However, there is increasing evidence in the literature that some anticancer agents may leave a mutational footprint in the cancer genome. Mutational signatures associated with prior treatment with platinum-based chemotherapy (SBS 25, SBS 31, SBS 35)<sup>5,49-54</sup>, temozolomide (SBS 11)<sup>43,49</sup>, capecitabine and fluorouracil (SBS 17b)<sup>5,49</sup>, and thiopurine (SBS 87)<sup>49,55,56</sup> have been described.

In fact, exposure to ionizing radiation may also have effects on the genome. However, there is no consensus on the impact of radiotherapy on the cancer genome, and several studies reported a variety of genomic effects. Prior studies analyzed radiation-induced tumors in humans<sup>31,57,58</sup> or mouse models<sup>59,60</sup> rather than therapeutic radiation.

For example, Behjati et al. demonstrated that radiation-induced secondary malignancies were enriched for small deletions and found considerable microhomology at junction sites, suggesting that DNA damage repair in secondary malignancies may be mediated by microhomology-mediated end joining<sup>31</sup>. In studies with a limited number of tumors derived from mouse models of radiation-induced malignancies, enrichment of single base substitutions rather than small deletions was observed<sup>59,61</sup>. A recent study by Li et al. described a broad spectrum of genomic alterations in ionizing radiation-induced tumors in mice, including single base substitutions, indels, larger structural variants, copy number gains/losses as well as whole chromosome gains/losses<sup>60</sup>. Interestingly, two studies on RT-induced secondary gliomas did not show a high burden of single base substitutions or indels, but rather an enrichment of copy number alterations<sup>57</sup> or driver alterations in PDGFRA<sup>58</sup>.

In summary, previous studies have shown a wide range of genomic changes in response to ionizing radiation and have suggested the involvement of various DNA double strand break repair mechanisms. The studies mentioned had limited sample sizes and studied radiation-induced malignancies rather than therapeutic radiation. I

## Mutational signatures of genotoxic cancer therapies

reasoned that there may be mechanistic differences between somatic variants in secondary malignancies and therapeutic radiation because the vast majority of irradiated individuals do not develop secondary malignancies. With the advantage of the addition of insertion-deletion (ID) signatures to a recent update of the mutational signature catalogue<sup>49</sup>, I was able to comprehensively characterize the mutational signatures associated with radiotherapy in pre- and post-treatment cancer samples.

### Outline of the thesis

Genotoxic cancer treatment strategies are widely used and effective. However, knowledge of how they are affecting the cancer genome is limited. To address this issue, I have comprehensively analyzed the genomic effects of gold standard therapies in pre- and post-treatment cancer samples, focusing particularly on gliomas. In **Chapter 3**, I highlight the genomic impact of the alkylating agent temozolomide in gliomas, while in **Chapter 4** I focus on the genomic effects of radiotherapy in a pan-cancer setting. For both treatments, I dissected specific mutational signatures and integrated comprehensive clinical data to assess their translational relevance as potential biomarkers. In **Chapter 5**, I analyze the longitudinal non-genetic changes in glioma associated with TMZ-associated hypermutation and RT-associated deletion signatures and integrate these results with the corresponding genomic markers.

One of the reasons for the lack of novel treatment opportunities in the field of gliomas is the lack of appropriate models for preclinical and translational science. This hinders the effective transfer of knowledge into the clinics. **Chapter 6** proposes the use of spontaneous canine glioma as a potential model to address this problem. Lastly, in **Chapter 7**, I summarize and discuss the results, focusing particularly on novel questions arising from my analyses and how the results can potentially be integrated into a precision oncology framework.

## Chapter 3

# Longitudinal molecular trajectories of diffuse gliomas

This chapter is based on the following publication<sup>62</sup>:

Barthel, F.P., Johnson, K.C., Varn, F.S., Moskalik, A.D., Tanner, G., **Kocakavuk, E.** *et al.* Longitudinal molecular trajectories of diffuse glioma in adults. *Nature* **576**, 112–120 (2019). <https://doi.org/10.1038/s41586-019-1775-1>

Article

# Longitudinal molecular trajectories of diffuse glioma in adults

<https://doi.org/10.1038/s41586-019-1775-1>

Received: 8 February 2019

Accepted: 1 October 2019

Published online: 20 November 2019

Floris P. Barthel<sup>1,2,72</sup>, Kevin C. Johnson<sup>1,72</sup>, Frederick S. Varn<sup>1</sup>, Anzhela D. Moskalik<sup>1</sup>, Georgette Tanner<sup>3</sup>, Emre Kocakavuk<sup>1,4,5</sup>, Kevin J. Anderson<sup>1</sup>, Olajide Abiola<sup>1</sup>, Kenneth Aldape<sup>6</sup>, Kristin D. Alfaro<sup>7</sup>, Donat Alpar<sup>8,9</sup>, Samirkumar B. Amin<sup>1</sup>, David M. Ashley<sup>10</sup>, Prati Bandopadhyay<sup>11,12</sup>, Jill S. Barnholtz-Sloan<sup>13</sup>, Rameen Beroukhi<sup>12,14</sup>, Christoph Bock<sup>8,15</sup>, Priscilla K. Brastianos<sup>16</sup>, Daniel J. Brat<sup>17</sup>, Andrew R. Brodbelt<sup>18</sup>, Alexander F. Bruns<sup>3</sup>, Ketan R. Bulsara<sup>19</sup>, Aruna Chakrabarty<sup>20</sup>, Arnab Chakravarti<sup>21</sup>, Jeffrey H. Chuang<sup>1,22</sup>, Elizabeth B. Claus<sup>23,24</sup>, Elizabeth J. Cochran<sup>25</sup>, Jennifer Connelly<sup>26</sup>, Joseph F. Costello<sup>27</sup>, Gaetano Finocchiaro<sup>28</sup>, Michael N. Fletcher<sup>29</sup>, Pim J. French<sup>30</sup>, Hui K. Gan<sup>31,32</sup>, Mark R. Gilbert<sup>33</sup>, Peter V. Gould<sup>34</sup>, Matthew R. Grimmer<sup>27</sup>, Antonio Iavarone<sup>35,36,37</sup>, Azzam Ismail<sup>20</sup>, Michael D. Jenkinson<sup>18</sup>, Mustafa Khasraw<sup>38</sup>, Hoon Kim<sup>1</sup>, Mathilde C. M. Kouwenhoven<sup>39</sup>, Peter S. LaViolette<sup>40</sup>, Meihong Li<sup>1</sup>, Peter Lichter<sup>29</sup>, Keith L. Ligon<sup>12,41</sup>, Allison K. Lowman<sup>40</sup>, Tathiane M. Malta<sup>42</sup>, Tali Mazor<sup>27</sup>, Kerrie L. McDonald<sup>43</sup>, Annette M. Molinaro<sup>27</sup>, Do-Hyun Nam<sup>44,45</sup>, Naema Nayyar<sup>16</sup>, Ho Keung Ng<sup>46</sup>, Chew Yee Ngan<sup>1</sup>, Simone P. Niclou<sup>47</sup>, Johanna M. Niers<sup>48</sup>, Houtan Noushmehr<sup>42</sup>, Javad Noorbakhsh<sup>1</sup>, D. Ryan Ormond<sup>48</sup>, Chul-Keek Park<sup>49</sup>, Laila M. Poisson<sup>50</sup>, Raul Rabadan<sup>51,52</sup>, Bernhard Radlwimmer<sup>29</sup>, Ganesh Rao<sup>53</sup>, Guido Reifenberger<sup>54</sup>, Jason K. Sa<sup>45</sup>, Michael Schuster<sup>8</sup>, Brian L. Shaw<sup>16</sup>, Susan C. Short<sup>3</sup>, Peter A. Silveis Smitt<sup>30</sup>, Andrew E. Sloan<sup>55,56,57</sup>, Marion Smits<sup>58</sup>, Hiromichi Suzuki<sup>59</sup>, Ghazaleh Tabatabai<sup>60</sup>, Erwin G. Van Meir<sup>61</sup>, Colin Watts<sup>62</sup>, Michael Weller<sup>63</sup>, Pieter Wesseling<sup>2,64</sup>, Bart A. Westerman<sup>65</sup>, Georg Widhalm<sup>66</sup>, Adelheid Woehrer<sup>67</sup>, W. K. Alfred Yung<sup>7</sup>, Gelareh Zadeh<sup>68</sup>, Jason T. Huse<sup>69,70</sup>, John F. De Groot<sup>7</sup>, Lucy F. Stead<sup>3</sup>, Roel G. W. Verhaak<sup>1\*</sup> & The GLASS Consortium<sup>71</sup>

The evolutionary processes that drive universal therapeutic resistance in adult patients with diffuse glioma remain unclear<sup>1,2</sup>. Here we analysed temporally separated DNA-sequencing data and matched clinical annotation from 222 adult patients with glioma. By analysing mutations and copy numbers across the three major subtypes of diffuse glioma, we found that driver genes detected at the initial stage of disease were retained at recurrence, whereas there was little evidence of recurrence-specific gene alterations. Treatment with alkylating agents resulted in a hypermutator phenotype at different rates across the glioma subtypes, and hypermutation was not associated with differences in overall survival. Acquired aneuploidy was frequently detected in recurrent gliomas and was characterized by IDH mutation but without co-deletion of chromosome arms 1p/19q, and further converged with acquired alterations in the cell cycle and poor outcomes. The clonal architecture of each tumour remained similar over time, but the presence of subclonal selection was associated with decreased survival. Finally, there were no differences in the levels of immunoediting between initial and recurrent gliomas. Collectively, our results suggest that the strongest selective pressures occur during early glioma development and that current therapies shape this evolution in a largely stochastic manner.

Diffuse glioma is the most common malignant brain tumour in adults and invariably relapse despite treatment with surgery, radiotherapy and chemotherapy. The molecular landscape of glioma at diagnosis has been extensively characterized<sup>3–9</sup>. Although these efforts have led to the identification of driver genes and clinically relevant subtypes<sup>10,11</sup>, how the glioma genetic landscape evolves over time and in response to therapy is unknown.

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<sup>12–15</sup>. Tumours are thought to circumvent these growth bottlenecks by dynamic competition of subclones that result in the most favourable environment for tumour sustenance<sup>1</sup>. Recent studies have suggested that stochastic changes in clone frequency (that is, neutral evolution) and immune surveillance

\*A list of affiliations appears at the end of the paper.

may further contribute to the observed intratumoural heterogeneity<sup>16,17</sup>. An understanding of evolutionary dynamics at several time points is needed to develop strategies aimed at delaying or preventing the onset of tumour 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<sup>18</sup>. We have analysed 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 towards highly variable and patient-specific trajectories of genomic alterations.

## GLASS cohort

We pooled existing and newly generated longitudinal DNA sequencing datasets from 288 patients treated at 35 hospitals (Supplementary Table 1, Extended Data Fig. 1). After applying quality filters, tumour 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 (hereafter referred to as IDH-mutant-codel;  $n = 25$ ); (2) IDH-mutant without co-deletion of chromosome 1p/19q (hereafter IDH-mutant-noncodel;  $n = 63$ ); and (3) IDH-wild-type ( $n = 134$ ), in alignment with the World Health Organization (WHO) classification of tumours of the central nervous system<sup>10,11</sup>. For each patient, we selected two time-separated tumour samples, henceforth termed initial and recurrence, for further analysis.

## 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 tumours were comparable with previously reported rates<sup>6,7,19</sup>. There were 2.20 mutations (single-nucleotide variants and small insertions or deletions) per megabase (Mb) for IDH-mutant-codels; 2.52 mutations per Mb for IDH-mutant-noncodels; and 2.85 mutations per Mb for IDH-wild-type glioma (Fig. 1a, Extended Data Fig. 2a). Excluding DNA hypermutation cases (more than 10 mutations per Mb,  $n = 35$ ), the mutation burden increased after recurrence in 70% of the cohort (Extended Data Fig. 2a). To study changes during tumour progression, we separated mutations into three fractions: initial only, recurrence only, or shared. Notably, the mutation burdens of the private fractions, but not the shared fraction, were comparable between subtypes (Extended Data Fig. 2b). Patient age at diagnosis was significantly associated with the shared mutational burden ( $P = 1.7 \times 10^{-7}$ ), and to a lesser extent with the burden of mutations private to the initial tumour ( $P = 0.0256$ ) (Extended Data Fig. 2c). On average, a longer time to recurrence was associated with a larger increase in mutation burden ( $P = 0.0043$ , Extended Data Fig. 2d).

These fraction-specific differences in mutational 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<sup>20</sup>. As expected, signature activity was closely related to subtype and fraction (Fig. 1b, Extended Data Fig. 3a). Signature 1 (ageing) was nearly always the dominant signature among shared mutations in IDH-wild-type tumours, whereas the shared fraction in IDH-mutant-noncodel and IDH-mutant-codel tumours—tumour subtypes that are associated with a younger age of diagnosis—also showed a strong presence of signature 16 (unknown aetiology). Signatures 3 (double-strand break repair), 15 (mismatch repair) and 8 (unknown

aetiology) were mostly confined to the private fractions, which suggests that these processes were of lesser importance to tumour maintenance than those associated with ageing.

The treatment of glioma includes alkylating agents that can induce hypermutations after treatment<sup>21–23</sup>. We observed enrichment of the associated signature 11 in recurrent tumours treated with alkylating agents and with a mutational load exceeding 10 mutations per Mb (Fig. 1a, Extended Data Fig. 3b). Treatment-associated hypermutation occurred most frequently among IDH-mutant-noncodels (47%), followed by IDH-mutant-codels (25%), and IDH-wild-type gliomas (16%) (Fig. 1c). The proportion of hypermutation events was significantly different between the three glioma subtypes (Fisher's exact test  $P = 2.0 \times 10^{-3}$ ), which suggests that IDH-mutant-noncodels are most sensitive to developing a hypermutator phenotype<sup>24</sup>.

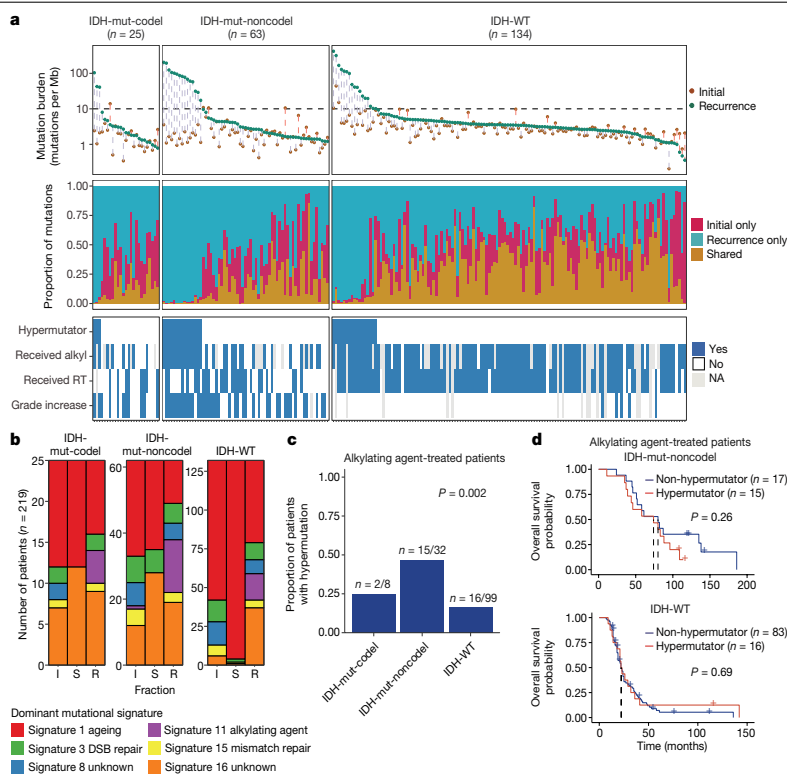
Treatment-induced hypermutation has been associated with disease progression<sup>23</sup>. We did not find any differences in overall survival between hypermutators and non-hypermutators treated with alkylating agents independent of age, subtype and *MGMT* methylation status (Fig. 1d, Supplementary Table 2a, b). To assess the pathogenicity of acquired mutations further, we studied their clonality<sup>25</sup>. Newly acquired clonal mutations have penetrated most of the tumour (that is, a selective sweep) between initial and recurrence and mark clonal expansion<sup>26</sup>. Conversely, acquired subclonal mutations are less prevalent, and therefore less likely to drive disease progression. Previous reports have suggested that mutations associated with alkylating agents are frequently clonal<sup>27</sup>. We found that in 48% of hypermutated tumours, most of the recurrence-only mutations were clonal, potentially reflecting cases in which a selective sweep occurred (Extended Data Fig. 4a). However, IDH-mutant-noncodel hypermutators with predominantly clonal mutations did not show differences in survival compared with those containing predominantly subclonal mutations (log-rank test  $P = 0.38$ , Extended Data Fig. 4b). Alkylating agents such as temozolomide prolong the survival of adult patients with glioma<sup>28,29</sup>. Our results show that treatment-induced hypermutation is common across subtypes and does not associate with reduced overall survival, supporting the noted benefit of alkylating agent therapy.

## 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 on the basis of their cancer cell fraction (CCF). CCF values represent the fraction of cancer cells that contain a given alteration and reflect the relative timing of events, because alterations that are present in a subset of cancer cells probably occurred later than events present in all cancer cells (Fig. 2a). Most tumours (84%) demonstrated a mutational cluster with a CCF greater than 50% that persisted from the initial tumour to recurrence, probably reflecting the tumour trunk and containing the tumour-initiating driver mutations<sup>30</sup> (Fig. 2b, Extended Data Fig. 5a). To determine changes in clonal dominance over time, we ranked clusters within each sample by their CCF value and found similarities in clonal architecture throughout the course of disease (Kendall rank correlation,  $\tau = 0.20$ ,  $P = 3.76 \times 10^{-24}$ ; Fig. 2b, Extended Data Fig. 5b–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 tumours by determining the normalized ratio between non-synonymous and synonymous mutations (dN/dS)<sup>31</sup>. Higher ratios (above one) 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 (Fig. 2c). Separating mutations into mutational fractions demonstrated that shared but not private mutations showed positive dN/dS ratios in all three glioma subtypes, which indicates that only shared

## Article



**Fig. 1 | Temporal changes in glioma mutational burden and processes.**

**a**, Each column represents a single patient ( $n = 222$ ) at two separate time points grouped by glioma subtype and ordered left-to-right by decreasing mutation frequency at recurrence. Top, mutation frequency differences between initial and recurrent tumours. Blue dotted line indicates increased mutation frequency, and the red dotted line indicates decreased mutational frequency. Middle, the proportion of total mutations shared (mustard), private to initial (magenta), or private to recurrence (blue). Bottom, clinical information including hypermutation status, therapy and grade changes. RT, radiation

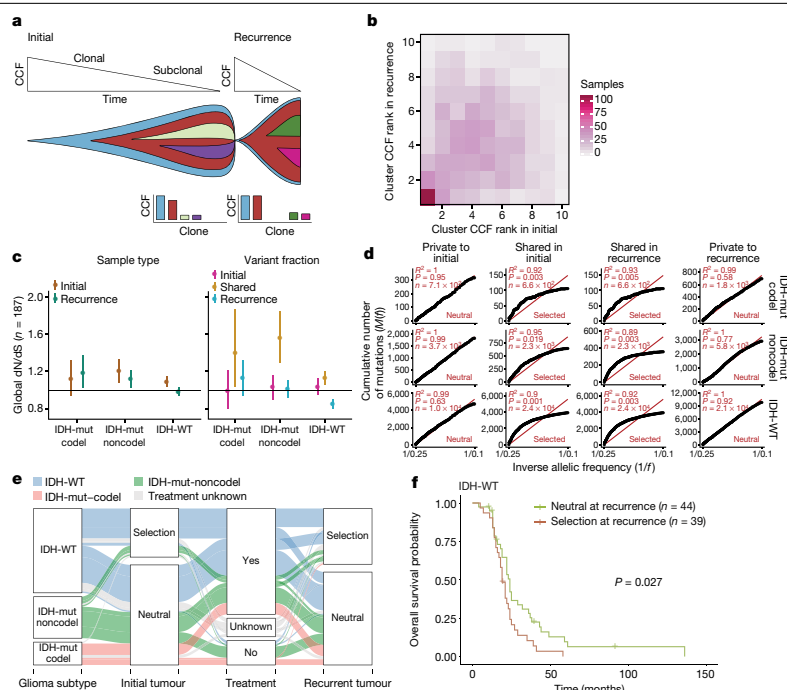
therapy. Alkyl, alkylating agent. **b**, Stacked bar plot ( $n = 219$ ) indicating the dominant mutational signature among initial, recurrent and shared mutation fractions stratified by glioma subtype. I, initial; S, shared; R, recurrence. **c**, The proportion of glioma recurrences with alkylating agent-related 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) patients treated with alkylating agent among IDH-wild-type (left,  $n = 99$ ) and IDH-mutant-noncodel (right,  $n = 32$ ) tumours.  $P$  values determined by log-rank test.

mutations (including truncal mutations) are likely to be subject to positive selection (Fig. 2c). The dN/dS ratio of initial-only mutations showed that these are neither positively nor negatively selected for, whereas recurrence-only mutations were subject to negative selection in IDH-wild-type gliomas.

To verify the reduced selective pressure in the private mutations, we used an orthogonal method to test for evidence of selection<sup>32</sup>. The method uses distributions of variant allele frequencies and estimated mutation rates to detect whether profiles significantly deviate from a model of neutral evolution (that is, as depicted by a linear relationship in Fig. 2d). In accordance with results of the dN/dS ratios, private mutations demonstrated dynamics that were consistent with neutral evolution (Fig. 2d). Shared subclonal mutations deviated from linearity and were consistent with selection both in non-hypermutators and hypermutators (Fig. 2d, Extended Data Fig. 6a, b), which provides further evidence that the strongest selective forces occur early in gliomagenesis.

Cohort-level analysis of selection masks the heterogeneity that exists in individual evolutionary trajectories. To determine the selective effects at each tumour time point, we used a Bayesian framework (SubClonalSelection algorithm) that simultaneously provides

sample-specific probabilities for both selection and neutrality while modelling 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 IDH-mutant-codels, 29 IDH-mutant-noncodels, 59 IDH-wild-type; Supplementary Table 3). Neutral-to-neutral was the most common evolutionary trajectory across all three subtypes (52%), and IDH-wild-type tumours displayed the highest observed selection at any time point, with selection detected in 64% of tumours (Fisher's exact test  $P = 0.01$ ; Fig. 2e, Supplementary Table 3). IDH-wild-type gliomas with evidence for selection at recurrence had a shorter overall survival than IDH-wild-type gliomas classified as neutral at recurrence ( $P = 0.027$ ; log-rank statistic, Fig. 2f), which suggests that subclonal competition associates with more aggressive tumour behaviour. 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



**Fig. 2 | Quantifying selective pressures during glioma evolution.**  
**a**, Schematic depiction of CCF values during tumour evolution indicating clonality and associated relative timing. **b**, Comparison of PyClone clusters ranked by CCF in matched initial and recurrent tumours. **c**, Left, dN/dS ratio for all variants (that is, global) in initial and recurrent tumours 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 time point, variant fraction and glioma subtype. Deviation from a linear relationship, significant Kolmogorov–Smirnov  $P$  values and Pearson’s  $R^2$  values 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 colours highlight the glioma subtype. Grey colouring reflects instances when treatment information was not available. **f**, Kaplan–Meier curve showing survival differences between IDH-wild-type recurrent tumours demonstrating selection ( $n = 39$ ) compared with neutrally evolving tumours ( $n = 44$ ).  $P$  value determined by log-rank test.

associated with shorter survival across subtypes (Hazard ratio = 1.53, 95% confidence interval 1.00–2.41,  $P = 0.048$ ; Supplementary Table 4). 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  $P > 0.05$ ; Supplementary Table 5), which suggests that standard therapeutic approaches for glioma have limited effect on the subclonal tumour architecture. Although high-depth sequencing datasets may be required to detect subtle selective effects<sup>26</sup>, our analyses raise the possibility that the survival benefit derived from standard chemoradiation results from the elimination of tumour cells in which treatment sensitivity of individual cells is not determined by genetic factors.

### Driver alteration frequencies across time

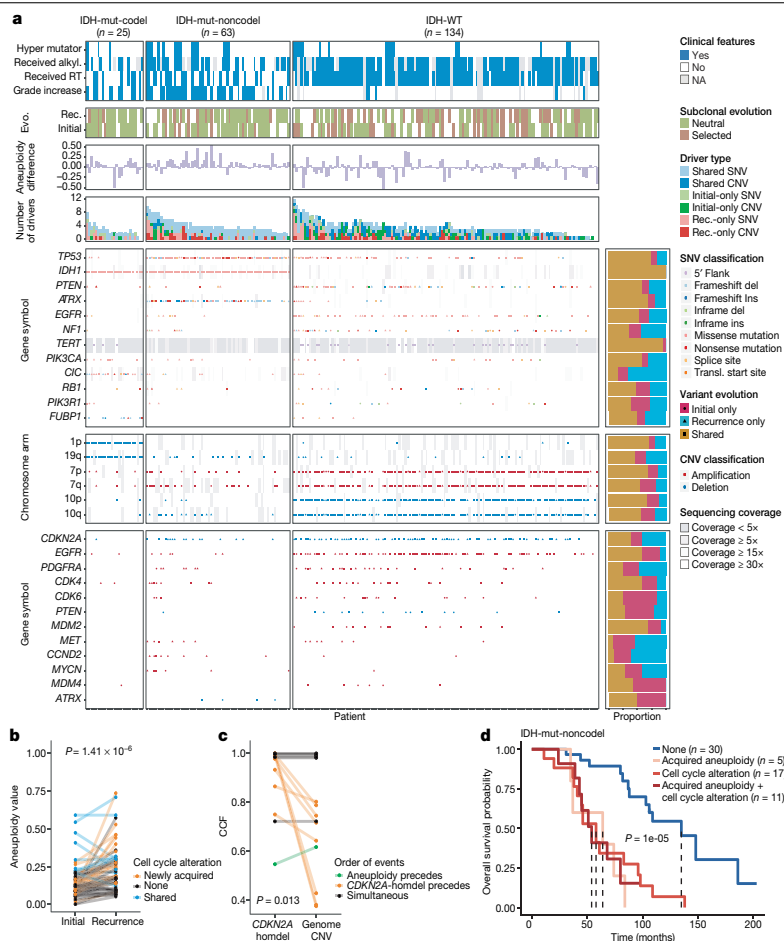
We evaluated how stability, acquisition and the loss of mutation and copy number drivers<sup>6</sup> over time affect glioma evolution. We used the dN/dS ratio to nominate 12 candidate mutation driver genes at both time points ( $Q < 0.05$ , Fig. 3a, Extended Data Fig. 7a) and determined significant alterations in copy number that recapitulated previously identified drivers (Extended Data Fig. 7b). Mutations in *IDH1* and

co-occurring loss of the 1p/19q chromosome arms have been suggested as glioma-initiating events<sup>1</sup>, which was corroborated by the observation that these events were not lost or acquired during the surgical interval (Fig. 3a, Extended Data Fig. 8a). Similarly, we observed that mutations in the *TERT* promoter were almost always shared in the IDH-mutant-codel and IDH-wild-type samples, although many samples lacked sufficient coverage in this GC-rich region. Chromosome 10 losses were present in a large majority of IDH-wild-type initial tumours and persisted into recurrence.

Shifts in the fraction of cancer cells containing 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 value (Extended Data Fig. 9). *ATRX* mutations in IDH-mutant-noncode initial tumours demonstrated lower CCFs than *TP53* ( $P = 0.03$ ) and *IDH1* ( $P = 0.10$ ) mutations, suggesting that *IDH1* and *TP53* mutations precede *ATRX* inactivation<sup>1</sup>. There was no difference in CCF values between *IDH1* and *TP53* among initial gliomas ( $P = 0.98$ ); however, *IDH1* mutations demonstrated significantly lower CCF values than *TP53* mutations ( $P = 0.0018$ ) in recurrent gliomas. We did not observe any CCF differences among driver mutations detected in IDH-wild-type tumours at either time point. Chromosome 10 deletion CCFs were higher than chromosome 7 amplifications ( $P = 0.0036$ ), which indicates that chromosome 10 deletions arise earlier<sup>33</sup>. Similarly,



Article



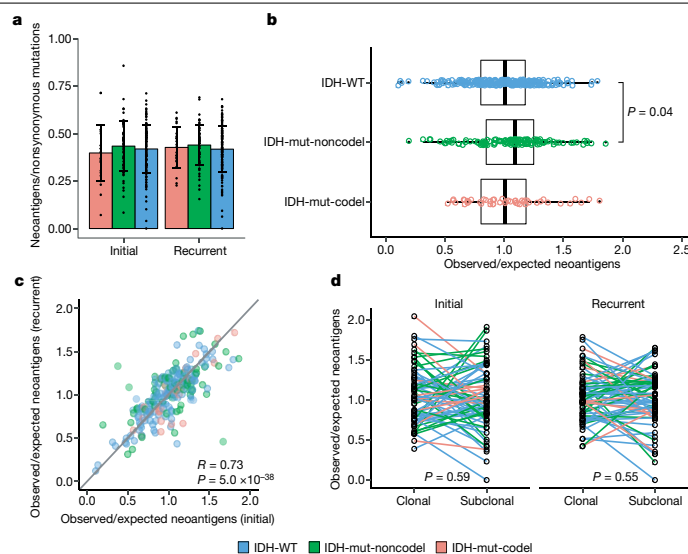
**Fig. 3 | Patterns of glioma driver frequencies over time.** **a**, Driver dynamics for single-nucleotide variants (SNVs) nominated by the dN/dS ratios and copy number alterations (CNVs) nominated by GISTIC ( $n = 222$ ). Each column represents a single patient at two separate time points stratified 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 colour. Stacked bar plots accompanying each gene/arm provide cohort-level proportions for whether the alteration was

shared, lost or acquired. Rec, recurrence; evo, evolution. **b**, Aneuploidy comparison in matching initial and recurrent IDH-mutant-noncodel tumours. **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.  $P$  value determined by Wilcoxon signed-rank test. **d**, Kaplan–Meier curve comparing survival in IDH-mutant-noncodel tumours with an alteration in the cell cycle, acquired aneuploidy, or both (shades of red) versus unaltered IDH-mutant-noncodel tumours (blue).  $P$  value determined by log-rank test.

there was no difference in CCF values between *CDKN2A* deletion and *EGFR* amplification ( $P = 0.70$ ). *EGFR* and chromosomal arm events significantly differed (that is, 10p del versus *EGFR* amp,  $P = 0.0019$ ) but not *CDKN2A* deletion and chromosomal events (that is, 10p del versus *CDKN2A* del,  $P = 0.33$ ). The consistently high CCF values 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*<sup>34</sup> 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 tumour and

recurrence (Extended Data Fig. 10a). A notable exception was the *TP53* mutation CCF that increased over time ( $P = 0.037$ ) in IDH-mutant-noncodels, but not IDH-wild-type gliomas ( $P = 0.13$ , Extended Data Fig. 10b). We did not observe any differences in *IDH1* CCF over time among IDH-mutant-noncodel tumours, possibly because the general trend of these tumours to increase in CCF is counteracted by the biological loss of relevance of mutant *IDH1* over time (Extended Data Fig. 10c). Indeed, a gross comparison of all shared mutation CCFs revealed an increase in recurrent IDH-mutant-noncodel tumours ( $P < 0.0001$ ), which may reflect increased clonality and a reduction in intratumoral heterogeneity (Extended Data Fig. 10d). By contrast, shared CCFs decreased in IDH-wild-type tumours, potentially indicating a general increase in intratumoral heterogeneity at recurrence ( $P < 0.0001$ , Extended Data



**Fig. 4 | Neoantigen selection during tumour progression.** **a**, Mean proportion of coding mutations giving rise to neoantigens (neoantigens/nonsynonymous mutations) stratified by glioma subtype and time point ( $n = 222$ ). Data are mean  $\pm$  s.d. **b**, Box plot depicting the distribution of observed-to-expected neoantigen ratios in the GLASS cohort stratified by glioma subtype.  $P$  value determined by  $t$ -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 tumours. Each point represents a single patient tumour pair.  $R$  denotes Pearson correlation coefficient. Panels **b** and **c** only include samples from pairs with at least three

neoantigens in the initial and recurrent tumours ( $n = 131, 63$  and  $24$  pairs for IDH-wild-type, IDH-mutant-noncode, and IDH-mutant-code, respectively). **d**, Ladder plot depicting the difference in observed-to-expected neoantigen ratio between a tumour's clonal and subclonal neoantigens. Each set of points connected by a line represents one tumour. Tumours are stratified by whether they were a patient's initial or recurrent tumour. Lines are coloured by each patient's glioma subtype. Panel **d** only includes samples from pairs with at least three clonal neoantigens and at least three subclonal neoantigens in both the initial and recurrent tumours ( $n = 35, 20$  and  $9$  for IDH-WT, IDH-mutant-noncode and IDH-mutant-code, respectively).  $P$  value determined by paired two-sided  $t$ -test. Colours in each panel represent the glioma subtype.

Fig. 10d). We confirmed that IDH-mutant-noncode CCF increases and IDH-wild-type decreases were not biased by patients with high mutational burden through the classification of patient-specific shared mutation CCF change (Extended Data Fig. 10e).

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 tumours, but there was no enrichment for somatic gene alterations in non-hypermutators, which suggests that glioma recurrence is not directed by particular sets of mutations (Extended Data Fig. 8b). Within subtypes, we detected an enrichment in *CDKN2A* homozygous deletions (Fig. 3a, Extended Data Fig. 8a) in recurrent IDH-mutant-noncode, which was corroborated by additional alterations to cell cycle genes (focal gain of *CCND2*, *CDK4* and *CDK6*, and mutation or homozygous loss of *RBI*). Mutations in cell cycle checkpoint control genes are associated with genomic instability<sup>35</sup>. Therefore, we analysed aneuploidy levels by determining the proportion of the genome that had undergone aneuploidy events (Extended Data Fig. 11a, b). We observed that IDH-mutant-noncode tumours had a higher level of aneuploidy at recurrence (Wilcoxon rank sum test  $P = 1.4 \times 10^{-6}$  total aneuploidy,  $P = 8.6 \times 10^{-3}$  arm-level aneuploidy; Extended Data Fig. 11c, d) with tumours carrying acquired cell cycle gene alterations displaying the largest increases in aneuploidy ( $P = 7.6 \times 10^{-6}$ ; Wilcoxon rank sum test, Fig. 3b). We reasoned that *CDKN2A* deletions may precede aneuploidy. Homozygous *CDKN2A* deletions had significantly higher CCFs than the average somatic copy number variation CCF across the genome (as a surrogate for aneuploidy-related copy number changes), suggesting that *CDKN2A* loss occurred before aneuploidy (Fig. 3c). These alterations may hasten

disease progression as patients with either alterations in cell cycle genes or the largest increases in aneuploidy at recurrence demonstrated significantly shorter survival than patients without these alterations (log-rank test  $P < 0.0001$ , Fig. 3d). Together, the persistence of drivers over time and the paucity of consistent change indicate 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 tumour 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 several cancer types, including glioma<sup>36-39</sup>, and suggests active immunosurveillance that may be therapeutically exploited<sup>40</sup>. We computationally predicted neoantigen-causing mutations<sup>41</sup>. 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 tumours ( $P > 0.05$ , Wilcoxon rank-sum test; Fig. 4a). The most common neoantigen arose from the clonal R132H mutation in *IDH1* and was present in 22 out of 88 IDH-mutant initial and recurrent tumours. Beyond mutations in *IDH1*, no mutations gave rise to a neoantigen found in more than three tumours at a given time point (Supplementary Table 6). Across the dataset, neoantigens and non-immunogenic mutations exhibited similar changes in CCF values between initial and

## Article

recurrent tumours indicating a lack of neoantigen-specific selection processes over time (Extended Data Fig. 12a).

We then examined the extent to which immunoeediting occurred by comparing the observed neoantigen rate of each sample to an expected rate that was empirically derived from our dataset. The output of this approach is a normally distributed set of ratios centred at 1. Samples with an observed-to-expected neoantigen ratio less than 1 exhibit evidence of neoantigen depletion relative to the rest of the dataset, and thus are more likely to have been immunoeedited. We found that none of the three glioma subtypes contained observed-to-expected ratios that significantly differed from 1 ( $P > 0.05$ , one sample  $t$ -test), although IDH-wild-type tumours exhibited significantly lower scores than IDH-mutant-noncodels ( $t$ -test,  $P = 0.04$ ; Fig. 4b). We also 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 tumours of each patient (Pearson's  $R = 0.73$ ,  $P = 5 \times 10^{-38}$ ), which suggests that the neoantigen depletion level in the recurrence reflects that of the initial tumour (Fig. 4c).

Immunoeediting is most likely to take place in the tumours with high cytolytic activity and low levels of immunosuppressive activity<sup>39</sup>. Hypermutators, which have high loads of neoantigens, have previously been associated with highly cytolytic microenvironments<sup>38</sup>. However, we did not observe any differences in the observed-to-expected neoantigen ratio between hypermutated recurrent tumours and their initial counterparts, nor did we observe differences between hypermutated and non-hypermutated recurrent tumours, indicating that immunoeediting activity is not related to the total number of mutations in a sample (Wilcoxon rank-sum test  $P > 0.05$ ; Extended Data Fig. 12b). To more directly determine whether there were immunological factors associated with neoantigen depletion, we analysed CIBERSORT immune cell fractions from a subset of samples that had undergone expression profiling in a previous study ( $n = 84$  from 42 tumour pairs)<sup>38,42</sup>. Initial tumours with an observed-to-expected neoantigen ratio greater than 1 exhibited significantly higher levels of CD4<sup>+</sup> T cells than those with a ratio less than 1, whereas recurrent tumours with a ratio greater than 1 exhibited significantly higher levels of macrophages and neutrophils, and significantly lower levels of plasma cells relative to those with a ratio less than 1 ( $P < 0.05$ , Wilcoxon rank-sum test; Extended Data Fig. 12c).

Although 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  $\rho = 0.28$ ,  $P = 2 \times 10^{-9}$ ), reflecting similar findings in lung cancer<sup>43</sup>. This may bias analyses comparing the ratio across patients. To determine whether immunoeediting varies over time in a patient-agnostic manner, we compared the observed-to-expected neoantigen ratio derived from the clonal mutations of a sample, which likely arose earlier in tumour evolution, to that derived from their subclonal mutations, which 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 tumour or recurrence ( $P > 0.05$ , paired  $t$ -test; Fig. 4d). 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 help to understand treatment failures and tumour progression. The longitudinal molecular profiles revealed common features such as acquired hypermutation and aneuploidy, and also highlighted the

individualistic paths of glioma evolution after treatment. Our results provide evidence that the current standard of care therapies do not frequently coerce glioma down predictable paths. Instead, an unexpected number of gliomas appeared to evolve stochastically after early driver events. We expect that continuing to profile patient tumours over time using comprehensive sequencing approaches will identify other common evolutionary paths. Our results highlight the 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 importance of glioma hypermutation uncertain<sup>21–24,27</sup>. Future analyses that consider the number of therapy cycles and *MGMT* DNA methylation status will help to determine factors that predispose tumours to hypermutation and identify therapies that effectively exploit the vulnerabilities of this phenotype (for example, high mutational burden). Acquired cell cycle alterations and aneuploidy in recurrent IDH-mutant-noncode gliomas also provide a rationale to target these more aggressive phenotypes with CDK inhibitors<sup>44</sup> or with compounds that disrupt microtubule dynamics<sup>45</sup>. Finally, our analyses revealed that immunoeediting activity does not vary in glioma over time, although we did observe variation between individual patients. Further molecular and immunological data are needed to fully understand the effect that this variability has on glioma evolution and to devise therapies directed at the glioma immune response<sup>17</sup>. To this end, we found that clonal neoantigens arising from the IDH1(R132H) mutation persisted from the initial tumour into the recurrence, justifying neoantigen vaccine approaches as treatments for initial and recurrent glioma<sup>46,47</sup>.

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 The Cancer Genome Atlas (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.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1775-1>.

1. Barthel, F. P., Wesseling, P. & Verhaak, R. G. W. Reconstructing the molecular life history of gliomas. *Acta Neuropathol.* **135**, 649–670 (2018).
2. Osuka, S. & Van Meir, E. G. Overcoming therapeutic resistance in glioblastoma: the way forward. *J. Clin. Invest.* **127**, 415–426 (2017).
3. Bettgowda, C. et al. Mutations in CIC and FUBP1 contribute to human oligodendroglioma. *Science* **333**, 1453–1455 (2011).
4. Zheng, S. et al. A survey of intragenic breakpoints in glioblastoma identifies a distinct subset associated with poor survival. *Genes Dev.* **27**, 1462–1472 (2013).
5. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
6. Ceccarelli, M. et al. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* **164**, 550–563 (2016).
7. The Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).
8. 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* **17**, 98–110 (2010).
9. Yan, H. et al. *IDH1* and *IDH2* mutations in gliomas. *N. Engl. J. Med.* **360**, 765–773 (2009).
10. Louis, D. N. et al. International Society of Neuropathology—Haarlem consensus guidelines for nervous system tumor classification and grading. *Brain Pathol.* **24**, 429–435 (2014).
11. Louis, D. N. et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* **131**, 803–820 (2016).
12. Venteicher, A. S. et al. Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**, eaai8478 (2017).
13. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).

# Mutational signatures of genotoxic cancer therapies

14. Snuderl, M. et al. Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. *Cancer Cell* **20**, 810–817 (2011).
15. Sottoriva, A. et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl Acad. Sci. USA* **110**, 4009–4014 (2013).
16. Williams, M. J. et al. Quantification of subclonal selection in cancer from bulk sequencing data. *Nat. Genet.* **50**, 895–903 (2018).
17. Nejo, T. et al. reduced neoantigen expression revealed by longitudinal multiomics as a possible immune evasion mechanism in glioma. *Cancer Immunol. Res.* **7**, 1148–1161 (2019).
18. The GLASS Consortium. Glioma through the looking GLASS: molecular evolution of diffuse gliomas and the Glioma Longitudinal Analysis Consortium. *Neuro-oncol.* **20**, 873–884 (2018).
19. Hu, H. et al. Mutational landscape of secondary glioblastoma guides met-targeted trial in brain tumor. *Cell* **175**, 1665–1678.e1618 (2018).
20. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
21. Wang, J. et al. Clonal evolution of glioblastoma under therapy. *Nat. Genet.* **48**, 768–776 (2016).
22. Kim, H. et al. Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Res.* **25**, 316–327 (2015).
23. Johnson, B. E. et al. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* **343**, 189–193 (2014).
24. Hunter, C. et al. A hypermutation phenotype and somatic MSH6 mutations in recurrent human malignant gliomas after alkylator chemotherapy. *Cancer Res.* **66**, 3987–3991 (2006).
25. Jolly, C. & Van Loo, P. Timing somatic events in the evolution of cancer. *Genome Biol.* **19**, 95 (2018).
26. Turajlic, S., Sottoriva, A., Graham, T. & Swanton, C. Resolving genetic heterogeneity in cancer. *Nat. Rev. Genet.* **20**, 404–416 (2019).
27. Choi, S. et al. Temozolomide-associated hypermutation in gliomas. *Neuro-oncol.* **20**, 1300–1309 (2018).
28. 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.* **17**, 1521–1532 (2016).
29. Buckner, J. C. et al. Radiation plus procarbazine, CCNU, and vincristine in low-grade glioma. *N. Engl. J. Med.* **374**, 1344–1355 (2016).
30. Yap, T. A., Gerlinger, M., Futreal, P. A., Pusztai, L. & Swanton, C. Intratumor heterogeneity: seeing the wood for the trees. *Sci. Transl. Med.* **4**, 127ps10 (2012).
31. Martincorena, J. et al. Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029–1041.e1021 (2017).
32. Williams, M. J., Werner, B., Barnes, C. P., Graham, T. A. & Sottoriva, A. Identification of neural tumor evolution across cancer types. *Nat. Genet.* **48**, 238–244 (2016).
33. 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* **35**, 692–704.e612 (2019).
34. deCarvalho, A. C. et al. Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. *Nat. Genet.* **50**, 708–717 (2018).
35. Giam, M. & Rancati, G. Aneuploidy and chromosomal instability in cancer: a jackpot to chaos. *Cell Div.* **10**, 3 (2015).
36. Marty, R., Thompson, W. K., Salem, R. M., Zanetti, M. & Carter, H. Evolutionary pressure against MHC class II binding cancer mutations. *Cell* **175**, 416–428.e413 (2018).
37. McGranahan, N. et al. Allele-specific HLA loss and immune escape in lung cancer evolution. *Cell* **171**, 1259–1271.e1211 (2017).
38. Wang, Q. et al. Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. *Cancer Cell* **32**, 42–56.e46 (2017).
39. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).
40. Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoevasion: from immunosurveillance to tumor escape. *Nat. Immunol.* **3**, 991–998 (2002).
41. Hundal, J. et al. pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. *Genome Med.* **8**, 11 (2016).
42. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
43. Rosenthal, R. et al. Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479–485 (2019).
44. 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.* **43**, 1360–1371 (2015).
45. 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.* **80**, 1209–1217 (2017).
46. Keskin, D. B. et al. Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature* **565**, 234–239 (2019).
47. Schumacher, T. et al. A vaccine targeting mutant IDH1 induces antitumour immunity. *Nature* **512**, 324–327 (2014).
48. Department of Neurosurgery, University Hospital Essen, Essen, Germany. 49. Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA. 50. Department of Neuro-Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. 51. CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria. 52. Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary. 53. Preston Robert Tisch Brain Tumor Center at Duke, Duke University Medical Center, Durham, NC, USA. 54. Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. 55. Broad Institute, Cambridge, MA, USA. 56. Department of Population and Quantitative Health Sciences, Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH, USA. 57. Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. 58. Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria. 59. Division of Neuro-Oncology, Massachusetts General Hospital, Boston, MA, USA. 60. Department of Pathology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. 61. Department of Neurosurgery, University of Liverpool & Walton Centre NHS Trust, Liverpool, UK. 62. Division of Neurosurgery, The University of Connecticut Health Center, Farmington, CT, USA. 63. Department of Cellular and Molecular Pathology, Leeds Teaching Hospital NHS Trust, St James's University Hospital, Leeds, UK. 64. Department of Radiation Oncology, The Ohio State Comprehensive Cancer Center—Arthur G. James Cancer Hospital, Columbus, OH, USA. 65. Department of Genetics and Genome Sciences, UConn Health, Farmington, CT, USA. 66. Yale University School of Public Health, New Haven, CT, USA. 67. Department of Neurosurgery, Brigham and Women's Hospital, Boston, MA, USA. 68. Department of Pathology & Laboratory Medicine, Medical College of Wisconsin, Milwaukee, WI, USA. 69. Department of Neurology, Medical College of Wisconsin, Milwaukee, WI, USA. 70. Department of Neurosurgery, University of California San Francisco, San Francisco, CA, USA. 71. Fondazione IRCCS Istituto Neurologico Besta, Milano, Italy. 72. Division of Molecular Genetics, Heidelberg Center for Personalized Oncology, German Cancer Research Consortium, German Cancer Research Center (DKFZ), Heidelberg, Germany. 73. Department of Neurology, Erasmus MC – University Medical Center Rotterdam, Rotterdam, The Netherlands. 74. Olivia Newton-John Cancer Research Institute, Austin Health, Melbourne, Victoria, Australia. 75. La Trobe University School of Cancer Medicine, Heidelberg, Victoria, Australia. 76. Neuro-Oncology Branch, National Institutes of Health, Bethesda, MD, USA. 77. Anatomic Pathology Service, Hôpital de l'Enfant-Jésus, CHU de Québec-Université Laval, Québec, Québec, Canada. 78. Department of Neurology, Columbia University Medical Center, New York, NY, USA. 79. Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY, USA. 80. Institute for Cancer Genetics, Columbia University Medical Center, New York, NY, USA. 81. Cooperative Trials Group for Neuro-Oncology (COGNO) NHMRC Clinical Trials Centre, The University of Sydney, Sydney, New South Wales, Australia. 82. Department of Neurology, Brain Tumor Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. 83. Department of Radiology, Medical College of Wisconsin, Milwaukee, WI, USA. 84. Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA, USA. 85. Department of Neurosurgery, Henry Ford Health System, Henry Ford Cancer Institute, Detroit, MI, USA. 86. Cure Brain Cancer Biomarkers and Translational Research Group, Prince of Wales Clinical School, University of New South Wales, Sydney, New South Wales, Australia. 87. Department of Neurosurgery, Sungkyunkwan University School of Medicine, Samsung Medical Center, Seoul, South Korea. 88. Institute for Refractory Cancer Research, Samsung Medical Center, Seoul, South Korea. 89. Department of Anatomical and Cellular Pathology, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, Hong Kong. 90. Department of Oncology, Luxembourg Institute of Health, Luxembourg, Luxembourg. 91. Department of Neurosurgery, University of Colorado School of Medicine, Aurora, CO, USA. 92. Department of Neurosurgery, Seoul National University College of Medicine, Seoul National University Hospital, Seoul, South Korea. 93. Department of Public Health Sciences, Henry Ford Health System, Henry Ford Cancer Institute, Detroit, MI, USA. 94. Department of Biomedical Informatics, Columbia University Medical Center, New York, NY, USA. 95. Department of Systems Biology, Columbia University, New York, NY, USA. 96. Department of Neurosurgery, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. 97. Institute of Neuro pathology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. 98. Department of Neurological Surgery, University Hospitals Cleveland Medical Center, Case Western Reserve University, Cleveland, OH, USA. 99. Department of Neurosurgery, Case Western Reserve University, Cleveland, OH, USA. 100. Seidman Cancer Center and Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH, USA. 101. Department of Radiology & Nuclear Medicine, Erasmus MC – University Medical Center Rotterdam, Rotterdam, The Netherlands. 102. The Hospital for Sick Children, Toronto, ON, Canada. 103. Interdisciplinary Division of Neuro-Oncology, Hertie Institute for Clinical Brain Research, DTK Partner Site Tübingen, Eberhard Karls University Tübingen, Tübingen, Germany. 104. Department of Neurosurgery, School of Medicine and Winship Cancer Institute of Emory University, Atlanta, GA, USA. 105. Institute of Cancer Genome Sciences, Department of Neurosurgery, University of Birmingham, Birmingham, UK. 106. Department of Neurology, University Hospital Zurich, Zurich, Switzerland. 107. Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands. 108. Department of Neurosurgery, Brain Tumor Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. 109. Department of Neurosurgery, Medical University of Vienna, Vienna, Austria. 110. Institute of Neurology, Medical University of Vienna, Vienna, Austria. 111. Division of Neurosurgery, Department of Surgery, University Health Network, Toronto, Ontario, Canada. 112. Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. 113. Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. 114. A list of participants and their affiliations appears in the online version of the paper. 115. These authors contributed equally: Floris P. Barthel, Kevin C. Johnson. \*e-mail: roel.verhaak@jax.org

© The Author(s), under exclusive licence to Springer Nature Limited 2019

<sup>1</sup>The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA. <sup>2</sup>Department of Pathology, Brain Tumor Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. <sup>3</sup>Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, UK. <sup>4</sup>DKFZ Division of Translational Neurooncology at the West German Cancer Center, German Cancer Consortium Partner Site, University Hospital Essen, Essen, Germany.

## Article

The Glioma Longitudinal Analysis (GLASS) Consortium

Floris P. Barthel<sup>1,2,7</sup>, Kevin C. Johnson<sup>1,7,2</sup>, Frederick S. Varn<sup>1</sup>, Anzhela D. Moskalik<sup>1</sup>, Georgette Tanner<sup>3</sup>, Emre Kocakavuk<sup>1,4,5</sup>, Kevin J. Anderson<sup>1</sup>, Kenneth Aldape<sup>6</sup>, Kristin D. Alfaro<sup>7</sup>, Samirkumar B. Amin<sup>1</sup>, David M. Ashley<sup>10</sup>, Pratiti Bandopadhyay<sup>11,12</sup>, Jill S. Barnholtz-Sloan<sup>13</sup>, Rameen Beroukhi<sup>12,14</sup>, Christoph Bock<sup>15,16</sup>, Priscilla K. Brastianos<sup>16</sup>, Daniel J. Brat<sup>17</sup>, Andrew R. Brodbelt<sup>18</sup>, Ketan R. Bulsara<sup>19</sup>, Aruna Chakrabarty<sup>20</sup>, Jeffrey H. Chuang<sup>1,22</sup>, Elizabeth B. Claus<sup>23,24</sup>, Elizabeth J. Cochran<sup>25</sup>, Jennifer Connelly<sup>26</sup>, Joseph F. Costello<sup>27</sup>, Gaetano Finocchiaro<sup>28</sup>, Michael N. Fletcher<sup>29</sup>, Pim J. French<sup>30</sup>, Hui K. Gan<sup>31,32</sup>, Mark R. Gilbert<sup>33</sup>, Peter V. Gould<sup>34</sup>, Antonio Iavarone<sup>35,36,37</sup>, Azzam Ismail<sup>20</sup>, Michael D. Jenkinson<sup>18</sup>, Mustafa Khasraw<sup>38</sup>, Hoon Kim<sup>1</sup>, Mathilde C. M. Kouwenhoven<sup>39</sup>, Peter S. LaViolette<sup>40</sup>, Peter Lichter<sup>29</sup>, Keith L. Ligon<sup>12,41</sup>, Allison K. Lowman<sup>40</sup>, Tathiane M. Malta<sup>42</sup>, Kerrie L. McDonald<sup>43</sup>, Annette M. Molinaro<sup>27</sup>, Do-Hyun Nam<sup>44,45</sup>, Ho Keung Ng<sup>46</sup>, Simone P. Nicolau<sup>47</sup>, Johanna M. Niers<sup>39</sup>, Houtan Noushmehr<sup>42</sup>, D. Ryan Ormond<sup>48</sup>, Chul-Keek Park<sup>49</sup>, Laila M. Poisson<sup>50</sup>, Raul Rabadan<sup>51,52</sup>, Bernhard Radlwimmer<sup>29</sup>, Ganesh Rao<sup>53</sup>, Guido Reifenberger<sup>54</sup>, Jason K. Sa<sup>45</sup>, Susan C. Short<sup>3</sup>, Peter A. Sillevs Smitt<sup>55</sup>, Andrew E. Sloan<sup>55,56,57</sup>, Marion Smits<sup>58</sup>, Hiromichi Suzuki<sup>59</sup>, Ghazaleh Tabatabai<sup>60</sup>, Erwin G. Van Meir<sup>41</sup>, Colin Watts<sup>62</sup>, Michael Weller<sup>63</sup>, Pieter Wesseling<sup>2,64</sup>, Bart A. Westerman<sup>65</sup>, Adelheid Woehrer<sup>67</sup>, W. K. Alfred Yung<sup>7</sup>, Gelareh Zadeh<sup>68</sup>, Jason T. Huse<sup>69,70</sup>, John F. De Groot<sup>7</sup>, Lucy F. Stead<sup>3</sup> & Roel G. W. Verhaak<sup>14</sup>

# Mutational signatures of genotoxic cancer therapies

## Methods

### Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

### 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 data 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 tumour samples, 17 standalone recurrences, and 19 patients with at least two geographically distinct tumour 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 1 fourth recurrence (Supplementary Table 7).

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 HiSeqX, whereas the NS and MD cohorts were sequenced using NovaSeq, according to Illumina's protocols. Whole-exome capture was performed using the following platforms as reported in previous publications<sup>7,21–23,48–52</sup>.

The Agilent SureSelect Human All Exon 50 Mb capture kit was used for patients SF-0001–SF-0021, and the Agilent SureSelect Human All Exon V4 capture kit was used for patients SF-0024–SF-0029 in the University of California San Francisco cohort. The Agilent SureSelect Human All Exon v4 or v5 kit was used to capture samples in the Kyoto University cohort. The Samsung Medical Center cohort reported using the Agilent SureSelect kit for patients SM-R056–SM-R071, SM-R075, SM-R076 and SM-R095–SM-R114, whereas the Illumina TruSeq Exome-capture kit was used for patient SM-R072. Exome capture was performed using the Agilent SureSelect Human All Exon 50 Mb kit in the TCGA glioblastoma (GBM) cohort and the Agilent SureSelect Human All Exon v2.0 44 Mb kit in the TCGA low grade glioma (LGG) cohort. Columbia University cases were captured using the Agilent V3 50 Mb kit, sequencing 90 bp paired-end reads for samples R009-TP, R009-R1, R011-TP, R011-R1, R014-TP, R014-R1, R017-R1, R018-R1 and R019-R1. Mapping files of initial tumour and normal samples of patients R017–R019 were obtained from the TCGA through the CG-hub. All other samples were captured using the Agilent SureSelect XT Human All Exon v.4 Kit, 80 million paired-end reads, 150× 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 using NextSeq in high output mode. In the Leeds Cohort (LU), the SureSelectXT V5 kit (PE100) was used to construct exome libraries. The Illumina TruSeq Exome capture kit was used for samples at the Medical University of Vienna – Research Center for Molecular Medicine (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 centre where the samples were either acquired or sequenced (Supplementary Table 7). This is followed by the four-character centre-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 are divided by a relative sample type, such as initial tumour (TP), recurrent tumour (R1), normal tissue (NB or NM, for example), or metastatic tumour 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). 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 sequencing (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<sup>53</sup>. Unless otherwise stated, all tools mentioned are part of the GATK 4 suite<sup>54</sup>. All data were collected at a central location (The Jackson Laboratory) and analysed using homogenous pipelines capable of processing raw fastq files as well as re-processing previously analysed bam files.

### Alignment and pre-processing

Data pre-processing was conducted in accordance to the GATK Best Practices using GATK 4.0.10.1. In brief, aligned BAM files were separated by readgroup, 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 optical duplicates across identical sequencing libraries. Lastly, base recalibration was performed using 'BaseRecalibrator' followed by 'ApplyBQSR'. Coverage statistics were gathered using 'CollectWgsMetrics'. Alignment quality control was performed running 'ValidateSamFile' on the final BAM file and quality control results were inspected using 'MultiQC 1.6a0<sup>55</sup>'. A haplotype database for fingerprinting was generated using a modified version of the code on [https://github.com/naumanjaved/fingerprint\\_maps](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 artefact detection mode and pooled into a cohort-wide panel of normals. Somatic variants were subsequently called in individual tumour 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 tumour and matching control

## Article

samples. Read orientation artefacts were evaluated using 'CollectFIR2Counts' 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<sup>56</sup>. 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 Functator and the v1.6.20190124s annotation data source. Allele frequencies from multi-sample Mutect2 were used to compare allele frequencies 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 Mb sequenced. A minimum coverage threshold of 15× was required for each base. DNA hypermutation was defined for recurrent tumours 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 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 non-negative-least squares problems 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 whole 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 1-kb padding and to merge overlapping intervals. In parallel, 'SelectVariants' was used to subset the gnomAD resource of germline variants to variants with a population allele frequency 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 allele frequency greater than 5%. The cohort was subsequently split into batches determined by sequencing centre and 'CreateReadCountPanelOfNormals' was used to create a panel of normal for each batch. Panel of normals were created separately for allosomes and autosomes, and allosomes were separated further by sex. To improve the panel of normals further, 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 panel of normal 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 'PlotModelledSegments' 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<sup>57,58</sup>.

Segments (from 'ModelSegments') with a non- $\log_2$  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 non- $\log_2$  copy ratio (once-filtered) were determined again. Outlier segments are removed and once again a weighted mean and standard deviation non- $\log_2$  copy ratio (twice-filtered) were determined. Segments with a non- $\log_2$  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 standard deviation of the non- $\log_2$  copy ratio (once-filtered) was then determined individually for each chromosome arm. Outlier segments were removed and the weighted mean and standard deviation of the non- $\log_2$  copy ratio (twice-filtered) was determined again. 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 standard deviation of the selected arms were used as high-level thresholds.

Gene level copy numbers were called by intersecting the gene boundaries with the segment intervals and by calculating the weighted non- $\log_2$  copy ratio for that gene. The copy number call for that gene was then determined by comparing the gene-level non- $\log_2$  copy ratio to the previously determined thresholds.

### dNdScv

The dN/dS ratios were estimated using the R package dNdScv<sup>31</sup> (<https://github.com/im3sanger/dndscv>) was run using the default and recommended parameters for all mutations in initial tumour samples, recurrent tumour samples, and for each mutational fraction (unique to initial, unique to recurrent and shared). All analyses were conducted separately within the three main tumour 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 modelled after a previously described method was computed<sup>59</sup>. In brief, 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 -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 v.0.0.2, <https://github.com/marcjwilliams1/neutralitytestr>)<sup>32</sup>. Individual variant allele frequency vectors were merged at the level of glioma subtype by variant status. Only mutations found in copy-neutral regions were included in these analyses. For all else, default parameters were used. Merged variant allele frequency distributions were deemed to be selected when the neutral null hypothesis was rejected using several metrics. Tests for neutrality required that both  $R^2 < 0.98$  and the area between the two curves of (1) merged variant allele frequency 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 tumour samples (Julia package: SubclonalSelection, <https://github.com/marcjwilliams1/SubClonalSelection.jl>)<sup>36</sup>. Patients that had samples at both time points with a TITAN-defined purity estimate  $\geq 0.5$  and  $\geq 25$  subclonal mutations in 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 1,000 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<sup>36</sup>. Classification based on the highest model probability yielded stable results as 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 IDH-wild-type tumours that were selected had a worse overall survival ( $P = 0.03$  ( $n = 81$ ),  $P = 0.01$  ( $n = 66$ ) and  $P = 0.01$  ( $n = 56$ ), respectively).

## Mutation clonality

Each patient's clonal architecture was inferred using PyClone (v.0.13.1) by grouping SNVs into clonal clusters (<https://github.com/arothon85/pyclone>)<sup>66</sup>. The patient-level input mutation matrix was reduced by limiting to sites with at least  $30\times$  coverage across all samples. PyClone was subsequently run using a binomial density model, connected initiation, and 10,000 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 1,000, 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 CCF values were  $\geq 0.5$ , subclonal for mutations with  $CCF \geq 0.1$  and  $CCF < 0.5$ , mutations were considered non-clonal when  $CCF < 0.1$ , as previously described<sup>61</sup>.

## CNV clonality

Allele-specific copy number, tumour purity and ploidy estimates were derived using a probabilistic model (TITAN, v.1.19.1) for both whole-genome and whole-exome sequencing samples<sup>62</sup>. TITAN was supplied with the tumour denoised read counts output by GATK DenoiseReadCounts and the tumour allelic counts at loci found to be heterozygous in control samples output by ModelSegments. An 'alphaK' (and 'alphaKHigh') parameter of 2,500 and 10,000 was used for exomes and genomes, respectively. The patient sex was provided to improve fitting allosomes. For each tumour–control pair, TITAN was run assuming an initial ploidy of two or three, and assuming one to three clusters, resulting in a total of six possible solutions per tumour/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 (half-maximal inhibitory concentration,  $IC_{50}$ ) that was  $\leq 500$  nM and better than or equal to the wild-type form of the peptide. Each patient's four-digit HLA class I types were inferred using OptiType (v.1.3.1, <https://github.com/FRED-2/OptiType>) run on each patient's matched normal sample<sup>63</sup>. VCF files for each tumour sample were annotated using Variant Effect Predictor (ensembl) with the 'downstream' and 'wildtype' plugins. Neoantigens from these VCFs were then called using pVACseq (v.4.0.10, <https://github.com/griffithlab/pVAC-Seq>)<sup>41</sup> run using netMHCpan (v.2.8, <http://www.cbs.dtu.dk/services/NetMHCpan-2.8/>)<sup>64</sup>. 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 tumour's observed neoantigen rate to an empirically derived expected rate that assumes no selection against neoantigen-causing mutations<sup>39</sup>. From the gold set samples in the GLASS cohort ( $n = 222$ ), define  $\bar{N}_s$  to be the expected number of nonsynonymous missense SNVs per synonymous SNV with trinucleotide context  $s$ .  $\bar{B}_s$  is then defined as the expected 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,  $N_{pred,i}$ , and neoantigen-causing mutations,  $B_{pred,i}$ , 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 the final neoantigen depletion ratio,  $R_i$ , of sample  $i$ , 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 the SNVs of a sample by the respective criteria and then recalculating the ratio as described above. To mitigate overfitting, all analyses presented here used samples from patients with at least three neoantigen-causing mutations in their primary and recurrent tumours.

## Immune cell analyses

CIBERSORT relative immune cell fraction data used in downstream neoantigen analyses were downloaded from a previous publication<sup>38</sup>.



# Mutational signatures of genotoxic cancer therapies

## Article

### 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.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

### Data availability

All de-identified, 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 in the Supplementary Information.

### Code availability

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

48. Brennan, C. W. et al. The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
49. 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* **142**, 1620–1626 (2018).
50. Mazar, T. et al. DNA methylation and somatic mutations converge on the cell cycle and define similar evolutionary histories in brain tumors. *Cancer Cell* **28**, 307–317 (2015).
51. Kim, J. et al. Spatiotemporal evolution of the primary glioblastoma genome. *Cancer Cell* **28**, 318–328 (2015).
52. Suzuki, H. et al. Mutational landscape and clonal architecture in grade II and III gliomas. *Nat. Genet.* **47**, 458–468 (2015).
53. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **34**, 3600 (2018).
54. 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* **43**, 11.10.11–11.10.33 (2013).
55. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
56. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
57. 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.* **12**, R41 (2011).
58. Beroukhi, R. et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl Acad. Sci. USA* **104**, 20007–20012 (2007).
59. Taylor, A. M. et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* **33**, 676–689.e673 (2018).
60. Roth, A. et al. PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11**, 396–398 (2014).
61. Turajlic, S. et al. Tracking cancer evolution reveals constrained routes to metastases: TRACERx Renal. *Cell* **173**, 581–594.e512 (2018).
62. Ha, G. et al. TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome Res.* **24**, 1881–1893 (2014).
63. Szolek, A. et al. OdiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* **30**, 3310–3316 (2014).
64. Hoof, I. et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* **61**, 1–13 (2009).

**Acknowledgements** This work 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 (CPRI) 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., T.M.M., H.N.), NCI R01CA222146 (L.M.P., H.K.N.), NCI R01CA230031 (J.H.C., J.N.), NCI R01CA188288 (J.S.B.-S., R.B., P.B., K.L.L., A. Chakravarty, A.E.S.), R01CA179044 (A. Iavarone), U54CA193313 (A. Iavarone), The National Brain Tumor Society (W.K.A.Y., J.F.d.G.). Brain Tumour Northwest tissue bank (including the Walton research tissue bank) is supported by the Sidney Driscoll Neuroscience Foundation and part of the Walton Centre and Lancashire Teaching Hospitals NHS Foundation Trusts (A.F.B., M.D.J.). This work was supported by a generous gift from the Dabbieri family (J.F.C.). Support is also provided by a Leeds Charitable Foundation grant (9R11/14-11 to L.F.S.), University of Leeds Academic Fellowship (11001061) (L.F.S.) and Studentship (11061191) (G. Tanner) as well as Leeds Teaching Hospitals NHS Trust (A. Chakravarti, A. 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-000590; B.A.W., 200450). Ghazaleh Tabatabai 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. Reifenberger, P.L.) and German Cancer Aid (DKH) 70-3163-W3 (M.W.). GLASS-NL receives support from KWF/Dutch Cancer Society project 11026 (M.C.M.K., P.W., R.G.W.V., P.J.F., J.M.N., M. Smits, B.A.W.). 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, G. Rao), 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-1741-01-DMC). We thank the Jackson Laboratory Clinical and Translation Support team for coordinating all data transfer agreements. We thank M. Wimsatt for assistance in graphic design.

**Author contributions** Sequencing data coordination was performed by H.K., F.P.B. and K.C.J., and clinical data coordination was by A.D.M. and O.A. Data analysis was led by F.P.B. and K.C.J. in collaboration with K.J.A., S.B.A., J.H.C., H.K., E.K., J.N., L.F.S., G. Tanner, F.S.V. and R.G.W.V. Clinical analysis was performed by F.P.B., K.C.J., A.D.M., L.M.P. and C.W. Pathology review was completed, in part, by A. Chakravarty, J.T.H., A. Ismail, A.W., H.K.N., K.L.L., G. Reifenberger and K.A. F.P.B., K.C.J., A.D.M., F.S.V. and R.G.W.V. wrote the manuscript. K.D.A., J.H. and J.F.d.G. coordinated the GLASS-MDACC cohort. L.F.S. was the lead coordinator of the GLASS-Leeds cohort and B.A.W. the lead coordinator of GLASS-Netherlands. D.M.A., D.A., P.B., J.S.B.-S., R.B., C.B., P.K.B., D.J.B., A.R.B. A. Chakravarty, A. Chakravarti, E.J.C., J.F.C., G.F., M.M.F., A. Iavarone, M.D.J., M.K., P.S.L., M.L., P.L., K.L.L., T.M.M., T.M., A.M.M., D.-H.N., N.N., H.K.N., C.Y.N., S.P.N., H.N., D.R.O., C.-K.P., L.M.P., G. Rao, B.R., J.K.S., S.C.S., A.E.S., M. Schuster, L.F.S., H.S., E.G.V.M., C.W., M.W., G.W. and A.W. contributed to sample acquisition and processing. All co-authors including K.A., P.B., A.F.B., K.R.B., E.B.C., J.C., P.J.F., H.K.G., M. R. Grimmer, P.V.G., M. R. Gilbert, A.K.L., K.L.M., J.M.N., R.R., G. Reifenberger, B.L.S., P.A.S.S., M. Smits, G. Tabatabai, P.W., W.K.A.Y. and G.Z. discussed the results and commented on the manuscript and Supplementary Information. R.G.W.V. was the project lead and coordinator.

**Competing interests** 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 Merck and Genentech-Roche. W.K.A.Y. serves in a consulting or advisory role at DNATrx 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. Reifenberger receives funding from Roche and Merck (EMD, Darmstadt) as well as honoraria from AbbVie. M. Smits is a central reviewer for Parexel Ltd and honoraria are paid to the institution. G. Tabatabai 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).

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-019-1775-1>.

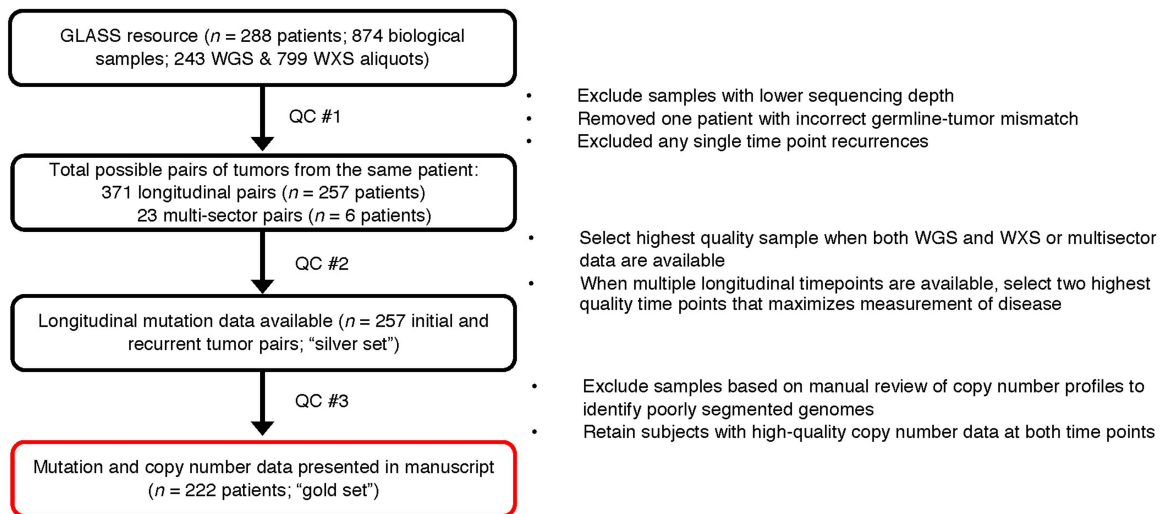
**Correspondence** and requests for materials should be addressed to R.G.W.V.

**Peer review information** Nature thanks Kamila Naxerova, Wolfgang Wick and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

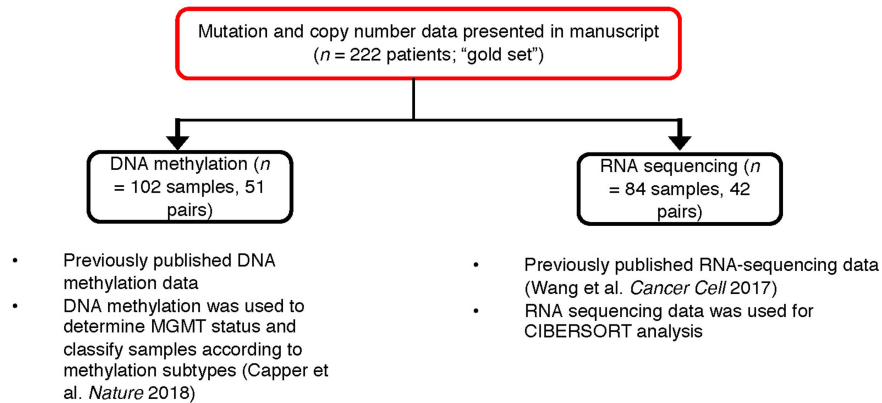
**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

# Mutational signatures of genotoxic cancer therapies

## a Quality Control (QC) filters



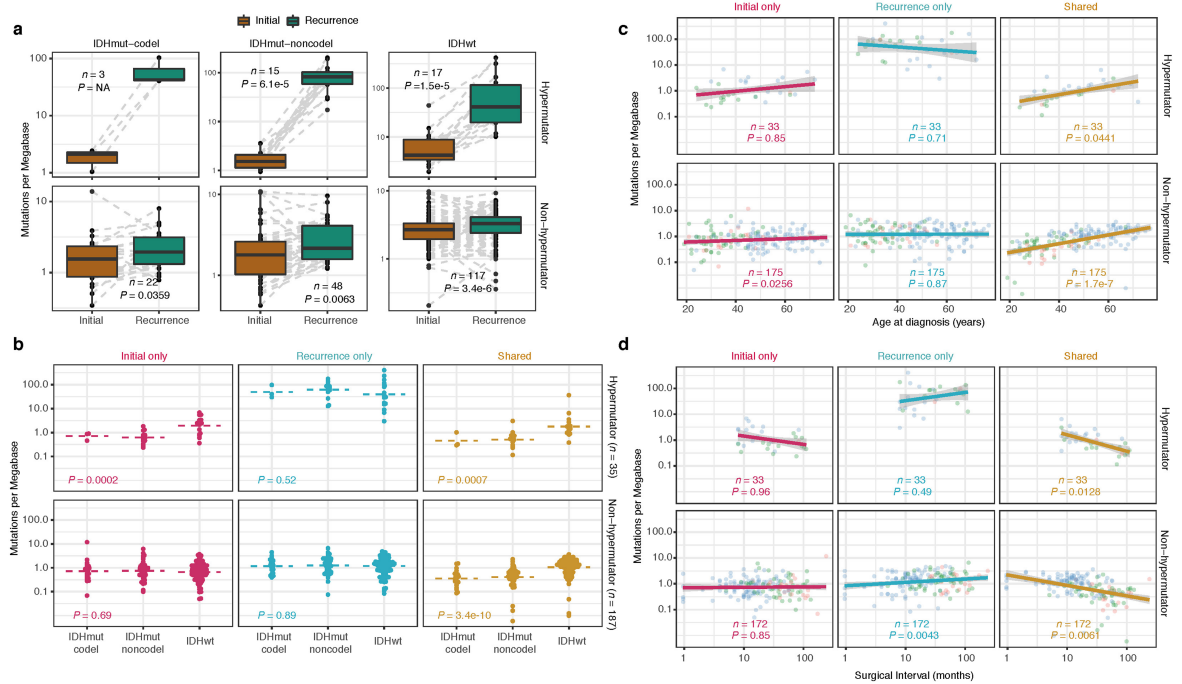
## b Overlapping datasets



**Extended Data Fig. 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.

# Mutational signatures of genotoxic cancer therapies

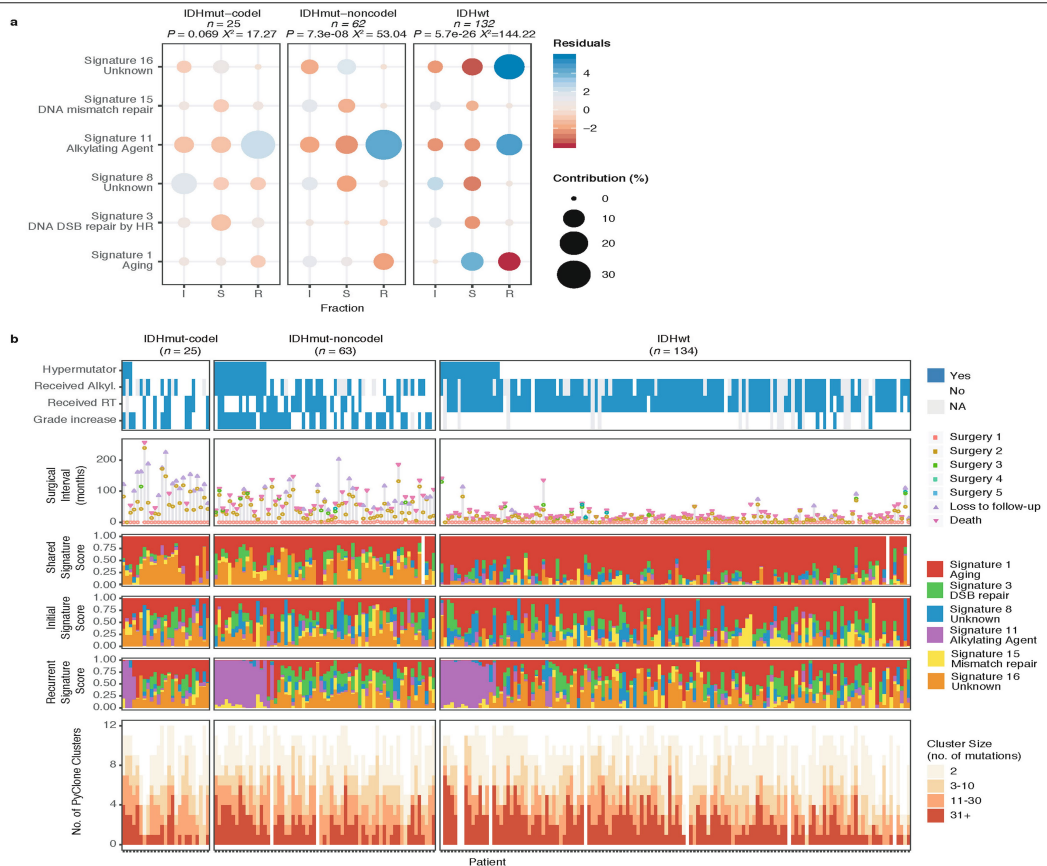
## Article



**Extended Data Fig. 2 | Mutation burden by time point and subtype.** **a**, Box plots 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.  $P$  values comparing three subtypes were determined by

one-way analysis of variance (ANOVA). **c**, Scatter plot showing the relationship between age at diagnosis and coverage adjusted mutation burdens by subtype and fraction.  $P$  values were determined by the linear model and adjusted by subtype. **d**, Similar to the analysis in **c** but showing the relationship between time to recurrence and coverage-adjusted mutation burdens.

# Mutational signatures of genotoxic cancer therapies

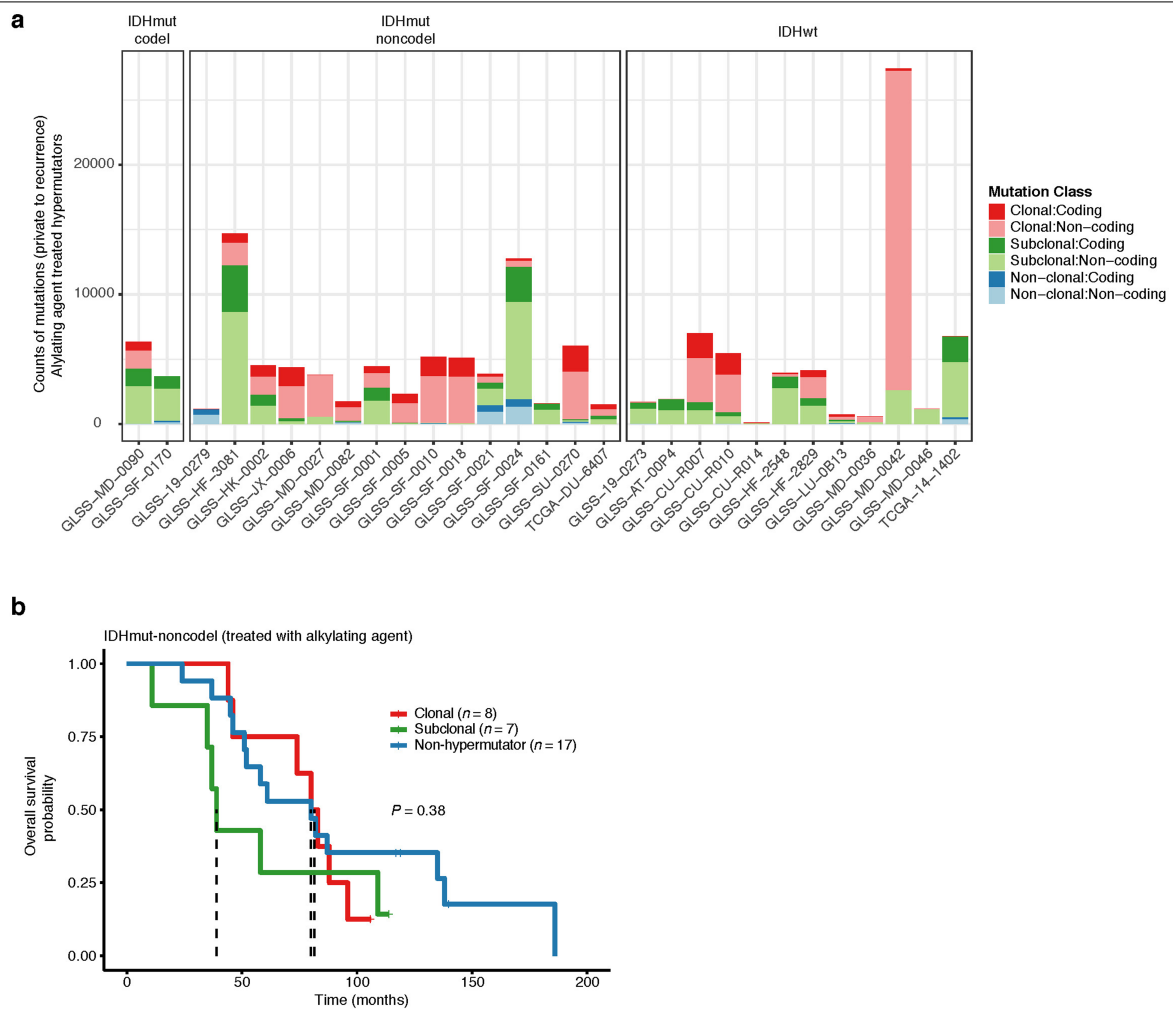


**Extended Data Fig. 3 | Mutational signatures by fraction and subtype.**  
**a.** Correlation plot showing the Pearson's chi-squared ( $\chi^2$ ) residuals for each signature by fraction and subtype. A  $\chi^2$  test was performed for each subtype and  $P$  values are indicated. Positive residuals (blue) indicate a positive correlation, whereas negative residuals (red) indicate an anti-correlation. The

point size reflects the contribution to the  $\chi^2$  estimate. **b.** Patients were ordered as in Fig. 1a, and relevant clinical information is provided alongside the fraction-specific mutational signatures. PyClone mutational clusters are also presented.

# Mutational signatures of genotoxic cancer therapies

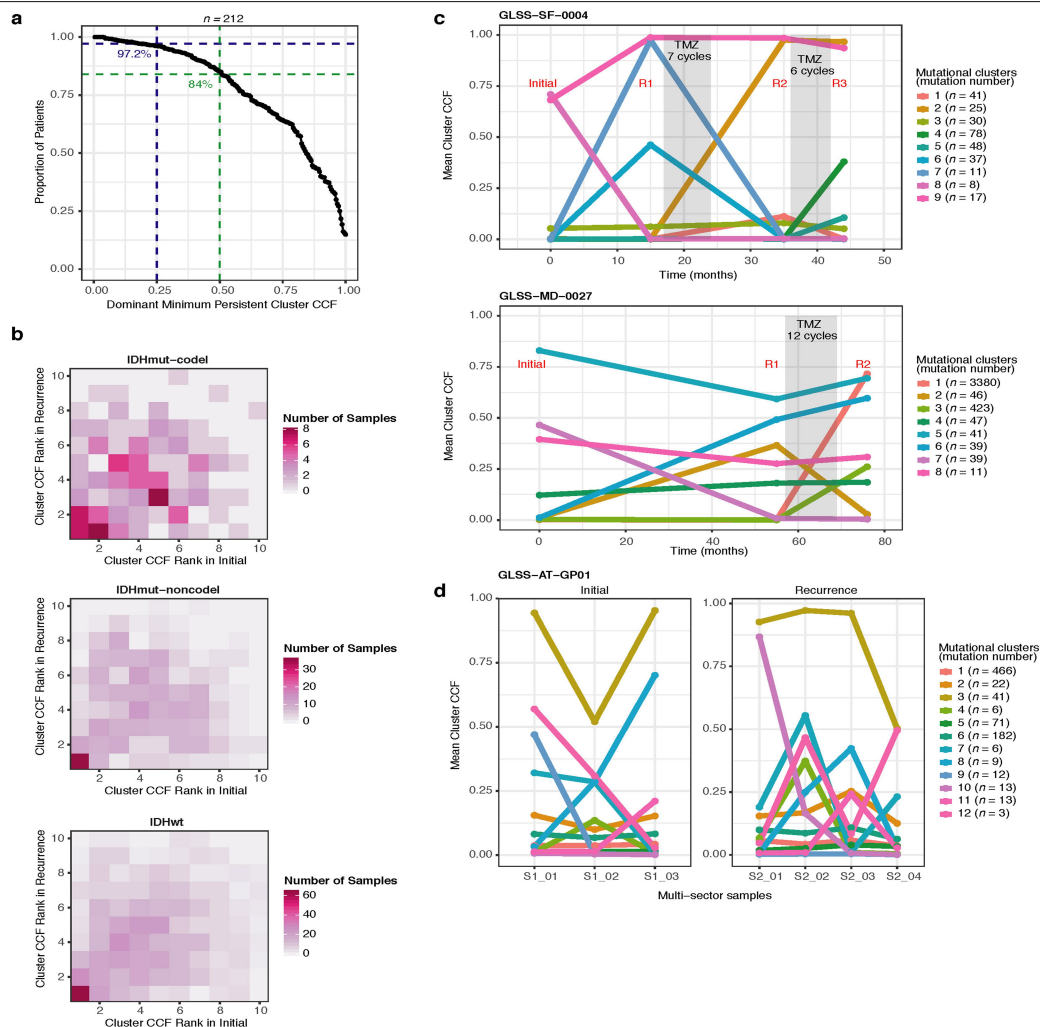
## Article



**Extended Data Fig. 4 | Hypermutator clonality. a**, Bar plots represent counts of recurrence-only mutations per hypermutator tumour that were known to receive treatment alkylating agent and were successfully run through the PyClone algorithm. Colours indicate mutation clonality and colour intensity indicates whether the mutations resulted in coding changes. **b**, Kaplan-Meier

curve comparing the survival of alkylating agent-treated IDH-mutant-noncodel hypermutator tumours that were predominantly clonal ( $n=8$ ), predominantly subclonal ( $n=7$ ) or non-hypermutator ( $n=17$ ). Limited to tumours with available PyClone data.  $P$  value determined by log-rank test.

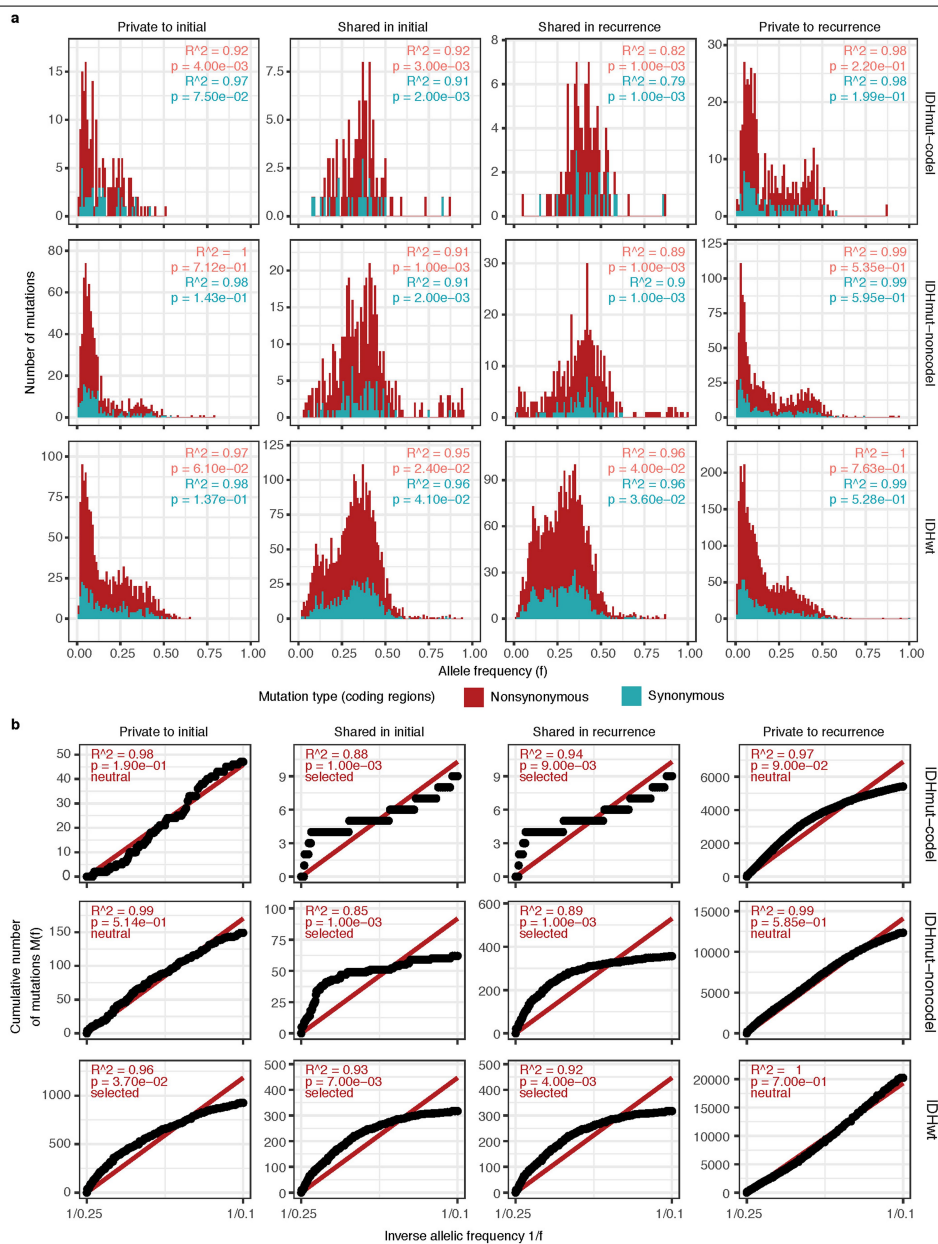
# Mutational signatures of genotoxic cancer therapies



**Extended Data Fig. 5 | Clonal structure evolution over time.** **a**, The minimum CCF of the most persistent (shared between initial and recurrence) PyClone cluster. **b**, Comparison of PyClone clusters ranked by CCF in matched initial and recurrent tumours, as in Fig. 2b, but separated by subtype. **c**, **d**, Examples of

cluster CCF dynamics over time in three separate samples, including two multi-time point samples (**c**) and one multi-sector sample (**d**). These additional data are available in the GLASS resource, but only two time-separated samples were used throughout to ensure clarity.

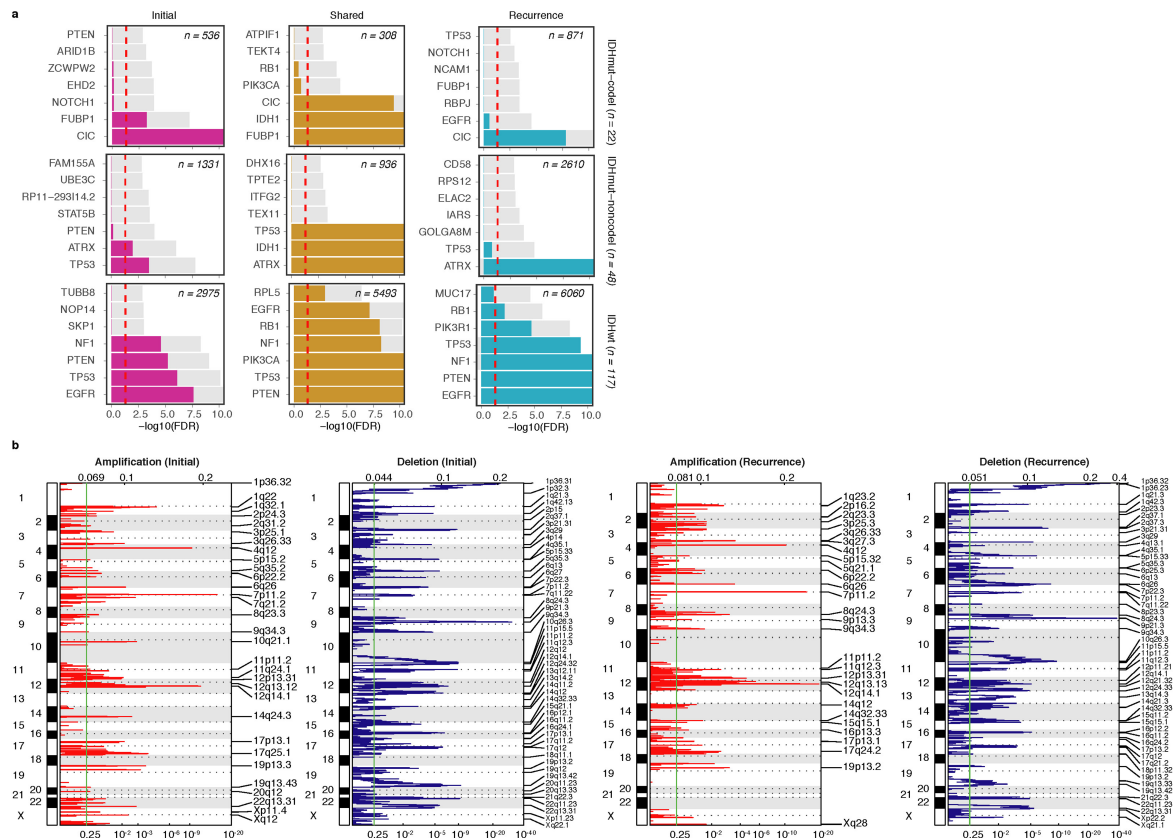
## Article



**Extended Data Fig. 6 | Distribution of variant allele fraction. a.** Distributions of non-hypermutator variant allele fraction for copy-neutral variants in coding regions ( $n=181$  patients). Variants are separated by subtype, fraction and also the variant was non-synonymous or synonymous mutation in a coding region.  $R^2$  goodness-of-fit measure and associated  $P$  values are shown. Note that these

data consider only the coding portion of genome, whereas Fig. 2d presents both coding and non-coding data. **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.

# Mutational signatures of genotoxic cancer therapies

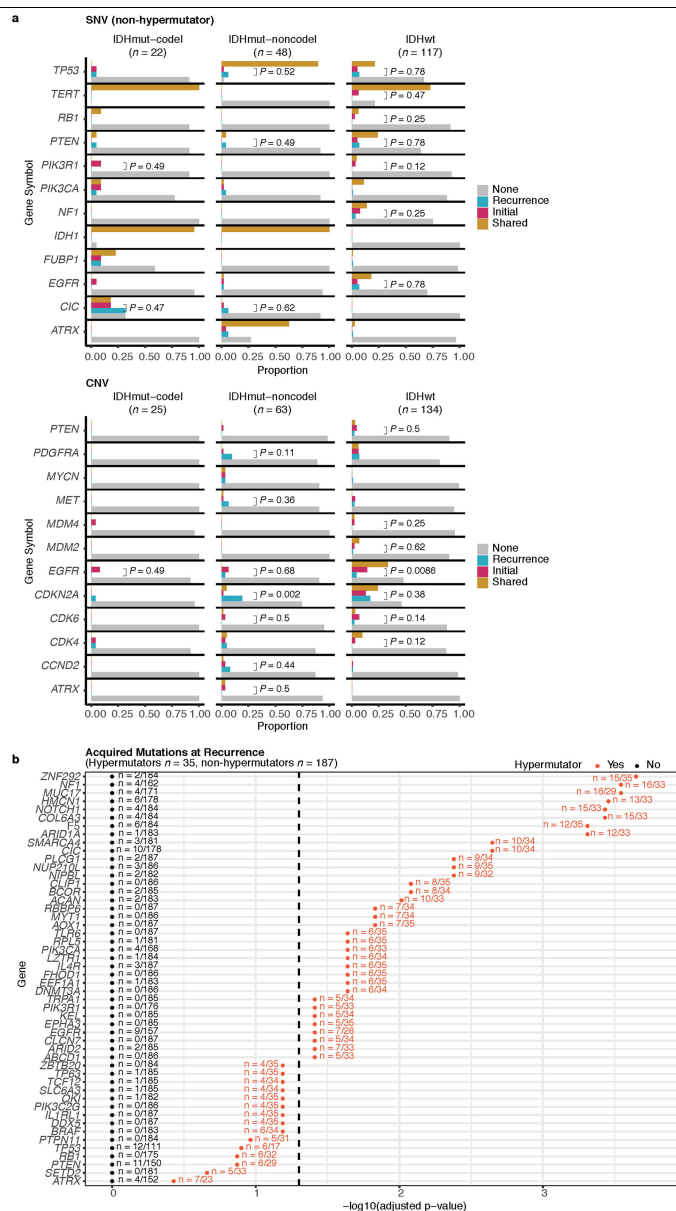


**Extended Data Fig. 7 | Driver gene nomination. a.** Local (gene-wise) dN/dS estimates by subtype (rows) and fraction (columns). Genes are sorted by *Q* value and *P* value. The *Q* value is shown in colour, whereas the *P* value is indicated in light grey. The *Q* value threshold of 0.05 is indicated by a horizontal

red line. **b.** GISTIC significant amplification (red) and deletion (blue) plots in initial (left) and recurrent tumours (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.



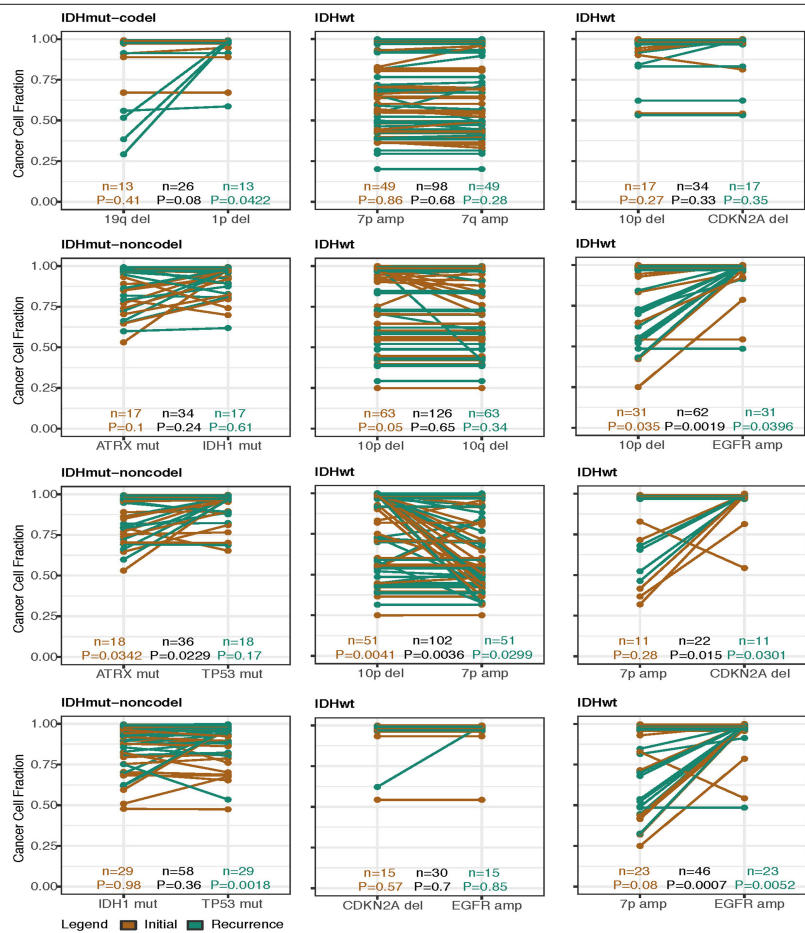
## Article



**Extended Data Fig. 8 | 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 determined by 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 among previously implicated glioma drivers testing for driver acquisition. *P* values were adjusted for multiple testing using the false discovery rate (x axis). Hypermutators (red) and non-hypermutators (black) were separately analysed.

# Mutational signatures of genotoxic cancer therapies

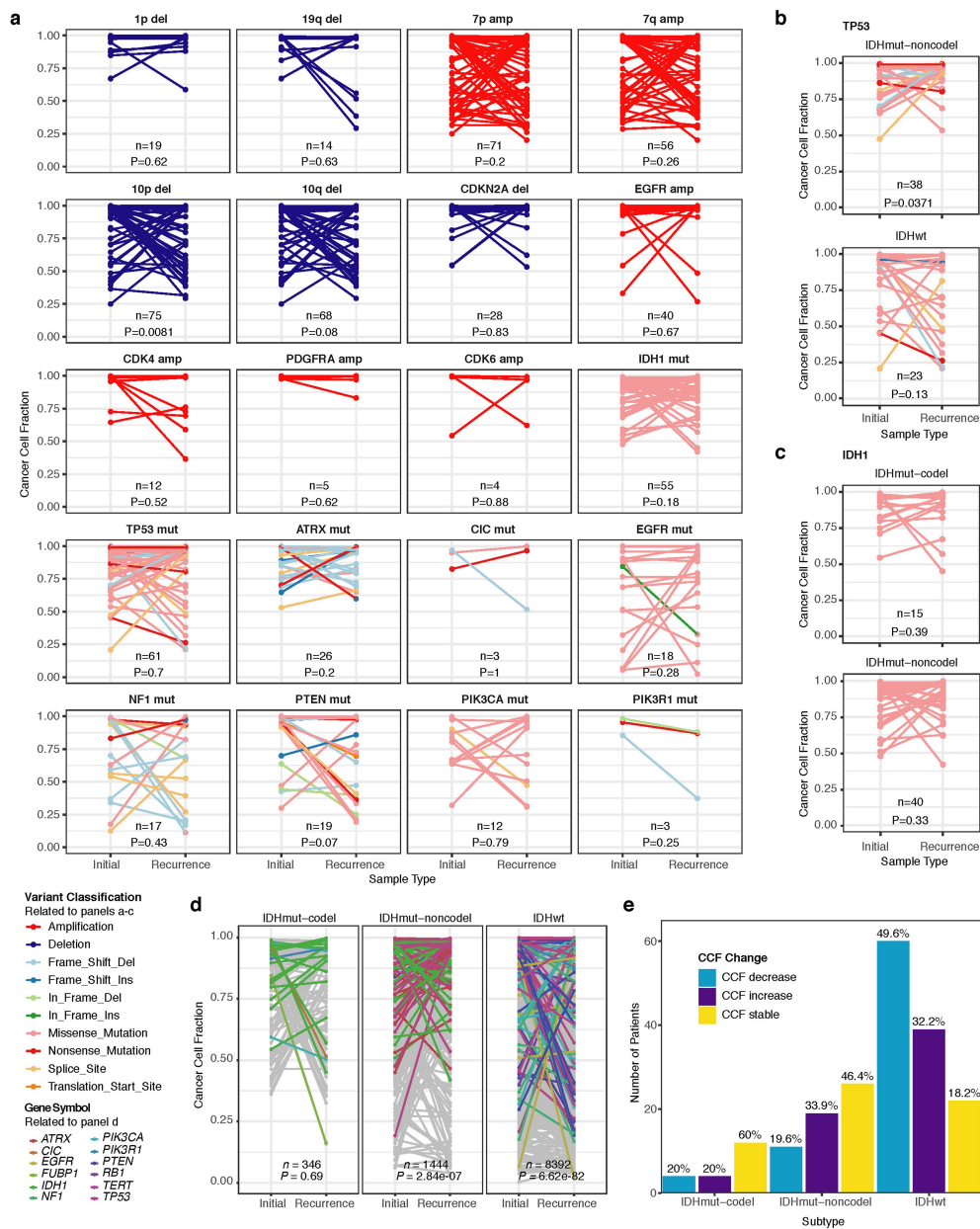


**Extended Data Fig. 9 | Intra-tumour CCF comparison.** Ladder plots comparing the CCF of co-occurring drivers in single tumour samples. The colour of the lines and points indicates whether the sample shown is an initial

(brown) or recurrent (green) tumour. *P* values determined by two-sided Wilcoxon rank-sum test for all initial samples, recurrent samples, as well as all samples (black).

# Mutational signatures of genotoxic cancer therapies

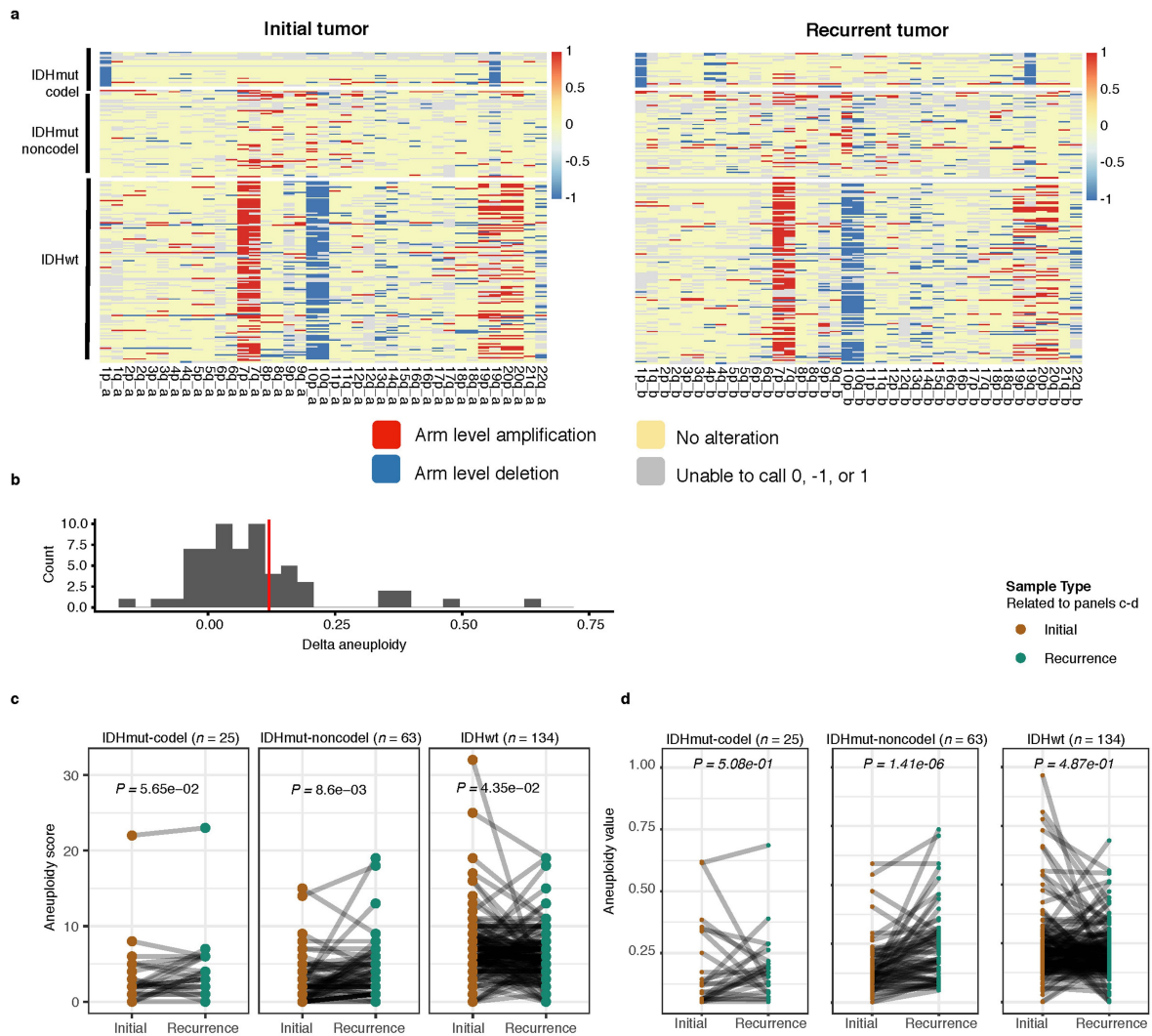
## Article



**Extended Data Fig. 10 | Between time point intra-patient CCF comparison.** **a**, Driver gene CCF comparison between initial and matched recurrences. Lines are coloured by variant classification. *P* values determined by two-sided Wilcoxon rank-sum test. **b**, *TP53* CCF by subtype, otherwise as in **a**. **c**, *IDH1* CCF by subtype, otherwise as in **a**. **d**, Ladder plot visualizing change in CCF across all SNVs between initial and recurrent tumours, separated by subtype. *P* values

determined by Wilcoxon rank-sum test. **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, and those with a significant increase or decrease are shown in dark and light blue, respectively.

# Mutational signatures of genotoxic cancer therapies

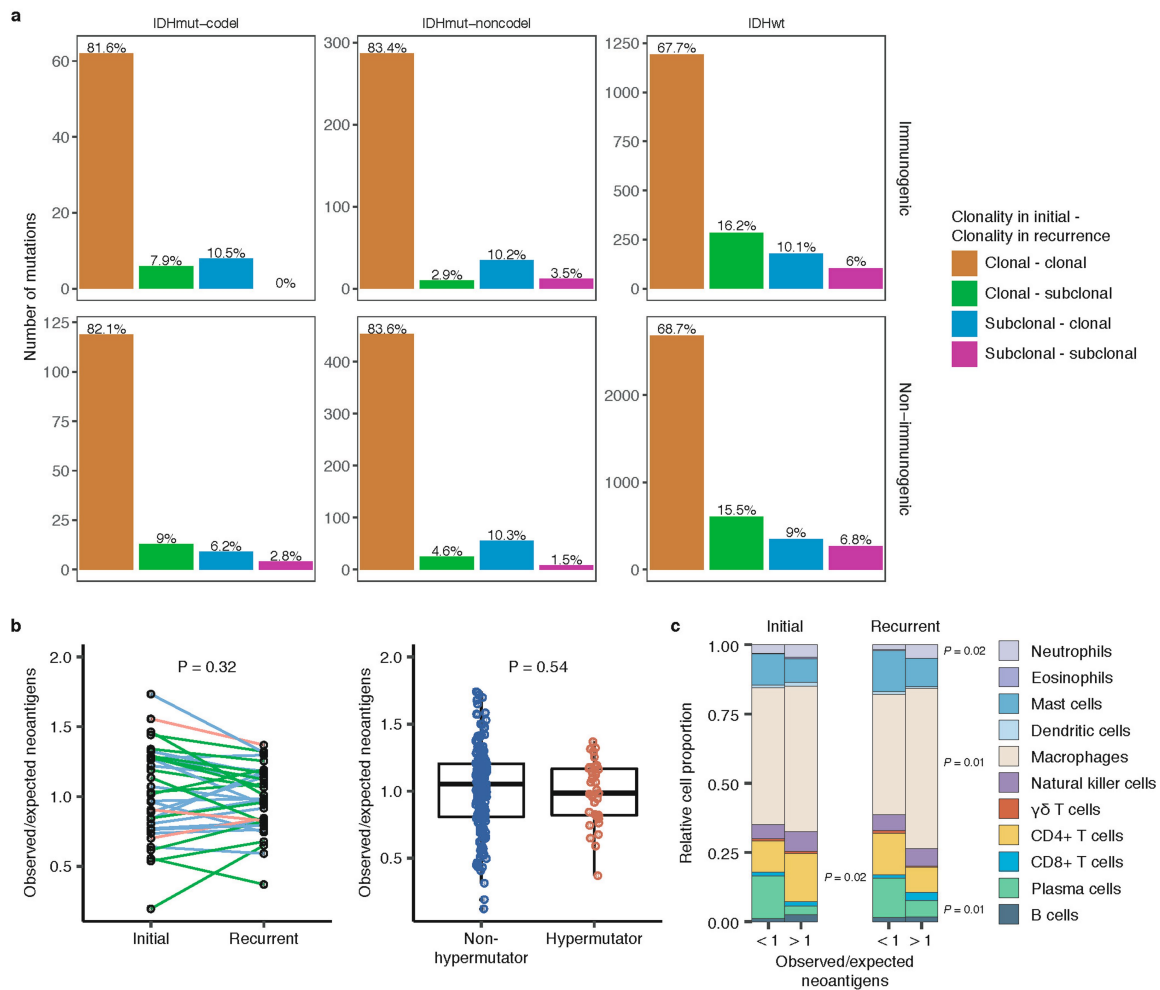


**Extended Data Fig. 11 | Aneuploidy calculation.** **a**, Heat map displaying the chromosomal arm-level events (x axis) with patients represented in each row. Patients are placed in the same order for both the initial (left) and recurrence (right). White space was inserted as a break between the three 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 tumours separated by subtype **d**. As in **c**, comparing aneuploidy value.

# Mutational signatures of genotoxic cancer therapies

## Article



**Extended Data Fig. 12 | 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. The percentage of longitudinal clonality per subtype and mutation is shown. **b**, Left, ladder plot depicting the difference in observed-to-expected neoantigen ratio between the initial and recurrent tumours of patients with hypermutated tumours at recurrence. Each set of points connected by a line represents one tumour ( $n = 70$ ). Right, box plot depicting the distribution of observed-to-expected neoantigen ratios in

recurrent tumours 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 were determined by a paired and an unpaired two-sided *t*-test, for left and right graphs, 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 tumour subgroups. *P* values to the right of each plot indicate a significant difference between the depleted and non-depleted groups for the noted cell type at that time.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- n/a Confirmed
- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
  - A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
  - The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
  - A description of all covariates tested
  - A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
  - A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
  - For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
  - For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
  - For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
  - Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

### Software and code

Policy information about [availability of computer code](#)

Data collection	No software was used for data collection.
Data analysis	<p>MultiQC version: 1.6a0 (quality assessment)            FastQC 0.11.7 (quality assessment)            BWA MEM 0.7.17 (alignment)            R 3.4.2 (general data analyses)            Python 2.7.15 (general data analysis)            Julia 0.7 (general data analysis)            PostgreSQL 10.5 (data management)            BCFTools 1.9 (normalize, sort and index variants)            snakemake 5.2.2 (pipeline development)            GATK (including Mutect2) version: 4.1.0.0 (SNV/CNV detection)            freebayes version: 1.2.0 (variant filtering)            vcf2maf version: 1.6.16 (variant filtering and annotation)            MutationalPatterns version: 1.6.1 (mutational signatures)            TITAN version: 1.19.1 (purity, ploidy, CNV clonality estimates)            dndscv (R package) version: 0.0.1.0 (selection strength, nominate driver genes)            alluvial (R package) version: 0.1-2 (visualize longitudinal neutrality)            DBI (R package) version: 1.0.0 (database management)            tidyverse (R package) version: 1.2.1 (data analysis and visualization)            survival (R package) version: 2.42-6 (survival analyses)            neutralitytestr version: 0.0.2 (subtype-level, variant-level selection)            SubClonalSelection version: 0.0.0 (sample-level selection)            PyClone version: 0.13.1 (mutational clusters)</p>

For manuscripts utilizing custom-made or commercially purchased software, research not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition to a public repository (e.g. GitHub) so that reviewers can confirm that your research data were analyzed using the published pipeline. For more information, see the Nature Research [guidelines for submitting code & software](#) for further information.

OptiType version: 3.1 (HLA class I types)  
 WACed version: 4.0.10 (neoantigen prediction)  
 netMHCpan version: 2.8 (neoantigen prediction)  
 All other custom scripts and pipelines are available on the project's github page (<https://github.com/TheJacksonLaboratory/GLASS>)

# Mutational signatures of genotoxic cancer therapies

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All deidentified, non-protected access somatic variant profiles and clinical data are accessible via Synapse (<http://synapse.org/glass>). A subset of whole genome and whole exome sequencing data has been deposited in the National Center for Biotechnology Information's Sequencing Read Archive and/or the European Genome/Phenome Archive (EGA). Please see Supplementary Table 1 for availability and accession codes.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](http://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. <b>Sample size was a function of availability.</b>
Data exclusions	We defined a quality control process to integrate whole exome and whole genome sequencing data collected from multiple cohorts. As shown in Extended Data Fig. 1, two datasets, Silver and Gold, were constructed to be used for each major analysis type, SNV and CNV, respectively. The two criteria used are intended to provide quality classifications for samples across fingerprinting, coverage, copy number variation (CNV) data and clinical annotation. Fingerprinting was performed using CrosscheckFingerprints (Picard), the purpose of this is to check that all of the input files (readgroups, libraries, samples, files) belong to the same patient, to remove duplicated cases, unmatched samples, and samples of poor quality. Any evidence of mismatch rendered the samples "blocked", otherwise the sample was annotated as "allow". To ensure suitable coverage for mutation calling, samples with near 0 mutation frequency as well as those 2 standard deviations below the mean for either WGS or WXS were annotated as "block". Samples were categorized as "allow", "review", or "block". Copy number data were excluded via manual review of all selected copy number solutions. Manual review consisted of identifying whether data had an atypical or noisy segmentation profile. While we recognize that this strategy is not objective it proved to be an effective strategy for identifying poor performing samples. Insufficient signal, noisy signal, TITAN run fail and unexpected genome stability (little to no copy number changes observed suggesting low purity) were the main reasons for sample exclusion or review. Clinical data was another source of sample filtering. Exclusion of samples was mostly related to sample pairs where surgical interval was very short (1-2 months) and thus did not appear to be a true recurrence. Caution should be used when considering whether a sample represents a true recurrence as no standard set time limits exist. Categories for clinical data include "allow", "interval 1 or less months", "interval 2 or less months", "different location" and "surgical indication" (including "further debulking"). Those interested in using the dataset for further analysis are encouraged to make their own judgments on the criteria they select. The Silver set is filtered to include those pairs with no fingerprinting mismatches and sufficient coverage and is made up of 257 pairs. The Gold set contains 222 pairs, which in addition to the previously mentioned criteria also contain acceptable CNV calls in both samples.
Replication	Replication was limited to select patient samples where both whole genome sequencing and whole exome sequencing was available. <b>All attempts at replication were successful.</b>
Randomization	There was no randomization in this study.
Blinding	All patient samples were deidentified and were assigned a study-specific barcode. Blinding was not relevant to our study since there was no randomization of groups.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

# Mutational signatures of genotoxic cancer therapies

## Materials & experimental systems

- n/a
- Involved in the study
- Antibodies
  - Eukaryotic cell lines
  - Palaeontology
  - Animals and other organisms
  - Human research participants
  - Clinical data

## Methods

- n/a
- Involved in the study
- ChIP-seq
  - Flow cytometry
  - MRI-based neuroimaging

## Human research participants

Policy information about [studies involving human research participants](#)

### Population characteristics

The dataset includes 271 sets of at least two time-separated tumor samples and 17 standalone recurrences. The majority of sets contain two tumor samples (n=246, 85%), with 19 (6.6%) three-tumor sample sets, three (1%) four-tumor sample sets, one (0.3%) with a total of five tumor samples and 17 (5.9%) standalone post-treatment tumor samples. Basic clinical information including age (years), gender, overall survival (months), tumor grade, and tumor histology was available for 90% (260/288) of patients and for 92% (536/584) of tumor samples of the dataset.

Temozolomide and radiation treatment information was available for 68% of the cohort (399/584), data on other treatment modalities was available for 119 patients. Median age at diagnosis of GLASS patients in the IDHmut-noncodel and IDHmut-codel subtypes were both 34 years old and in the IDHwt group age at diagnosis was 53 years old. This is compared with 46 years for IDHmut-codels, 38 years for the IDHmut-noncodels and 59 years in the TCGA cohort respectively. Patients in our dataset were biased toward longer survival as 261 patients were deemed fit for surgical resection or biopsy at recurrence. Median survival for primary glioblastoma patients was 21 months (95% CI 19–23) in the GLASS cohort versus 15 months in historical cohorts.

Patients in this cohort were predominantly treated at teaching/academic centers, which have been shown to be an independent predictive factor of longer survival compared with non-teaching/community hospital settings

All other relevant patient demographics for the GLASS cohort are presented in the Supplement.

### Recruitment

Informed consent was obtained from all study subjects as part of each institution's individual IRB.

### Ethics oversight

All tissue source centers listed in **Supplementary Table 1** obtained study approval by the corresponding institutional review board (IRB) and informed consent from all patients in the cohort. Data pooling at the Jackson Laboratory was performed under the oversight of the IRB at the Jackson Laboratory.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

NA.

Study protocol

NA.

Data collection

NA.

Outcomes

NA.



## Chapter 4

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

This chapter is based on the following publication<sup>63</sup>:

**Kocakavuk, E.**, Anderson, K.J., Varn, F.S. *et al.* Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer. *Nat Genet* **53**, 1088–1096 (2021). <https://doi.org/10.1038/s41588-021-00874-3>



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

Emre Kocakavuk<sup>1,2,3</sup>, Kevin J. Anderson<sup>1</sup>, Frederick S. Varn<sup>1</sup>, Kevin C. Johnson<sup>1</sup>, Samirkumar B. Amin<sup>1</sup>, Erik. P. Sulman<sup>4</sup>, Martijn P. Lolkema<sup>5</sup>, Floris P. Barthel<sup>1,6,7,8</sup> ✉ and Roel G. W. Verhaak<sup>1,8</sup> ✉

**Ionizing radiation causes DNA damage and is a mainstay for cancer treatment, but understanding of its genomic impact is limited. We analyzed mutational spectra following radiotherapy in 190 paired primary and recurrent gliomas from the Glioma Longitudinal Analysis Consortium and 3,693 post-treatment metastatic tumors from the Hartwig Medical Foundation. We identified radiotherapy-associated significant increases in the burden of small deletions (5–15 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 classical non-homologous end-joining-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.**

Radiation therapy or radiotherapy (RT) is used in the clinical management of more than half of all patients with cancer<sup>1,2</sup>. 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) and single-strand annealing (SSA)<sup>3</sup>. 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 (also called microhomology-mediated end joining) uses a length of 2–20 base pairs (bp) of microhomology and SSA uses >20 bp of (micro-)homology. These repair processes may cause changes in the post-treatment cancer genome that can be detected through 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<sup>4</sup> and in particular following temozolomide (TMZ) chemotherapy of gliomas<sup>5,6</sup>. Similarly, an increased burden of small deletions has been observed in radiation-induced malignancies<sup>7</sup>. Despite these advancements, the mutational footprints of palliative and curative radiation in sporadic tumors are not well understood.

To address this gap in knowledge, we analyzed pre- and post-treatment datasets from the Glioma Longitudinal Analysis (GLASS) cohort as well as post-treatment metastatic tumor datasets from the Hartwig Medical Foundation (HMF)<sup>5,8,9</sup>. We identified a significant increase of small (5–15-bp) deletions, large (>20 bp) deletions and inversions in response to ionizing radiation, which we genomically characterized. Finally, we observed that the identified signatures were associated with worse clinical outcomes.

### Results

**RT drives small-deletion burden.** RT and TMZ are the post-surgical standard of care for patients with glioma<sup>10</sup>. We assessed the impact of RT and/or TMZ on the somatic mutation burden, including somatic single-nucleotide variants (sSNVs) and small insertions/deletions (indels, length of 1–20 bp), in matched pre- and post-treatment glioma samples ( $n=190$  patients). Of the patients from which these samples were obtained, 119 (63%) received RT and TMZ, 19 (10%) received RT alone, 13 (7%) received only TMZ, and 16 (8%) received neither RT nor TMZ. 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 (acquired, present only in the recurrent tumor). We then calculated the mutation burden (average mutation frequency per megabase) of post-treatment mutations. A median of 0.68 new small deletions per megabase was acquired in recurrent RT-treated (RT+) glioma, which was significantly higher than the median of 0.19 new small deletions per megabase acquired

<sup>1</sup>The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA. <sup>2</sup>DKFZ Division of Translational Neurooncology at the West German Cancer Center, German Cancer Consortium Partner Site, University Hospital Essen, Essen, Germany. <sup>3</sup>Department of Neurosurgery, University Hospital Essen, Essen, Germany. <sup>4</sup>NYU Langone Health, New York, NY, USA. <sup>5</sup>Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands.

<sup>6</sup>Department of Pathology, Brain Tumor Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. <sup>7</sup>Present address: Cancer and Cell Biology Division, The Translational Genomics Research Institute, Phoenix, AZ, USA. <sup>8</sup>These authors jointly supervised this work: Floris P. Barthel, Roel G. W. Verhaak. ✉e-mail: fbarthel@tgen.org; roel.verhaak@jax.org

in recurrent RT-naive (RT<sup>-</sup>) gliomas (Fig. 1a,  $P=5.1 \times 10^{-3}$ , Mann-Whitney *U* test), and significantly higher than the small-deletion burden detected at diagnosis (Fig. 1b). RT was not associated with a significant increase in sSNV burden (Extended Data Fig. 1a,  $P=4.7 \times 10^{-1}$ , Mann-Whitney *U* test) or small-insertion burden (Extended Data Fig. 1a,  $P=6.7 \times 10^{-1}$ , Mann-Whitney *U* test). The small-deletion increase was particularly pronounced in gliomas marked by the presence of mutations in *IDH1*, a clinically relevant subtype<sup>11</sup> predominantly consisting of grade 2 and 3 gliomas (Extended Data Fig. 1b,  $P=1.4 \times 10^{-2}$ , Mann-Whitney *U* test). The number of RT<sup>-</sup> recurrent cases among *IDH1*-wild-type glioma was too small to test for differences ( $n=2$ , versus  $n=107$  RT-treated cases). To ensure that these changes were not due to TMZ-associated hypermutation (>10 mutations per megabase at recurrence)<sup>2</sup>, we stratified the cohort by hypermutation status. Hypermutation associated with small-deletion increase independent of RT treatment, whereas among non-hypermutators only RT<sup>+</sup> patients showed a significant increase in small deletions (Fig. 1b,  $P=5.0 \times 10^{-11}$ , paired Wilcoxon signed-rank test). 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 (Fig. 1c,  $P=3 \times 10^{-3}$ , *t*-test), directly attributing the observed small-deletion increase to RT treatment. Acquired small deletions were not more clonal/subclonal (Extended Data Fig. 1c, hypermutant:  $P=9.3 \times 10^{-1}$ , non-hypermutant:  $P=8.7 \times 10^{-1}$ , Mann-Whitney *U* test). Comparing the pre-treatment mutation burden and aneuploidy scores between tumors with a high and low post-treatment deletion burden revealed no significant differences, suggesting that these pre-RT tumor genomic characteristics are not predictive of RT-induced small-deletion acquisition.

Importantly, 30% (41/136) of non-hypermutant samples gained >1 deletion per megabase following RT, compared to 7% (2/27) of RT-naive non-hypermutators ( $P=1.6 \times 10^{-2}$ , Fisher's exact test). The association between RT treatment and mutational burden was significant for small deletions but not for insertions and sSNVs. Conversely, TMZ-associated hypermutation was correlated with a significant increase in the burden of all types of mutations (Extended Data Fig. 1d).

Following these observations, we hypothesized that RT may have similar effects in other tumor types. We evaluated whole-genome sequencing-derived mutational profiles from 3,693 metastatic tumors with complete treatment annotation (Extended Data Fig. 1e), available via the HMF<sup>3</sup>. We separated tumors by site of origin and compared the small-deletion burden between RT<sup>+</sup> and RT<sup>-</sup> tumors (Fig. 1d). RT<sup>+</sup> tumors were further stratified depending on whether the treatment intent was curative (RT<sup>cur</sup>,  $n=739$ ) or palliative (RT<sup>pall</sup>,  $n=689$ ), which differ in cumulative radiation dosage<sup>12</sup>. While this analysis was restricted to single-time-point mutational profiles, we observed a significantly higher small-deletion burden in RT<sup>+</sup> patients across multiple tumor types, including bone/soft tissue (RT<sup>cur</sup>: median 0.15 deletions per megabase, RT<sup>-</sup>: median 0.08 deletions per megabase,  $P=6.2 \times 10^{-4}$ , Kruskal-Wallis test),

lung (RT<sup>cur</sup>: median 0.56 deletions per megabase, RT<sup>-</sup>: median 0.43 deletions per megabase,  $P=3.4 \times 10^{-3}$ , Kruskal-Wallis test) and breast (RT<sup>cur</sup>: median 0.18 deletions per megabase, RT<sup>-</sup>: median 0.12 deletions per megabase,  $P=1.2 \times 10^{-4}$ , Kruskal-Wallis test) cancers (Fig. 1d). The observed patterns were present in both non-small-cell and small-cell lung cancer but were restricted to estrogen-receptor-positive breast cancer subtypes (Extended Data Fig. 1f). Tumors receiving palliative RT frequently presented an intermediate state in between the RT<sup>-</sup> and RT<sup>cur</sup> cohorts, suggesting an association between acquired small-deletion burden and RT dose.

DNA repair deficiency associates with increased mutational load<sup>4</sup>. We compared the small-deletion burden between HMF tumors with and without microsatellite instability (MSI) or HR deficiency (HRD)<sup>13</sup>. Notably, HRD<sup>+</sup> and particularly MSI<sup>+</sup> tumors harbored significantly more small deletions compared to HRD<sup>-</sup>/MSI<sup>-</sup> samples (Extended Data Fig. 1g,  $P<2.2 \times 10^{-16}$ , Kruskal-Wallis test). RT treatment was associated with an increase in small-deletion burden in HRD<sup>-</sup>/MSI<sup>-</sup> (Extended Data Fig. 1g,  $P=6.0 \times 10^{-8}$ , Mann-Whitney *U* test) and HRD<sup>+</sup> ( $P=3.5 \times 10^{-2}$ ) tumors, but not in MSI<sup>+</sup> tumors ( $P=7.1 \times 10^{-1}$ ). To account for the possibility that HRD and MSI confounded the association between RT treatment and the small-deletion burden, we have included HRD and MSI status in a multivariable log-linear regression analysis and found that the association between RT treatment and small-deletion burden is independent of HRD/MSI status (Extended Data Fig. 1h).

Next, we assessed whether the small-deletion burden was associated with mutations in selected genes (*ATM*, *ATR*, *CHEK1*, *CHEK2*, *PARP1*, *PRKDC*, *TP53* and *WEE1*) involved in the DNA damage response (DDR). This analysis indicated that DDR mutations in these genes 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, DDR gene mutations and various cytotoxic treatment regimens (for example, taxane, platinum, anthracyclines and alkylating agents) that have previously been associated with increased mutation burdens<sup>14</sup>. Results from this analysis confirmed a robust association with an increased small-deletion burden for both palliative and curative RT (Extended Data Fig. 1h, RT<sup>cur</sup> versus RT-naive: odds ratio = 1.25,  $P<1 \times 10^{-3}$ , *t*-test).

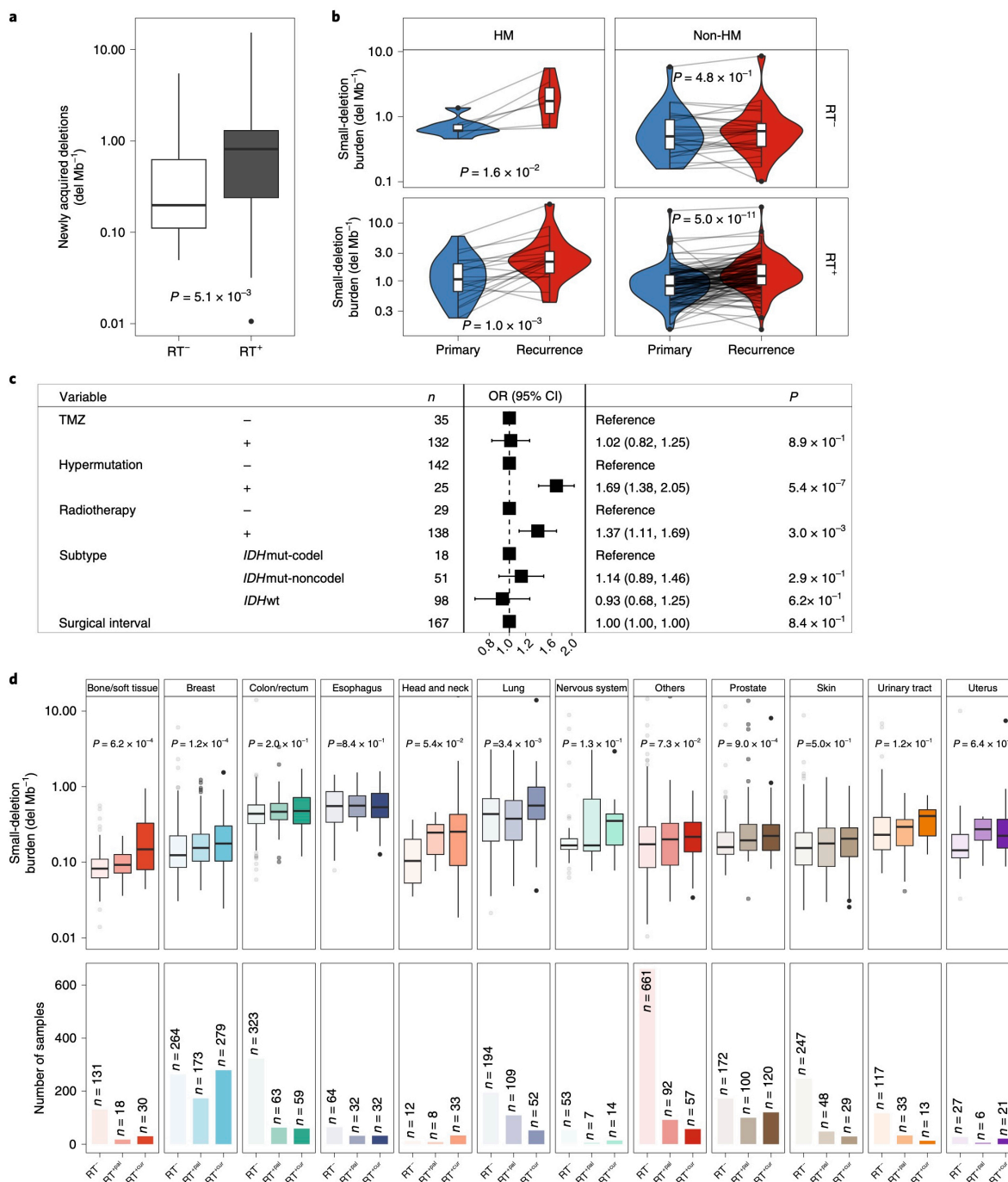
To verify the causal association between RT and acquired small deletions, we reanalyzed whole-genome sequencing data from 324 human induced pluripotent stem cells (iPSCs) exposed to various carcinogens, including two iPSCs treated with ionizing radiation<sup>15</sup>. Small-deletion burden was significantly higher in the RT<sup>+</sup> iPSCs compared to controls (Extended Data Fig. 1i,  $P=2.0 \times 10^{-2}$ , Mann-Whitney *U* test). In contrast, there was no significant difference in small-insertion burden ( $P=1.8 \times 10^{-1}$ ). Strikingly, the ionizing radiation group showed the highest median burden of small deletions across all treatment modalities, further substantiating our results (Extended Data Fig. 1k).

**RT-associated deletions harbor a specific genomic signature.** Characteristics of RT-associated small deletions, such as length

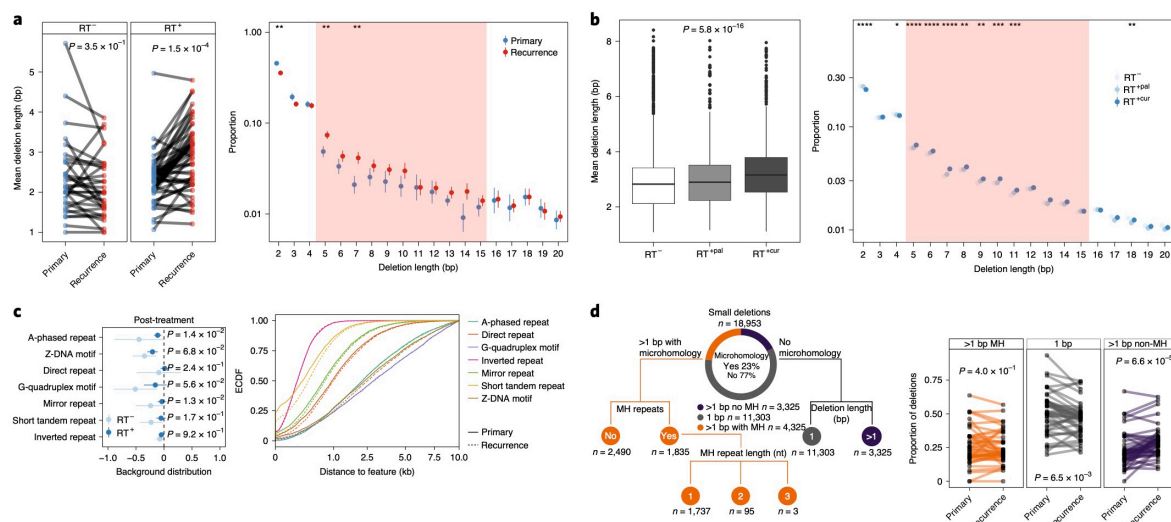
**Fig. 1 | RT is associated with an increased small-deletion burden.** **a**, A boxplot (in this and all following figures: boxes span quartiles, center lines represent medians, and whiskers represent absolute range, excluding outliers) depicting the burden of newly acquired/post-treatment small deletions (deletions per megabase (del Mb<sup>-1</sup>)) in RT<sup>-</sup> ( $n=34$ ) and RT<sup>+</sup> ( $n=156$ ) patients from the GLASS cohort. A two-sided 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 RT. A two-sided paired Wilcoxon signed-rank test was applied for statistical testing. **c**, A forest plot showing a multivariable log-linear regression model of newly acquired small-deletion burden (deletions per megabase) including TMZ treatment, hypermutation, RT treatment, molecular subtype and surgical interval (in months) as variables. A two-sided *t*-test was applied. OR, odds ratio. **d**, Top: for the metastatic cohort, boxplots depicting small-deletion burden (deletions per megabase) in RT<sup>-</sup> tumor samples (left), tumor samples treated with RT with palliative intent (RT<sup>pall</sup>, middle) and tumor samples treated with RT with curative intent (RT<sup>cur</sup>, right), separated by primary tumor location. A two-sided Kruskal-Wallis test was applied for statistical testing. Bottom: sample sizes of the metastatic cohort separated by primary tumor location.

distribution and breakpoint microhomology, may provide insights into their etiology. We explored such features in GLASS *IDH*-mutant gliomas (RT<sup>+</sup>, *n* = 49; RT<sup>-</sup>, *n* = 32) as only two GLASS *IDH*-wild-type gliomas were not RT treated (versus 107 RT treated). Small deletions in RT<sup>+</sup> recurrent tumor samples showed increased deletion lengths (Fig. 2a, left; RT<sup>+</sup>, *P* = 1.5 × 10<sup>-4</sup>; RT<sup>-</sup>,

*P* = 3.5 × 10<sup>-1</sup>; paired Wilcoxon signed-rank test), which was even more pronounced in acquired deletions (Extended Data Fig. 2a, *P* = 1.3 × 10<sup>-4</sup>, Mann–Whitney *U* test), supporting the idea that RT results in longer deletions (Fig. 2a, left). Moreover, the size distribution of RT-associated deletions shifted towards deletions of length ~5–15 bp (Fig. 2a, right).







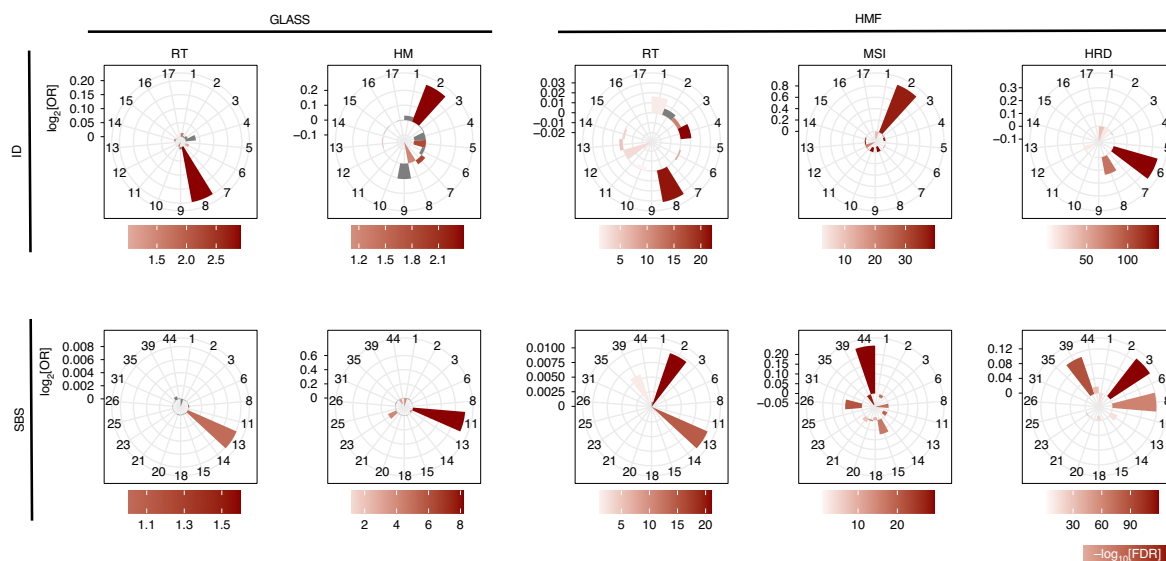
**Fig. 2 | RT-associated small deletions harbor a characteristic genomic signature. a**, Length distribution in GLASS. Left: mean deletion lengths in primary versus recurrent *IDH*-mutant glioma ( $n = 81$ ), separated by RT treatment ( $RT^+$ ,  $n = 32$ ,  $RT^-$ ,  $n = 49$ ). The  $P$  values were calculated using a two-sided paired Wilcoxon signed-rank test. Right: y axis, proportion of deletions; x axis, deletion length  $>1$  bp. Proportions were calculated for each patient, with the mean (point) and 95% CI (line range) compared longitudinally in  $RT^+$  non-hypermutable glioma ( $n = 44$ ).  $**P < 0.01$ ; two-sided paired Wilcoxon signed-rank test. Shaded area (5–15 bp): size range for which the most apparent differences were observed. **b**, Length distribution in HMF. Left: comparison of mean deletion lengths in  $RT^-$  versus  $RT^{+pal}$  versus  $RT^{+cur}$  samples. The  $P$  values were calculated using a two-sided Kruskal–Wallis test. Right: y axis, proportion of deletions; x axis, deletion length  $>1$  bp. Proportions were calculated for each patient, with the mean (point) and 95% CI (line range) compared between  $RT^-$  versus  $RT^{+pal}$  versus  $RT^{+cur}$  samples.  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ ,  $****P < 0.0001$ ; two-sided Kruskal–Wallis test. **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:  $\log_{10}$  ratio of mean distance of non-radiation-associated and radiation-associated post-treatment deletions to genomic feature over background distribution in non-hypermutable glioma samples ( $n = 69$ ). Distribution of radiation-associated deletions shows little variability (narrow 95% CI) and resembles background distribution more closely (closer to 0). The  $P$  values were calculated using a two-sided Mann–Whitney  $U$  test. Right: empirical cumulative distribution function (ECDF, y axis) of distance to non-B-DNA features in kilobases (x axis) in post-irradiated non-hypermutable recurrent samples ( $n = 44$ ). Neither in hypermutable, nor in  $RT^-$  non-hypermutable gliomas were longitudinal differences observed (Extended Data Fig. 2c). **d**, Small-deletion categories in GLASS. Left: separation of small deletions in the GLASS cohort into 3 major categories: 1 bp (gray),  $>1$  bp without microhomology (MH; purple) and  $>1$  bp with microhomology (orange) in *IDH*-mutant glioma ( $n = 81$ ). 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-hypermutable glioma sample ( $n = 44$ , further comparisons in Extended Data Fig. 2e) using a two-sided paired Wilcoxon signed-rank test.

Comparing  $RT^+$  and  $RT^-$  metastatic tumor samples from the single-time-point HMF dataset confirmed larger deletions in palliative and curative RT-treated tumors (Fig. 2b and Extended Data Fig. 2b), and a shift in deletion span from 1–4 bp towards 5–15 bp deletions (Fig. 2b). Deletion length was larger following curative compared to palliative RT treatment, further substantiating a dose and exposure association. Taken together, these findings suggest that RT drives a burden of small deletions with distinct characteristics.

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<sup>16</sup>, which may include RT-induced deletions. We compared the distribution of small deletions across these genomic features against a random background distribution. In the *IDH*-mutant GLASS cohort, deletions following RT showed less variability and higher similarity to the random background distribution compared to non-RT-induced deletions, and larger distances to non-B-DNA features (Fig. 2c, right, and Extended Data Fig. 2c). The lack of or reduced association between RT-associated deletions and the analyzed genomic features, such as repeats and G-quadruplex motifs, suggests that RT-associated small deletions occur in a stochastic manner, independent of the intrinsic mutagenicity of the fragile genome regions 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), to identify selection of mutations at the level of individual genes and separately for GLASS pre- and post-treatment fractions (Extended Data Fig. 2d)<sup>17</sup>. We did not find evidence for significant selection for any genes in the post-treatment fraction following RT. In the HMF set, where pre-treatment samples are unavailable, we could not perform this analysis. Our results in *IDH*-mutant glioma further support the notion that RT-associated deletions do not occur at particular genomic loci.

Small deletions can be the result of error-prone DSB repair mechanisms such as c-NHEJ and a-EJ<sup>3</sup>. We characterized deletions on the basis of size, microhomology and repeat content to investigate which mechanism is used for DSB repair following RT (Fig. 2d and Extended Data Fig. 2e). Deletions without microhomology comprised the majority of deletions in the dataset (77%, Fig. 2d). However, in non-hypermutable gliomas receiving ionizing radiation, we observed a significant increase in  $>1$ -bp deletions without microhomology (Fig. 2d,  $P = 6.6 \times 10^{-5}$ , paired Wilcoxon signed-rank test) and conversely a decrease in 1-bp deletions (Fig. 2d,  $P = 6.5 \times 10^{-3}$ , paired Wilcoxon signed-rank test). Comparison of  $RT^+$  and  $RT^-$  metastatic tumors from the HMF dataset demonstrated comparable



**Fig. 3 | ID8 and APOBEC SBS signatures associated with RT.** Indel (ID) and SBS mutational signatures in the GLASS and HMF cohorts associated with RT, hypermutation, MSI and HRD. RT<sup>+</sup> mean contribution = 0.22 versus RT<sup>-</sup> mean contribution = 0.07,  $P = 7.4 \times 10^{-5}$ ,  $Q = 3.8 \times 10^{-3}$ , two-sided Mann-Whitney  $U$  test and FDR, respectively. Bars in the petal plots not reaching statistical significance (defined as  $FDR < 0.01$ ) are indicated in gray.

results (Extended Data Fig. 2f). These data suggest that c-NHEJ is the preferred pathway for repairing radiation-induced DNA damage.

**Distinct mutational signatures associated with RT.** Cancer cells accumulate somatic mutations through mechanisms that may leave distinct genomic scars, termed mutational signatures<sup>18</sup>. To determine the mutational processes of RT, we compared pre- and post-treatment mutations in the GLASS dataset to previously defined mutational signatures<sup>18</sup>. The comparison of signature contributions between post-recurrence mutations in RT<sup>+</sup> and RT<sup>-</sup> *IDH*-mutant glioma samples revealed a strong enrichment of indel signature 8 (ID8; Fig. 3 and Extended Data Fig. 3d; RT<sup>+</sup> mean contribution = 0.22 versus RT<sup>-</sup> mean contribution = 0.07,  $P = 7.4 \times 10^{-5}$ ,  $Q = 3.8 \times 10^{-3}$ , Mann-Whitney  $U$  test and false discovery rate (FDR), respectively). In RT<sup>+</sup> but not RT<sup>-</sup> patients, comparing ID8 values before and after treatment revealed significant increases in absolute (Extended Data Fig. 3e,  $P = 4.5 \times 10^{-7}$ , paired Wilcoxon rank-signed test) and relative (Extended Data Fig. 3e,  $P = 2.3 \times 10^{-3}$ ) ID8 contributions, post-treatment. ID8 is composed of  $\geq 5$ -bp deletions without microhomology and has previously been linked to DSB repair by c-NHEJ, suggesting radiation-induced DSB repair via c-NHEJ<sup>18</sup>. As expected, hypermutation due to TMZ treatment in *IDH*-mutant gliomas was associated with ID2, which is reported to be elevated in mismatch-repair-deficient tumors (Fig. 3 and Extended Data Fig. 3a,b)<sup>18</sup>.

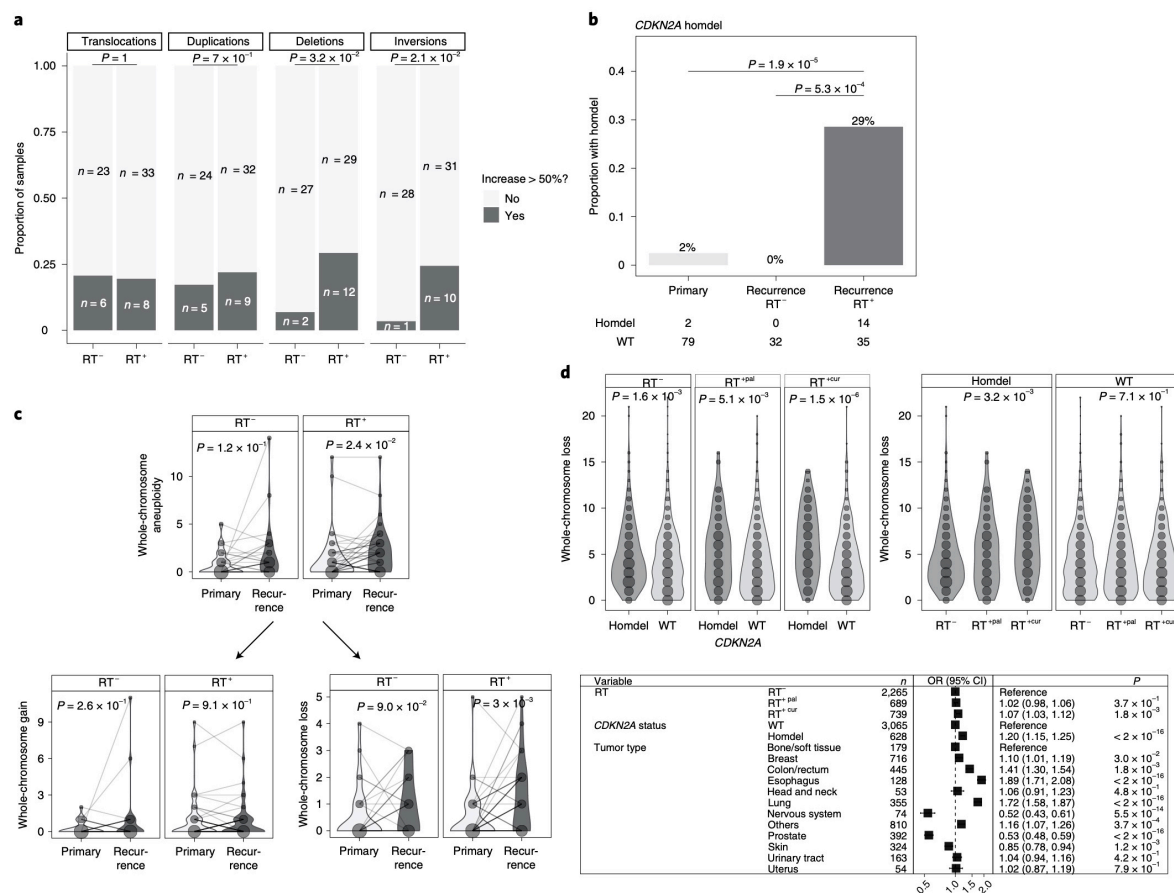
A previous mutational signature analysis in the HMF dataset observed the strongest association between RT and ID6<sup>14</sup>. Confirming our findings in the GLASS cohort, we observed the strongest association with ID8 in the HMF dataset, and significant but less pronounced association for ID6 (Fig. 3). Both absolute and relative ID8 values were significantly higher in RT<sup>+</sup> samples when compared to RT<sup>-</sup> samples, and a significant association was observed in nine of twelve tumor types (Extended Data Fig. 3f). The comparison of HRD<sup>+</sup> and HRD<sup>-</sup> samples associated HRD with ID6. ID6 comprises  $>5$ -bp deletions with microhomology at breakpoints and

is elevated in HRD<sup>+</sup> breast cancers<sup>19</sup>. Analogous to TMZ-associated hypermutators in GLASS, MSI samples in HMF were enriched for ID2 (Fig. 3).

To summarize, while MSI leads to an increased small-deletion burden due to hypermutability from impaired DNA mismatch repair at microsatellites/homopolymers, DSBs due to HRD and RT are repaired via error-prone DSB repair mechanisms. Our results implicate the a-EJ pathway that utilizes microhomologies at breakpoints in HRD<sup>+</sup> samples (signature ID6) and the c-NHEJ pathway which does not require breakpoint microhomology in RT<sup>+</sup> samples (signature ID8).

We sought to identify single-base-substitution (SBS) signature associations in both datasets. We confirmed an enrichment of SBS11 in hypermutant *IDH*-mutant glioma samples<sup>5,20</sup>, an enrichment of signatures SBS44, SBS26, SBS21, SBS20 and SBS15 in MSI<sup>+</sup> samples<sup>18</sup> and enrichment of SBS3 and SBS8 in HRD<sup>+</sup> cases<sup>18,19,21</sup> along with a previously undescribed enrichment of SBS39 (Fig. 3). In addition, in the GLASS cohort, RT treatment was significantly associated with SBS13, and in the HMF cohort, it was significantly associated with SBS2 and SBS13. SBS2 and SBS13 are APOBEC signatures<sup>18,22</sup>. APOBEC cytosine deaminases are involved in retrovirus and retrotransposon restriction, and the enrichment of APOBEC signatures in RT<sup>+</sup> samples in both datasets implicated APOBEC-mediated mutagenesis in RT-associated DSB repair<sup>23-25</sup>. Our results support the hypothesis that mutational signatures are shaped by cycles of DNA damage and DNA repair<sup>26</sup>. While RT causes DSBs that are repaired via c-NHEJ resulting in specific small deletions (ID8), APOBEC cytosine deaminases may be activated during the repair process leading to SBS2 and/or SBS13.

**RT associates with aneuploidy and larger deletions.** We evaluated whether RT could be associated with other types of genomic variants. We detected large structural variants, including large deletions, duplications, inversions and translocations, in the longitudinal GLASS cohort. We observed an increase in large deletions (length  $> 20$  bp to chromosome-arm length) post-therapy in RT<sup>+</sup>



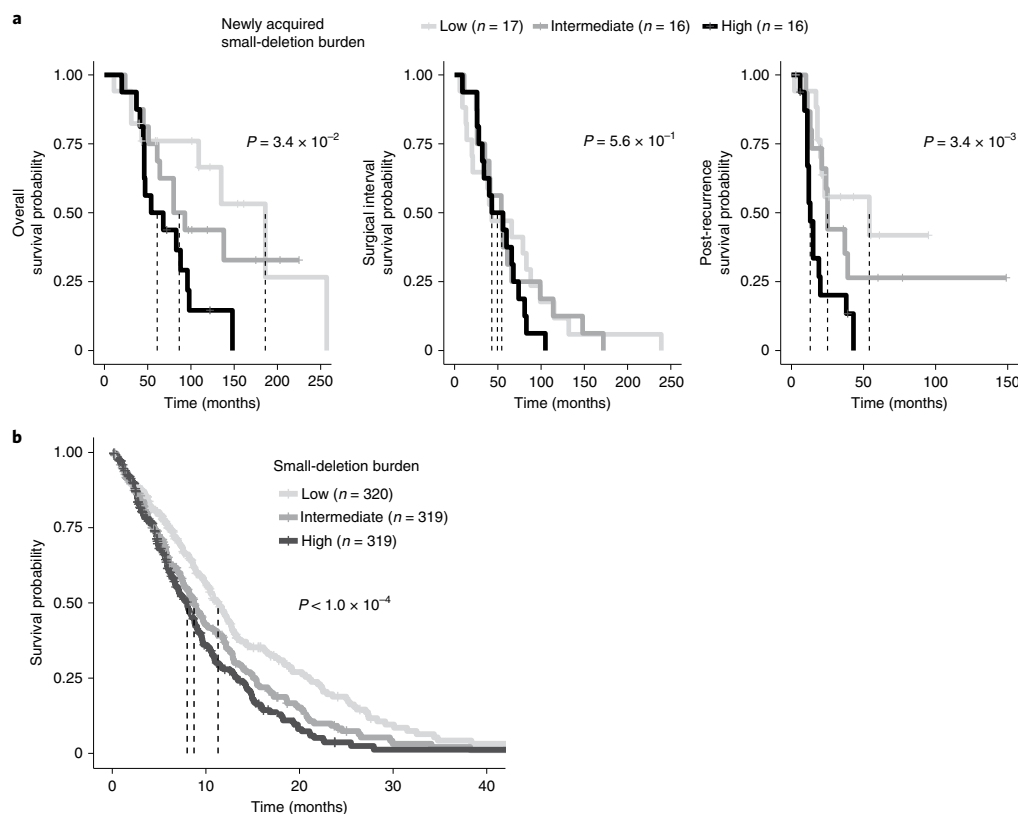
**Fig. 4 | RT is associated with aneuploidy and larger deletions.** **a**, RT-associated increase in large deletions and inversions. The plot shows results of an analysis of structural variants after RT in *IDH*-mutant glioma samples with sufficient quality for calling ( $n = 70$ ): translocations, duplications, deletions and inversions. For each patient, the number of structural variants was calculated pre- and post-treatment. Based on the distribution of percentage increase from primary to recurrence, a cutoff was set for  $>50\%$  increase (Extended Data Fig. 4a). Shown is a comparison of the proportion of samples with/without an increase of given structural variants between RT<sup>+</sup> versus RT<sup>-</sup>. The *P* values were calculated using a two-sided Fisher's exact test. **b**, RT-associated *CDKN2A* homozygous deletions (homdel). Depicted are proportions of *IDH*-mutant glioma samples ( $n = 81$ ) harboring a homozygous deletion in *CDKN2A*. Using a two-sided Fisher's exact test, proportions were compared between RT-received recurrence (RT<sup>+</sup>) versus RT-naive recurrence (RT<sup>-</sup>) versus samples before treatment (primary). Detailed distributions of whole-chromosome deletion scores are provided in Extended Data Fig. 4f. wt, wild type. **c**, RT-associated whole-chromosome aneuploidy. Upper: longitudinal comparison of whole-chromosome aneuploidy scores separated by RT treatment for *IDH*-mutant glioma samples with sufficient quality for calling and complete treatment annotation (total  $n = 69$ , RT<sup>+</sup>  $n = 42$ , RT<sup>-</sup>  $n = 27$ ). Bottom: separation of whole-chromosome aneuploidy into whole-chromosome gain (left) and whole-chromosome loss (right) scores, respectively. The size of the dots is proportional to the frequency of whole-chromosome loss integers for each subgroup. The *P* values were calculated using a two-sided paired Wilcoxon rank-signed test. **d**, Validation of structural variant and aneuploidy results in HMF. Upper: comparison of whole-chromosome deletion scores between RT<sup>-</sup> versus RT<sup>+pal</sup> versus RT<sup>+cur</sup> and/or *CDKN2A* homozygous deletion versus wild-type samples. Note that *CDKN2A* homozygous deletion is associated with higher whole-chromosome deletion scores, independent of RT. Within samples with *CDKN2A* homozygous deletion, samples that were RT treated with curative intent show the highest deletion scores. The size of the dots is proportional to the frequency of whole-chromosome loss integers for each subgroup. The *P* values were calculated using a two-sided Kruskal-Wallis test. Detailed distributions of whole-chromosome deletion scores are provided in Extended Data Fig. 4g. Bottom: a multivariable Poisson regression model for whole-chromosome deletion scores integrating RT, *CDKN2A* and tumor type as variables. A two-sided Wald test was applied.

patients, compared to RT<sup>-</sup> patients (Fig. 4a,  $P = 3.2 \times 10^{-2}$ , Fisher's exact test). We also found a statistically significant increase in inversions (Fig. 4a,  $P = 2.1 \times 10^{-2}$ ) and no differences in translocations (Fig. 4a,  $P = 1$ ) and duplications (Fig. 4a,  $P = 7 \times 10^{-1}$ ). These associations remained significant after accounting for potentially confounding factors such as TMZ treatment and molecular subtype

(Extended Data Fig. 4b). While radiation-induced secondary malignancies were reported to contain increased rates of inversions<sup>7</sup>, a concomitant increase in large deletions in association with RT has not previously been observed.

We next evaluated whether deletions at specific loci were associated with RT. In the *IDH*-mutant GLASS cohort, where *CDKN2A*





**Fig. 5 | RT-associated genomic changes are linked to poor survival.** **a**, Association of RT-related deletions with survival in GLASS. Left: Kaplan–Meier survival plots comparing overall survival dependent on deletion burden at recurrence using a log-rank test in RT<sup>+</sup> *IDH*-mutant glioma samples ( $n = 49$  with available survival information). Samples were separated into 3 tertiles based on deletion burden at recurrence: high (top tertile), intermediate (middle tertile) and low (bottom tertile). Dotted lines indicate median overall survival times. Note the stepwise association of tertiles with survival. Middle: Kaplan–Meier survival plots comparing surgical interval/time to second surgery dependent on deletion burden at recurrence using a two-sided log-rank test. Right: Kaplan–Meier survival plots comparing post-recurrence survival dependent on deletion burden at recurrence using a two-sided log-rank test. **b**, Association of RT-related deletions with survival in HMF. Shown are Kaplan–Meier survival plots comparing survival time dependent on deletion burden at metastasis using a two-sided log-rank test in RT<sup>+</sup> metastases ( $n = 958$  with available survival information). Samples were separated into 3 tertiles based on deletion burden: high (top tertile), intermediate (middle tertile) and low (bottom tertile). Dotted lines indicate median survival times. Note the stepwise association of tertiles with survival.

loss at initial diagnosis is rare, acquired *CDKN2A* homozygous deletions occurred exclusively in RT<sup>+</sup> recurrences (Fig. 4b, 29% versus 0%,  $P = 5.3 \times 10^{-4}$ , Fisher's exact test)<sup>27</sup>. This result nominated acquired *CDKN2A* homozygous loss as a potential biomarker for RT resistance among recurrent *IDH*-mutant gliomas, but not in *IDH*-wild-type gliomas where *CDKN2A* homozygous deletion at diagnosis is common.

Ionizing radiation can promote mitotic chromosome segregation errors through non-disjunction events causing aneuploidy<sup>28–31</sup>. We investigated the association between RT and aneuploidy, 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, Extended Data Fig. 4c). In the *IDH*-mutant GLASS cohort, we observed a significant association between RT and chromosome losses, but not for simple gains or complex events (Fig. 4c and Extended Data Fig. 4d), which was no longer significant after adjusting for covariates in a multivariable Poisson regression. Instead, this analysis highlighted a significant association between chromosome losses and *CDKN2A* deletions

(Extended Data Fig. 4e), implying 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 (Fig. 4d and Extended Data Fig. 4f). In fact, both curative RT treatment and *CDKN2A* homozygous deletions were independently associated with increased number of chromosomal losses in the HMF datasets (Fig. 4d and Extended Data Fig. 4f). However, testing for interactions between *CDKN2A* deletions and RT treatment indicated a trend towards interaction between palliative/curative RT and *CDKN2A* deletions (Supplementary Table 1,  $P = 9.75 \times 10^{-2}$  and  $P = 4.92 \times 10^{-2}$ , respectively, *t*-test). In summary, aneuploidy may not be directly associated with RT but through interactions with *CDKN2A* deletions, requiring further investigation.

**RT-associated genomic changes are linked to poor survival.** Finally, we wanted to ascertain whether the genomic effects of RT were relevant to patient outcomes. As expected, *CDKN2A*



homozygous deletion at recurrence was significantly associated with worse overall survival in *IDH*-mutant glioma samples (Extended Data Fig. 5a,  $P < 1 \times 10^{-4}$ , log-rank test)<sup>5</sup>. To test for a survival association of *CDKN2A* deletions among RT<sup>+</sup> patients in the HMF dataset, we selected 958 samples that received RT and had sufficient survival information available from 11 tumor types (Extended Data Fig. 1e). Patients whose tumors harbored a *CDKN2A* homozygous deletion showed worse outcomes compared to patients with *CDKN2A*-wild-type tumors (Extended Data Fig. 5b, left). Stratification of the cohort into tertiles based on genome-wide aneuploidy frequency demonstrated that low aneuploidy was linked to favorable outcomes and high aneuploidy was linked to poor outcomes (Extended Data Fig. 5b, middle). 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 acquired *CDKN2A* deletions, GLASS patients with tumors carrying a high small-deletion burden at recurrence (top tertile) had significantly shorter overall survival (Fig. 5a,  $P = 3.4 \times 10^{-2}$ , 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 (Extended Data Fig. 5c, HR = 1.19 (95% confidence interval (CI): 1.01–1.14);  $P = 4.3 \times 10^{-2}$ , Wald test). Multivariable modeling using a limited subset of patients with detailed dosage information in the GLASS cohort ( $n = 21$ ) showed that the association between small-deletion burden and survival is dose independent ( $P = 2 \times 10^{-2}$ ). 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 (Fig. 5a,  $P = 3.4 \times 10^{-3}$ , log-rank test). Surgical interval times did not differ significantly between the three tertiles (Fig. 5a,  $P = 5.6 \times 10^{-1}$ ), suggesting that patients with glioma may initially benefit equally from RT, but after exposure to RT and acquisition of the deletion signature, tumors may lose sensitivity to further RT. This pattern is reminiscent of the association between hypermutant glioma and TMZ therapy<sup>30</sup>.

In 958 RT<sup>+</sup> samples from the HMF cohort (Extended Data Fig. 1g), we also found that patients harboring a high small-deletion burden (top tertile) had significantly shorter survival than other RT-treated patients (Fig. 5b,  $P < 4 \times 10^{-4}$ , log-rank test). Similarly, stratifying HMF patients into tertiles by ID8 burden associated an intermediate or high ID8 burden with poor survival and a low ID8 burden with more favorable outcomes (Extended Data Fig. 5b). Therefore, the presence of a high number of RT-associated small deletions identifies a tumor that has initially responded to therapy, but which may have lost some or all of the treatment sensitivity. Combined, 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

Prior studies on radiation-induced tumors have shown a range of genomic effects and have suggested the involvement of various DNA DSB repair mechanisms<sup>30,32–35</sup>. We identified a unique signature of RT-associated deletions carrying characteristics of DSB repair by c-NHEJ. This work extends our knowledge on the genomic response to RT and provides direction for the development of effective radiosensitizers.

The notable expansion of clones harboring RT-associated genomic events depends on clonal selection or drift<sup>36</sup>. 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. We found that a higher load of RT-associated deletions was linked to worse patient outcomes. Thus, additional rounds of RT in patients with recurrent or metastatic tumors containing a significant increase in small-deletion burden are unlikely to further extend progression-free survival. The ability to effectively repair RT-induced damage implies that sensitivity to RT is intrinsically diminished or has progressively been lost due to, for example, clonal selection. Inhibiting these repair processes could potentially sensitize tumors to the tumor-killing effect of ionizing radiation. *CDKN2A* homozygous deletions were acquired in RT<sup>+</sup> *IDH*-mutant gliomas but not in RT<sup>-</sup> recurrent *IDH*-mutant gliomas, suggesting that RT-induced DNA damage promotes the acquisition of this poor prognostic marker, and implicating a convergence between RT-induced DSB repair and cell cycle checkpoints. A biomarker able to readily detect an increased small-deletion burden may help reduce treatment costs and avoid RT-associated patient comorbidities and side effects.

We note several limitations to our study. The HMF metastatic dataset comprises samples from a single time point, preventing 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 was also irradiated. Considering these caveats, the effects of RT described in this study might be more pronounced than what we have observed. Additionally, evolutionary pressures for local and distant metastases, which were analyzed homogeneously in this study, might be fundamentally different<sup>37</sup>. RT may have no or a different impact on metastases that are not immediately within the field of the radiation, requiring further investigations in dedicated datasets.

Compounds that inhibit DNA repair may improve the response of cancer cells to RT. Numerous clinical and preclinical studies have shown efficacy in targeting DNA repair. Inhibitors directed at poly(ADP-ribose) polymerase (*PARP*) in HRD<sup>+</sup> tumors (synthetic lethality) were shown to be effective in the treatment of various cancer types<sup>38–40</sup>. Effective inhibitors of NHEJ have not yet been reported but may sensitize tumors to RT. Inhibitors of ATM serine/threonine kinase (*ATM*), a protein kinase that activates 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 preclinical studies and phase 1 trials of diffuse glioma<sup>41–44</sup>. The identification of enrichment of APOBEC-associated mutational signatures may warrant further evaluation of targeting cytosine deaminases for cancer therapy<sup>45</sup>. The present study highlights the importance of effective DNA repair in therapy resistance.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-021-00874-3>.

Received: 18 September 2020; Accepted: 21 April 2021;

Published online: 27 May 2021

## References

- Barton, M. B. et al. Estimating the demand for radiotherapy from the evidence: a review of changes from 2003 to 2012. *Radiother. Oncol.* **112**, 140–144 (2014).
- 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.* **79**, 1507–1515 (2011).
- Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**, 495–506 (2017).

4. Campbell, B. B. et al. Comprehensive analysis of hypermutation in human cancer. *Cell* **171**, 1042–1056 (2017).
5. Barthel, F. P. et al. Longitudinal molecular trajectories of diffuse glioma in adults. *Nature* **576**, 112–120 (2019).
6. Touat, M. et al. Mechanisms and therapeutic implications of hypermutation in gliomas. *Nature* **580**, 517–523 (2020).
7. Behjati, S. et al. Mutational signatures of ionizing radiation in second malignancies. *Nat. Commun.* **7**, 12605 (2016).
8. Priestley, P. et al. Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **575**, 210–216 (2019).
9. Consortium, G. Glioma through the looking GLASS: molecular evolution of diffuse gliomas and the Glioma Longitudinal Analysis Consortium. *Neuro Oncol.* **20**, 873–884 (2018).
10. Stupp, R. et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **352**, 987–996 (2005).
11. Louis, D. N. et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* **131**, 803–820 (2016).
12. Lutz, S. T., Jones, J. & Chow, E. Role of radiation therapy in palliative care of the patient with cancer. *J. Clin. Oncol.* **32**, 2913–2919 (2014).
13. Nguyen, L., Martens, J. W. M., Van Hoeck, A. & Cuppen, E. Pan-cancer landscape of homologous recombination deficiency. *Nat. Commun.* **11**, 5584 (2020).
14. Pich, O. et al. The mutational footprints of cancer therapies. *Nat. Genet.* **51**, 1732–1740 (2019).
15. Kucab, J. E. et al. A compendium of mutational signatures of environmental agents. *Cell* **177**, 821–836 (2019).
16. Georgakopoulos-Soares, I., Morganella, S., Jain, N., Hemberg, M. & Nik-Zainal, S. Noncanonical secondary structures arising from non-B DNA motifs are determinants of mutagenesis. *Genome Res.* **28**, 1264–1271 (2018).
17. Martincorena, I. et al. Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029–1041 (2017).
18. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
19. Davies, H. et al. HRDetect is a predictor of *BRCA1* and *BRCA2* deficiency based on mutational signatures. *Nat. Med.* **23**, 517–525 (2017).
20. Touat, M. et al. Mechanisms and therapeutic implications of hypermutation in gliomas. *Nature* **580**, 517–523 (2020).
21. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
22. Roberts, S. A. et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **45**, 970–976 (2013).
23. Lei, L. et al. APOBEC3 induces mutations during repair of CRISPR–Cas9-generated DNA breaks. *Nat. Struct. Mol. Biol.* **25**, 45–52 (2018).
24. Nowarski, R. & Kotler, M. APOBEC3 cytidine deaminases in double-strand DNA break repair and cancer promotion. *Cancer Res.* **73**, 3494–3498 (2013).
25. Nowarski, R. et al. APOBEC3G enhances lymphoma cell radioresistance by promoting cytidine deaminase-dependent DNA repair. *Blood* **120**, 366–375 (2012).
26. Volkova, N. V. et al. Mutational signatures are jointly shaped by DNA damage and repair. *Nat. Commun.* **11**, 2169 (2020).
27. Ceccarelli, M. et al. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* **164**, 550–563 (2016).
28. Adewoye, A. B., Lindsay, S. J., Dubrova, Y. E. & Hurles, M. E. The genome-wide effects of ionizing radiation on mutation induction in the mammalian germline. *Nat. Commun.* **6**, 6684 (2015).
29. Bakhom, S. F. et al. Numerical chromosomal instability mediates susceptibility to radiation treatment. *Nat. Commun.* **6**, 5990 (2015).
30. Rose, Li, Y. et al. Mutational signatures in tumours induced by high and low energy radiation in *Trp53* deficient mice. *Nat. Commun.* **11**, 394 (2020).
31. Toul, N., Elhajoui, A., Thierens, H. & Kirsch-Volders, M. 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* **15**, 1–7 (2000).
32. Behjati, S. et al. Mutational signatures of ionizing radiation in second malignancies. *Nat. Commun.* **7**, 12605 (2016).
33. Davidson, P. R., Sherborne, A. L., Taylor, B., Nakamura, A. O. & Nakamura, J. L. A pooled mutational analysis identifies ionizing radiation-associated mutational signatures conserved between mouse and human malignancies. *Sci. Rep.* **7**, 7645 (2017).
34. Lopez, G. Y. et al. The genetic landscape of gliomas arising after therapeutic radiation. *Acta Neuropathol.* **137**, 139–150 (2019).
35. Phi, J. H. et al. Genomic analysis reveals secondary glioblastoma after radiotherapy in a subset of recurrent medulloblastomas. *Acta Neuropathol.* **135**, 939–953 (2018).
36. Hu, Z. et al. Quantitative evidence for early metastatic seeding in colorectal cancer. *Nat. Genet.* **51**, 1113–1122 (2019).
37. Reiter, J. G. et al. Lymph node metastases develop through a wider evolutionary bottleneck than distant metastases. *Nat. Genet.* **52**, 692–700 (2020).
38. Wiggans, A. J., Cass, G. K., Bryant, A., Lawrie, T. A., & Morrison, J. Poly(ADP-ribose) polymerase (PARP) inhibitors for the treatment of ovarian cancer. *Cochrane Database Syst. Rev.* **5**, CD007929 (2015).
39. 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.* **16**, 1661–1668 (2014).
40. O’Neil, N. J., Bailey, M. L. & Hieter, P. Synthetic lethality and cancer. *Nat. Rev. Genet.* **18**, 613–623 (2017).
41. 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.* **11**, 10463–10476 (2019).
42. 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-2*H*-pyran-4-yl)-7,9-dihydro-8*H*-purin-8-one (AZD7648), a potent and selective DNA-dependent protein kinase (DNA-PK) inhibitor. *J. Med. Chem.* **63**, 3461–3471 (2020).
43. Thijssen, R. et al. Dual TOR/DNA-PK inhibition blocks critical signaling pathways in chronic lymphocytic leukemia. *Blood* **128**, 574–583 (2016).
44. Timme, C. R., Rath, B. H., O’Neill, J. W., Camphausen, K. & Tofilon, P. J. The DNA-PK inhibitor VX-984 enhances the radiosensitivity of glioblastoma cells grown in vitro and as orthotopic xenografts. *Mol. Cancer Ther.* **17**, 1207–1216 (2018).
45. Li, M. et al. First-in-class small molecule inhibitors of the single-strand DNA cytosine deaminase APOBEC3G. *ACS Chem. Biol.* **7**, 506–517 (2012).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

## 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<sup>5</sup>. 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*)<sup>11</sup>. The GLASS cohort used in this manuscript consists of  $n = 106$  whole-genome sequencing samples ( $n = 53$  primary samples,  $n = 53$  matched first-recurrence samples) and  $n = 274$  whole-exome sequencing samples ( $n = 106$  primary samples,  $n = 106$  matched first-recurrence samples). Detailed information on sequence platforms, capture kits and read length is outlined in the GLASS marker paper<sup>5</sup>.

For validation analyses, we curated a metastatic cohort from the HMF comprising a total of 4,549 samples<sup>8</sup>. 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 HMF. Whole-genome sequencing was performed for each sample according to standardized protocols<sup>16</sup>. Detailed information on sequence platforms, capture kits and read length is outlined in the HMF marker paper<sup>8</sup>. VCF files with mutations and associated metadata were downloaded from The Hartwig Medical Database (<https://database.hartwigmedicalfoundation.nl>). After application of filtering criteria using BCFTools 1.9 (as described in detail in Extended Data Fig. 1e), a set of  $n = 3,693$  were defined and used for the majority of analyses throughout the manuscript. For survival analyses, we selected RT<sup>+</sup> samples with sufficient survival information ( $n = 958$ ). All prior RT data were extracted using clinical data as present in the CPCT-02 eCRF on 8 December 2020. These data were not cleaned and represent the data entered by the clinical sites. The prior RT 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 RT was considered curative intent RT. All local RT for pain relief or other symptom-directed goals was considered as palliative. Some items were not specified, and those events were not included in our analysis. We also excluded all RT for non-malignant disease states, specifically for gynecocoma treatment after castration. We cannot exclude overrepresentation or underrepresentation of the radiation signatures as we are unaware whether the metastases that were biopsied were not already present at the time of RT.

**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<sup>5</sup>. 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 variant allele frequency 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.

**Mutation burden comparison.** The mutation burden was calculated as the number of mutations per megabase with at least 10× coverage and stratified by variant type. The overall tumor mutation burden was calculated as the sum of the burden of small deletions, small insertions and SNVs. Recurrent tumors with greater than 10 mutations per megabase were considered hypermutated as previously described<sup>5</sup>. 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 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 did not necessitate 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<sup>17</sup>. 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 base pairs) or categorical (position in or up to 100 bp to genomic feature—yes/no) value. We used a two-sided 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 RT, we calculated dN/dS ratios as previously described<sup>3</sup>. Briefly, the R package dNdScv<sup>17</sup> 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 RT-naïve and RT-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 2 nucleotides of homology was considered microhomology mediated. The homologous sequence was characterized and further analyzed for the presence of 1-, 2- and 3-nucleotide repeats. The repeat unit and number of repeats were quantified.

**Mutational signatures.** SigProfiler was used to extract and plot mutational signatures of SBSs, double-base substitutions and indels<sup>18</sup>. Absolute and relative contributions of signatures were determined using modified functions from the MutationalPatterns R package<sup>19</sup>. 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 SBS signatures SBS31 and SBS35 have previously been linked to platinum therapy<sup>14,18</sup>. Analysis of the HMF cohort using the signatures we extracted confirmed these previously established associations, providing further credence to the identified signatures. SigProfilerPlotting<sup>19</sup> was used to visualize the distribution of indel characteristics (Extended Data Fig. 3a–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<sup>20</sup>. 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<sup>21</sup>. Copy number variation predictions inferred from read depth using CNVator 0.3.3 were additionally provided to garner further support for identified variants<sup>22</sup>. The resulting call set was post-processed using SVtyper 0.6.0 to genotype structural variants for each individual sample belonging to a patient<sup>23</sup>. Finally, we used GATK VariantFiltration to filter all variants with fewer than four reads of support and those with quality scores less than  $10^4$ . 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 20 bp were considered.

**MSI and HRD.** MSI and HRD status were assigned according to previously defined criteria<sup>15</sup>. In short, MSI was determined in samples with >14,000 indel repeats; HRD was classified on the basis of a probability score of  $\geq 0.5$  according to the CHORD algorithm.

**Aneuploidy calculation.** Arm-level aneuploidy data from the GLASS dataset were obtained from a previous publication and copy number segmentation files from HMF were processed into arm-level copy number calls using the same methods<sup>5</sup>. 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 (broadly using tidyverse 1.3.0), Python 3.7.3 and PostgreSQL 10.5. R was interfaced with the PostgreSQL database used for data storage using the unixODBC 2.3.6 driver plus the DBI 1.0.0 and odbc 1.1.6 R packages. All survival analyses including Kaplan–Meier plots and Cox proportional hazards models were conducted using the R packages survival and survminer. For unpaired group comparisons, the two-sided Mann–Whitney *U* test and two-sided Kruskal–Wallis test were used and for paired longitudinal comparisons the two-sided 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<sup>5</sup>. 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.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Processed sequencing data from the GLASS project used in this study are available on Synapse, at <https://www.synapse.org/glass>. The whole-genome sequencing, RNA sequencing and corresponding clinical data used in this study were made available by the HMF and were accessed under a license agreement (HMF DR-057 version 3.0). Data access can be obtained by filling out a data request form. The form and detailed application procedures can be found at <https://www.hartwigmedicalfoundation.nl/applying-for-data/>. The repeatmasker database used in this manuscript is available at <https://www.repeatmasker.org/>.

## Code availability

Pipeline scripts can be found at <https://github.com/fpbarthel/GLASS>. Custom scripts for analyses performed in this manuscript can be found at <https://github.com/EmreKocakavuk/RTscars>.

## References

46. 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* **22**, 33–40 (2017).
47. 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 Res.* **41**, D94–D100 (2013).
48. Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med.* **10**, 33 (2018).
49. Bergstrom, E. N. et al. SigProfilerMatrixGenerator: a tool for visualizing and exploring patterns of small mutational events. *BMC Genomics* **20**, 685 (2019).
50. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
51. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
52. Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **21**, 974–984 (2011).
53. Chiang, C. et al. SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat. Methods* **12**, 966–968 (2015).
54. 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* **43**, 11.10.1–11.10.33 (2013).

## Acknowledgements

This publication and the underlying study have been made possible partly on the basis of the data that HMF and the Center of Personalised Cancer Treatment (CPCT) have made available to the study. This work was supported by the NIH grants R01 CA190121, R01 CA237208, R21 NS114873 and Cancer Center Support Grant P30 CA034196, grants from the Musella Foundation, the B\*CURVED Foundation and 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 the recipient of an MD fellowship by the Boehringer Ingelheim Fonds and is supported by the German National Academic Foundation.

## Author contributions

E.K., F.P.B. and R.G.W.V. designed the project. Data processing and analysis was performed by E.K. and F.P.B.; data visualization was performed by E.K., E.K., K.J.A., F.S.V., K.C.J., S.B.A., F.P.B. and R.G.W.V. participated in the design of analyses and interpretation of results. M.P.L. provided clinical data. E.K., F.P.B. and R.G.W.V. wrote the manuscript. All coauthors including M.P.L. and E.P.S. discussed the results and commented on the manuscript.

## 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 the Scientific Advisory Board of the HMF. 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. The remaining authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41588-021-00874-3>.

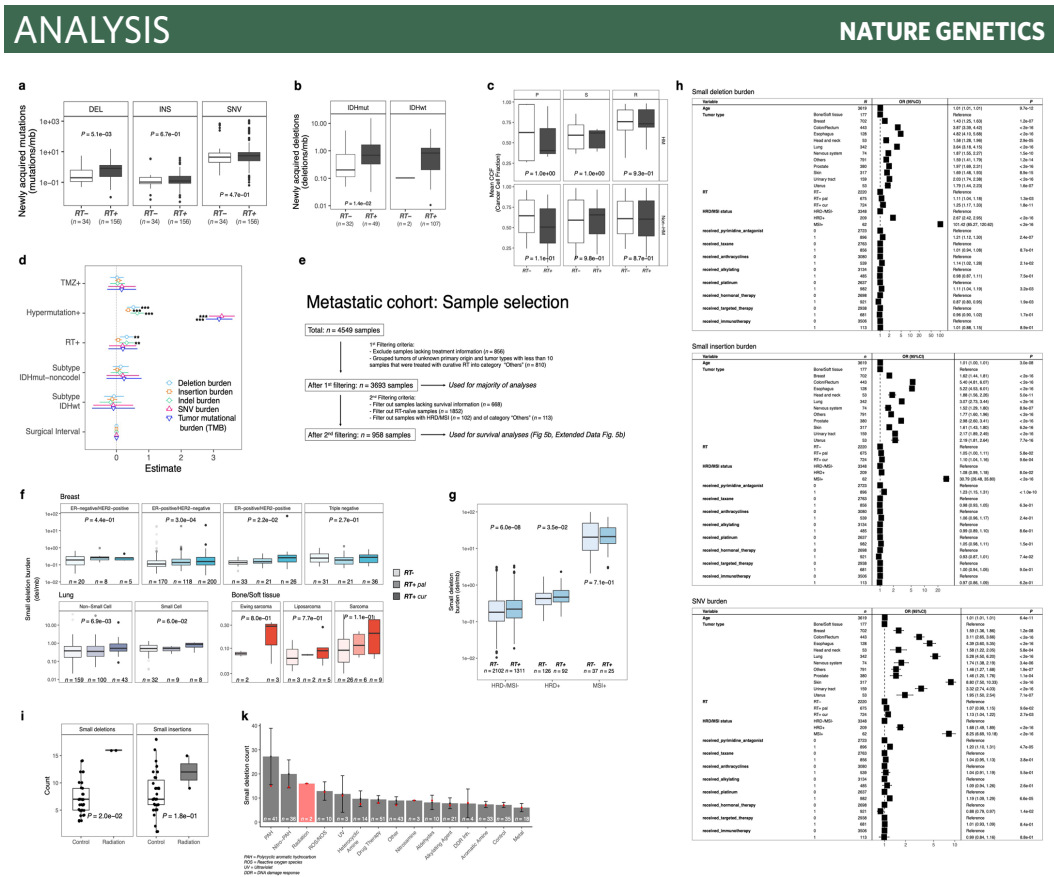
**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41588-021-00874-3>.

**Correspondence and requests for materials** should be addressed to F.P.B. or R.G.W.V.

**Peer review information** *Nature Genetics* thanks Moritz Gerstung, Simon Powell and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

# Mutational signatures of genotoxic cancer therapies



**Extended Data Fig. 1 | Radiotherapy specifically drives small deletion burden independent of multiple variables.** **a**, Boxplot (in this and all following figures: boxes span quartiles, center lines as medians, whiskers represent absolute range, excluding outliers): burden of post-treatment mutations (mutations/mb) in RT-naïve (n=34) and RT-received (n=156) patients from GLASS cohort. Mutations separated by DEL (deletions), INS (insertions) and SNV (single nucleotide variants). Two-sided Mann-Whitney U test. **b**, Acquired small deletion burden comparison between RT-naïve and RT-received cases separated by molecular subtype. Two-sided Mann-Whitney U test. **c**, Comparison of mean cancer cell fraction of small deletions per patient in GLASS separated by P, primary-only fraction, S, shared fraction and R, recurrence-only fraction and by HM, hypermutation. Two-sided Mann-Whitney U test. **d**, Forest plots: multivariable log-linear regression model of acquired mutation burden (mutations/mb) in GLASS. Point, mean estimate; lines, 95%-confidence-interval. Two-sided t-test (\*\*=p < 0.01, \*\*\*=p < 0.001). **e**, Sample selection and filtering criteria for HMF including a detailed description of the usage for specific figures. **f**, Separation of lung, breast and bone/soft tissue cancers into respective subtypes. Comparison of small deletion burden between RT-, RT + pal and RT + cur samples. Two-sided Kruskal-Wallis test. **g**, Boxplots depicting burden of small deletions in HRD-/MSI- (n=3,413), HRD+ (n=218) and MSI+ (n=62) samples from the HMF cohort separated by RT-status. Two-sided Mann-Whitney U test. **h**, Forest plots depicting multivariable log-linear regression model for mutation burdens in HMF. Two-sided t-test. Mutations separated into small deletions/insertions and SNVs. Independent variables: age, primary tumor location, DNA repair deficiency background and treatment including radiotherapy, taxane, alkylating agents, platinum and others. **i**, Comparison of small deletion counts between control vs ionizing radiation groups (PMID:30982602). Two-sided Mann-Whitney U test. **k**, Distribution of small deletion counts per treatment group (PMID:30982602). Data presented as mean values +/- standard error of the mean, and red dots indicate median count of small deletions.

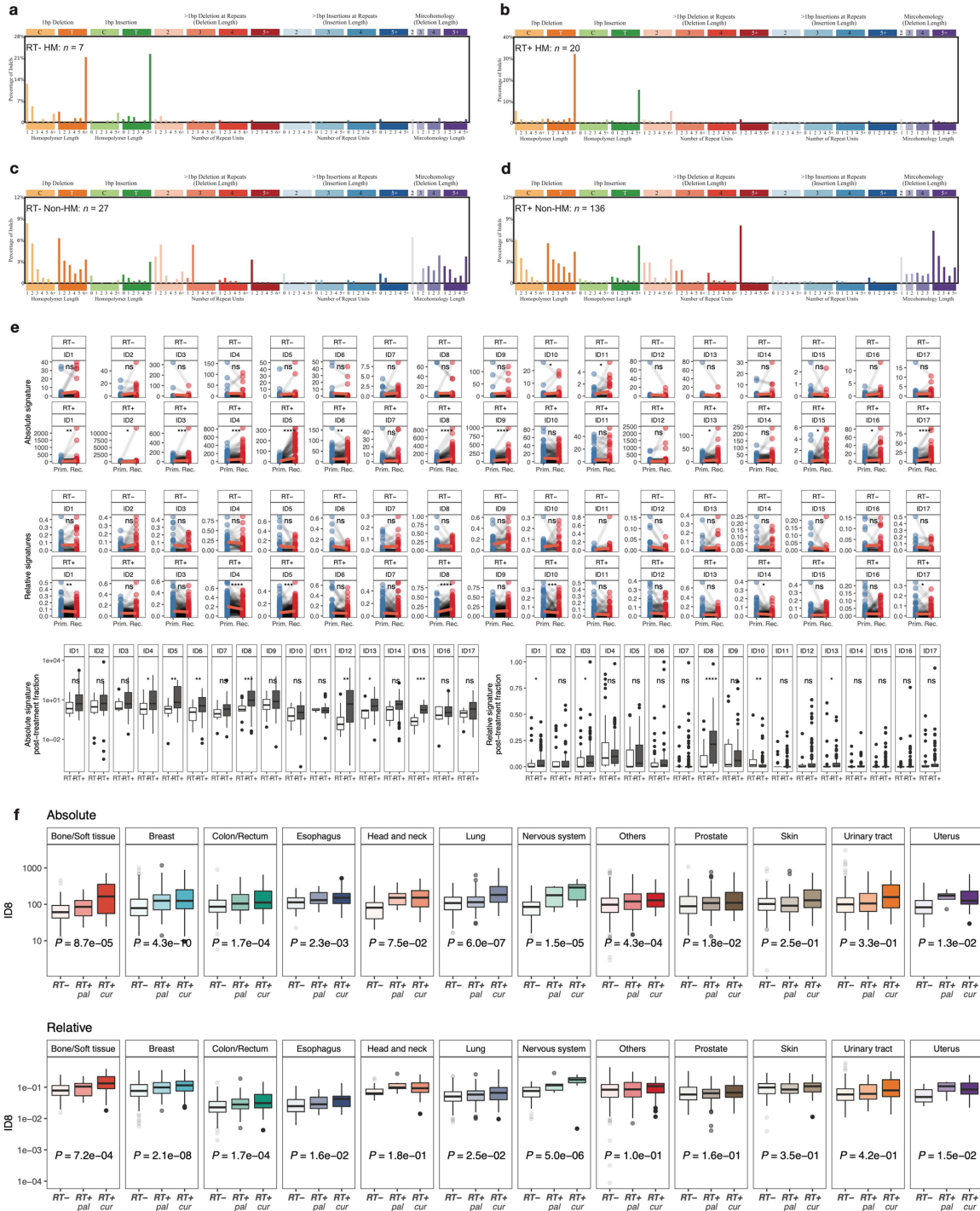




# Mutational signatures of genotoxic cancer therapies

ANALYSIS

NATURE GENETICS



Extended Data Fig. 3 | See next page for caption.

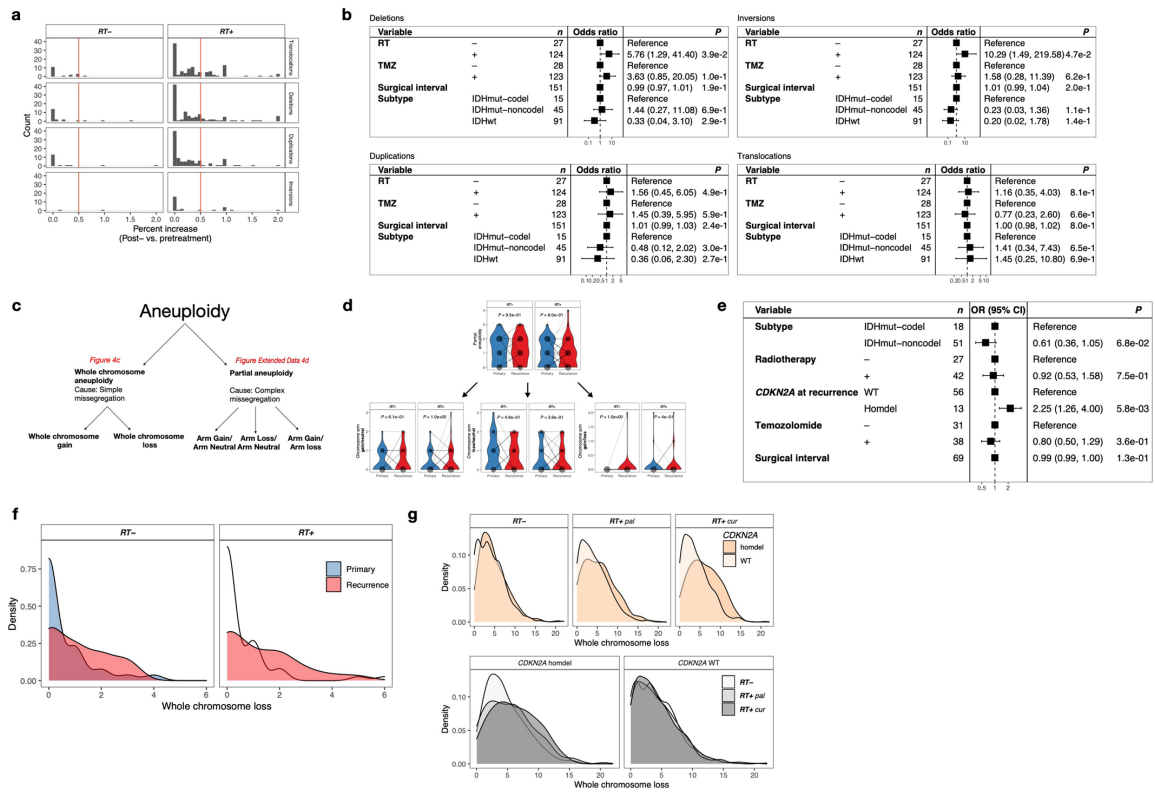
**Extended Data Fig. 3 | Mutational signatures associated with RT.** **a-d**, Distribution of indel types for post-treatment mutations in the GLASS cohort, separated by RT (**a, c**, RT- negative; **b, d**, RT-treated) and HM (**a-b**, Hypermutator; **c-d**, Non-Hypermutator). Note that patterns of indels 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 *IDH*mut 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 two-sided 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 two-sided 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. Boxes span quartiles, center lines as medians, whiskers represent absolute range, excluding outliers. **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. Two-sided Kruskal-Wallis test was applied for statistical testing. Boxes span quartiles, center lines as medians, whiskers represent absolute range, excluding outliers.



# Mutational signatures of genotoxic cancer therapies

## ANALYSIS

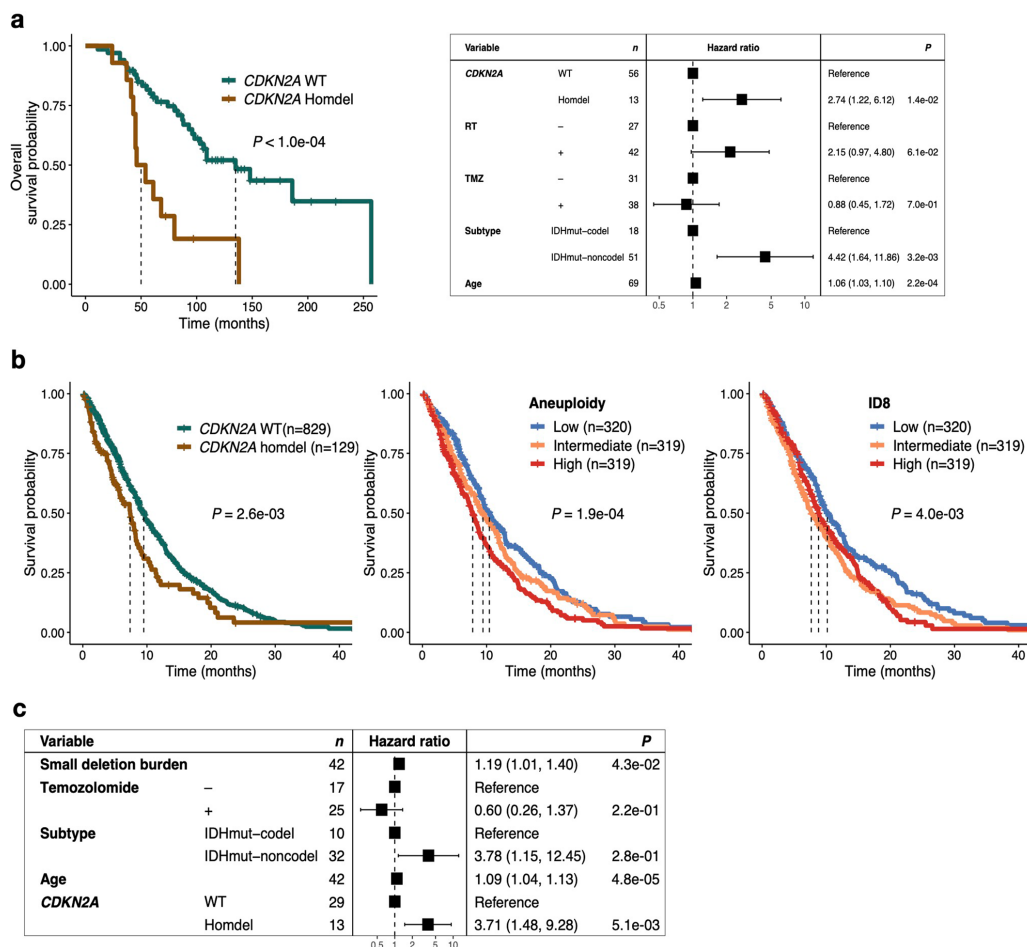
## NATURE GENETICS



**Extended Data Fig. 4 | Effects of radiotherapy on structural variants.** **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 Fig. 4a. **b**, To support analyses presented in Fig. 4a, a multivariable logistic regression model was fitted for the >50% increase values of the structural variant types. Two-sided Wald test. This model includes radiation therapy, temozolomide therapy, molecular subtype and surgical interval as variables. **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. Dots are proportional to the frequency of whole chromosome loss integer for each subgroup. Two-sided paired Wilcoxon rank-signed test. **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. Two-sided Wald test. Note that 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-naive 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.

# Mutational signatures of genotoxic cancer therapies

## NATURE GENETICS ANALYSIS



**Extended Data Fig. 5 | Radiotherapy-associated genomic scars linked to poor survival.** **a**, Left: Kaplan-Meier survival plot comparing overall survival time dependent on *CDKN2A* status at recurrence using two-sided 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. Two-sided Wald test was applied. **b**, Left: Kaplan Meier survival plot comparing survival time dependent on *CDKN2A* status at metastasis using two-sided 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 two-sided log-rank test in RT-treated metastases (n = 958 with available survival information). Samples were separated into 3 tertiles based on whole chromosome loss aneuploidy scores: high (top tertile), intermediate (middle tertile) and low (bottom tertile). Right: Kaplan Meier survival plot comparing survival time dependent RT signature ID8 burden at metastasis using two-sided log-rank test in RT-treated metastases (n = 958 with available survival information). Samples were separated into 3 tertiles based on ID8 burden: high (top tertile), intermediate (middle tertile) and low (bottom tertile). 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 *IDH* mutant samples.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Confirmed  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection	PostgreSQL 10.5 (data collection and storage, data analysis) DBI (R package) version: 1.0.0 (database management) odbc (R package) version 1.1.6 (database connectivity) unixODBC 2.3.6 (database connectivity)
-----------------	--

# Mutational signatures of genotoxic cancer therapies

Data analysis	BCFTools 1.9 (normalize, sort and index variants) snakemake 5.2.2 (pipeline development) GATK (including Mutect2) version: 4.1.0.0 (SNV/CNV detection) samtools 1.7 (sam/bam processing) SigProfilerMatrixGenerator version: 1.1 (mutational signatures) SigProfilerPlotting version: 1.1 (mutational signatures) MutationalPatterns version: 1.6.1 (mutational signatures) lumpy version: 0.2.13 (calling structural variants) CNVnator version: 0.3.3 (copy number variations) SVtyper version: 0.6.0 (post-processing structural variant calls) PyClone version: 0.13.1 (cancer cell fraction) dndscv (R package) version: 0.0.1.0 (selection strength, nominate driver genes) tidyverse (R package) version: 1.3.0 (data analysis and visualization) survival (R package) version: 3.1-12 (survival analyses) survminer (R package) version 0.4.6 (survival analysis) forestmodel (R package) version: 0.5.0 (visualization forest plots) All custom scripts relevant to this paper can be found on our github page at <a href="https://github.com/EmreKocakavuk/RTscars">https://github.com/EmreKocakavuk/RTscars</a> . Data pre-processing pipelines can be found at <a href="https://github.com/fpbarthel/GLASS">https://github.com/fpbarthel/GLASS</a> .
---------------	--

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Processed sequencing data from the GLASS project used in this study are available on Synapse, at <https://www.synapse.org/glass>. The WGS, RNA-seq, and corresponding clinical data used in this study were made available by the Hartwig Medical Foundation and were accessed under a license agreement (HMF DR-057 version 3.0). Data access can be obtained by filling out a data request forms. The form and detailed application procedures can be found online at <https://www.hartwigmedicalfoundation.nl/applying-for-data/>. The RepeatMasker database used in this manuscript is available online at <https://www.repeatmasker.org/>.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. This is a retrospective analysis and all available data was initially included to maximize sample size and power. A detailed sample selection is described in the "Data exclusions" section below.
Data exclusions	GLASS: From the high-quality longitudinal DNA sequencing data of n=222, we selected matched samples from primary and matched first recurrence glioma samples from n=190 patients (n=380 samples) with fully annotated clinical information on radiotherapy treatment and survival data.  HMF: We downloaded WGS data from 4549 metastatic samples from the Hartwig Medical Foundation. We applied 2 major filtering steps. First, we excluded samples lacking treatment information (n=856), and grouped tumors of unknown origin of primary tumor and tumor types with less than n=10 samples that were treated with curative RT into the category "Others" (n=810). After application of filtering criteria (as described in detail in Supplementary Fig. 1e) a set of 3693 samples were defined and used for the majority of analyses throughout the manuscript. In a second filtering step, we excluded samples lacking treatment information (n=668) and filtered out samples that were radiotherapy-naïve (n=1852), harbored homologous recombination deficiency (HRD) or microsatellite instability (MSI) backgrounds (n=102) or were categorized as "Others" as described above (n=113). We used the curated n=958 samples with radiotherapy treatment and sufficient survival information for survival analyses.
Replication	Findings were found to be consistent in two independent datasets. Furthermore, data and code are available as outlined above and execution of this code on the matching dataset will accurately reproduce the noted findings.
Randomization	There was no randomization in this work. This work is a retrospective analysis and does not concern a controlled trial where patients or samples are assigned to an intervention group.
Blinding	All patient samples were deidentified and were assigned a study-specific barcode. This concerns a retrospective study and blinding was not relevant to our work as patients were not randomly assigned to an intervention group.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

- | n/a                                 | Included in the study                                  |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern  |

### Methods

- | n/a                                 | Included in the study                           |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

## Chapter 5

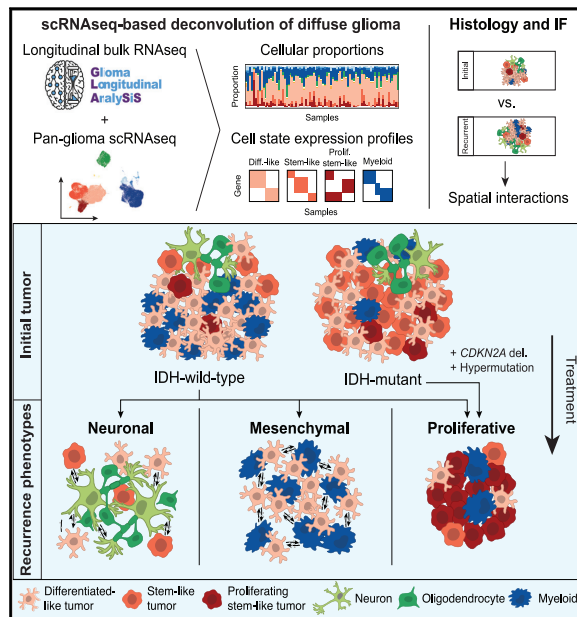
# Glioma progression is shaped by genetic evolution and microenvironment interactions

This chapter is based on the following publication<sup>64</sup>:

Varn, F.S., Johnson, K.C., Martinek J., Huse J.T., Nasrallah, M.P., Wesseling, P., Cooper, L.A.D., Malta, T.M., Wade, T.E., Sabedot, T.S., Brat, D., Gould, P.V., Wöhrer, A., Aldape, K., Ismail, A., Sivajothi, S.K., Barthel, F.P., Kim, H., **Kocakavuk, E.**, *et al.* Glioma progression is shaped by genetic evolution and microenvironment interactions. *Cell* **185**, 2184-2199 (2022).  
<https://doi.org/10.1016/j.cell.2022.04.038>

# Glioma progression is shaped by genetic evolution and microenvironment interactions

Graphical abstract



Authors

Frederick S. Varn, Kevin C. Johnson, Jan Martinek, ..., Antonio Iavarone, Roel G.W. Verhaak, The GLASS Consortium

Correspondence

roel.verhaak@jax.org

In brief

Integrating longitudinal transcriptomic and genomic data from paired diffuse glioma samples with complementary single-cell RNA-seq and multiplex immunofluorescence datasets reveals recurrence-associated genetic and microenvironmental changes that are dependent on IDH mutation status.

Highlights

- Longitudinal glioma evolution follows an IDH mutation-dependent trajectory
- Hypermutation and *CDKN2A* deletions underlie increased proliferation at recurrence
- Recurrent IDH-wild-type neoplastic cells up-regulate neuronal signaling programs
- Mesenchymal transitions associate with distinct myeloid cell interactions



Varn et al., 2022, Cell 185, 2184–2199  
 June 9, 2022 © 2022 Elsevier Inc.  
<https://doi.org/10.1016/j.cell.2022.04.038>





## Resource

# Glioma progression is shaped by genetic evolution and microenvironment interactions

Frederick S. Varn,<sup>1</sup> Kevin C. Johnson,<sup>1</sup> Jan Martinek,<sup>1</sup> Jason T. Huse,<sup>2,3</sup> MacLean P. Nasrallah,<sup>4</sup> Pieter Wesseling,<sup>5,6</sup> Lee A.D. Cooper,<sup>7</sup> Tathiane M. Malta,<sup>8</sup> Taylor E. Wade,<sup>1</sup> Thais S. Sabedot,<sup>9</sup> Daniel Brat,<sup>7</sup> Peter V. Gould,<sup>10</sup> Adelheid Wöhrer,<sup>11</sup> Kenneth Aldape,<sup>12</sup> Azzam Ismail,<sup>13</sup> Santhosh K. Sivajothi,<sup>1</sup> Floris P. Barthel,<sup>1,14</sup> Hoon Kim,<sup>1,15</sup> Emre Kocakavuk,<sup>1,16</sup> Nazia Ahmed,<sup>17</sup> Kieron White,<sup>18</sup> Indrani Datta,<sup>19</sup> Hyo-Eun Moon,<sup>20</sup> Steven Pollock,<sup>17</sup> Christine Goldfarb,<sup>1</sup> Ga-Hyun Lee,<sup>1</sup> Luciano Garofano,<sup>21</sup> Kevin J. Anderson,<sup>1</sup> Djamel Nehar-Belaid,<sup>1</sup> Jill S. Barnholtz-Sloan,<sup>22,23</sup> Spyridon Bakas,<sup>4,24</sup> Annette T. Byrne,<sup>18</sup> Fulvio D'Angelo,<sup>21</sup> Hui K. Gan,<sup>25</sup> Mustafa Khasraw,<sup>26</sup>

(Author list continued on next page)

<sup>1</sup>The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA

<sup>2</sup>Department of Pathology, University of Texas MD Anderson Cancer Center, Houston, TX, USA

<sup>3</sup>Department of Translational Molecular Pathology, University of Texas MD Anderson Cancer Center, Houston, TX, USA

<sup>4</sup>Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA

<sup>5</sup>Amsterdam University Medical Centers/VUmc, Amsterdam, the Netherlands

<sup>6</sup>Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands

<sup>7</sup>Department of Pathology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

<sup>8</sup>School of Pharmaceutical Sciences of Ribeirao Preto, University of São Paulo, Ribeirao Preto, São Paulo, Brazil

<sup>9</sup>Hermelin Brain Tumor Center, Henry Ford Health System, Detroit, MI, USA

<sup>10</sup>service d'anatomopathologie, Hôpital de l'Enfant-Jésus du Centre hospitalier universitaire de Québec, Université Laval, Quebec City, QC, Canada

<sup>11</sup>Division of Neuropathology and Neurochemistry, Department of Neurology, Medical University of Vienna, Vienna, Austria

<sup>12</sup>National Cancer Institute, Bethesda, MD, USA

<sup>13</sup>Department of Cellular and Molecular Pathology, Leeds Teaching Hospital NHS Trust, St James's University Hospital, Leeds, UK

<sup>14</sup>Cancer and Cell Biology Division, the Translational Genomics Research Institute, Phoenix, AZ, USA

<sup>15</sup>Department of Biopharmaceutical Convergence, Department of Pharmacy, Sungkyunkwan University, Suwon-si, Gyeong gi-do, South Korea

<sup>16</sup>Department of Hematology and Stem Cell Transplantation, West German Cancer Center, University Hospital Essen, Essen, Germany

<sup>17</sup>University of Leeds, Leeds, UK

<sup>18</sup>Precision Cancer Medicine Group, Department of Physiology and Medical Physics, Royal College of Surgeons in Ireland, D02 YN77 Dublin, Ireland

(Affiliations continued on next page)

## SUMMARY

The factors driving therapy resistance in diffuse glioma remain poorly understood. To identify treatment-associated cellular and genetic changes, we analyzed RNA and/or DNA sequencing data from the temporally separated tumor pairs of 304 adult patients with isocitrate dehydrogenase (IDH)-wild-type and IDH-mutant glioma. Tumors recurred in distinct manners that were dependent on IDH mutation status and attributable to changes in histological feature composition, somatic alterations, and microenvironment interactions. Hypermutation and acquired *CDKN2A* deletions were associated with an increase in proliferating neoplastic cells at recurrence in both glioma subtypes, reflecting active tumor growth. IDH-wild-type tumors were more invasive at recurrence, and their neoplastic cells exhibited increased expression of neuronal signaling programs that reflected a possible role for neuronal interactions in promoting glioma progression. Mesenchymal transition was associated with the presence of a myeloid cell state defined by specific ligand-receptor interactions with neoplastic cells. Collectively, these recurrence-associated phenotypes represent potential targets to alter disease progression.

## INTRODUCTION

Diffuse gliomas in adults are primary tumors of the central nervous system that are characterized by a poor prognosis and

the development of resistance to a surgery and chemoradiation treatment regimen (Wen et al., 2020). Genomic profiling has identified drivers of glioma progression and led to the definition of clinically relevant subtypes based on the





Simona Migliozi,<sup>21</sup> D. Ryan Ormond,<sup>27</sup> Sun Ha Paek,<sup>20</sup> Erwin G. Van Meir,<sup>28</sup> Annemiek M.E. Walenkamp,<sup>29</sup> Colin Watts,<sup>30</sup> Tobias Weiss,<sup>31</sup> Michael Weller,<sup>31</sup> Karolina Palucka,<sup>1</sup> Lucy F. Stead,<sup>17</sup> Laila M. Poisson,<sup>19</sup> Houtan Noushmehr,<sup>9</sup> Antonio Iavarone,<sup>21,32,33,34</sup> and Roel G.W. Verhaak<sup>1,35,36,\*</sup> The GLASS Consortium

<sup>19</sup>Department of Public Health Sciences, Hermelin Brain Tumor Center, Henry Ford Health System, Detroit, MI, USA

<sup>20</sup>Seoul National University College of Medicine and Seoul National University Hospital, Seoul, Republic of Korea

<sup>21</sup>Institute for Cancer Genetics, Columbia University Medical Center, New York, NY, USA

<sup>22</sup>Case Western Reserve University School of Medicine and University Hospitals of Cleveland, Cleveland, OH, USA

<sup>23</sup>Center for Biomedical Informatics and Information Technology & Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA

<sup>24</sup>Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA

<sup>25</sup>Olivia Newton-John Cancer Research Institute, Austin Health, Melbourne, Australia

<sup>26</sup>Preston Robert Tisch Brain Tumor Center at Duke, Department of Neurosurgery, Duke University Medical Center, Durham, NC, USA

<sup>27</sup>Department of Neurosurgery, University of Colorado School of Medicine, Aurora, CO, USA

<sup>28</sup>Department of Neurosurgery, School of Medicine and O'Neal Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, AL, USA

<sup>29</sup>Department of Medical Oncology, University Medical Center Groningen, Groningen, the Netherlands

<sup>30</sup>Academic Department of Neurosurgery, Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK

<sup>31</sup>Department of Neurology, Clinical Neuroscience Center, University Hospital Zurich and University of Zürich, Switzerland

<sup>32</sup>Department of Neurology, Columbia University Medical Center, New York, NY, USA

<sup>33</sup>Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY, USA

<sup>34</sup>Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY, USA

<sup>35</sup>Department of Neurosurgery, Amsterdam University Medical Centers/VUmc, Amsterdam, the Netherlands

<sup>36</sup>Lead contact

\*Correspondence: [roel.verhaak@jax.org](mailto:roel.verhaak@jax.org)

<https://doi.org/10.1016/j.cell.2022.04.038>

presence of somatic mutations in the isocitrate dehydrogenase (IDH) genes and co-deletion of chromosome arms 1p and 19q (Cancer Genome Atlas Research Network et al., 2015; Ceccarelli et al. 2016; Louis et al., 2016). Bulk and single-cell transcriptional profiling has revealed that the gene expression programs in neoplastic glioma cells are influenced by underlying somatic alterations and interactions with the tumor microenvironment. These cells additionally exhibit high plasticity that enables them to respond dynamically to diverse challenges (Johnson et al., 2021; Nefitel et al., 2019; Venteicher et al., 2017; Wang et al., 2017). Bulk genomics approaches have revealed therapy-related mesenchymal transitions and both branching and linear evolutionary patterns (Barthel et al., 2019; Kim et al., 2015a, 2015b; Korber et al., 2019; Wang et al., 2016, 2017). However, the extent to which individual neoplastic and microenvironmental host cells interact and evolve over time to facilitate therapy resistance remains poorly understood.

To identify the drivers of treatment resistance in glioma, we established the Glioma Longitudinal Analysis Consortium (GLASS) (GLASS Consortium, 2018). We previously assembled longitudinal whole-exome and whole-genome sequencing data from 222 patients to define the clonal dynamics of glioma under therapy (Barthel et al., 2019). Here, we expand upon these analyses by integrating longitudinal genomic data with overlapping transcriptomic data. We apply single-cell-based deconvolution and spatial imaging approaches to infer a tumor's physical structure and identify the cell-state interactions across IDH-wild-type and IDH-mutant glioma. Collectively, we find that gliomas exhibit several common transcriptional and compositional changes at recurrence that represent promising therapeutic targets for delaying disease progression.

## RESULTS

### GLASS cohort

We expanded the GLASS cohort, with an emphasis on collecting orthogonal RNA sequencing profiles, to 381 patients treated across 37 hospitals (Table S1). After applying quality control filters, the final cohort comprised 304 patients, with 168 having RNA sequencing data for two or more time points, 256 having DNA sequencing data for two or more time points, and 115 having overlapping RNA and DNA available for at least two time points. The cohort of 168 patients used for RNA sequencing analyses comprised each of the three major glioma subtypes, with 128 IDH-wild-type (IDHwt), 31 IDH-mutant 1p/19q intact (IDHmut-noncode), and 9 IDH-mutant 1p/19q co-deleted (IDHmut-codel) glioma pairs (Figure 1A; Table S2). Given the limited number of IDHmut-codel cases, we pooled the IDH-mutant categories (IDHmut), unless specified otherwise. All processed genomic data and clinical annotation are available via <https://www.synapse.org/glass>.

### Longitudinal cellular heterogeneity in glioma

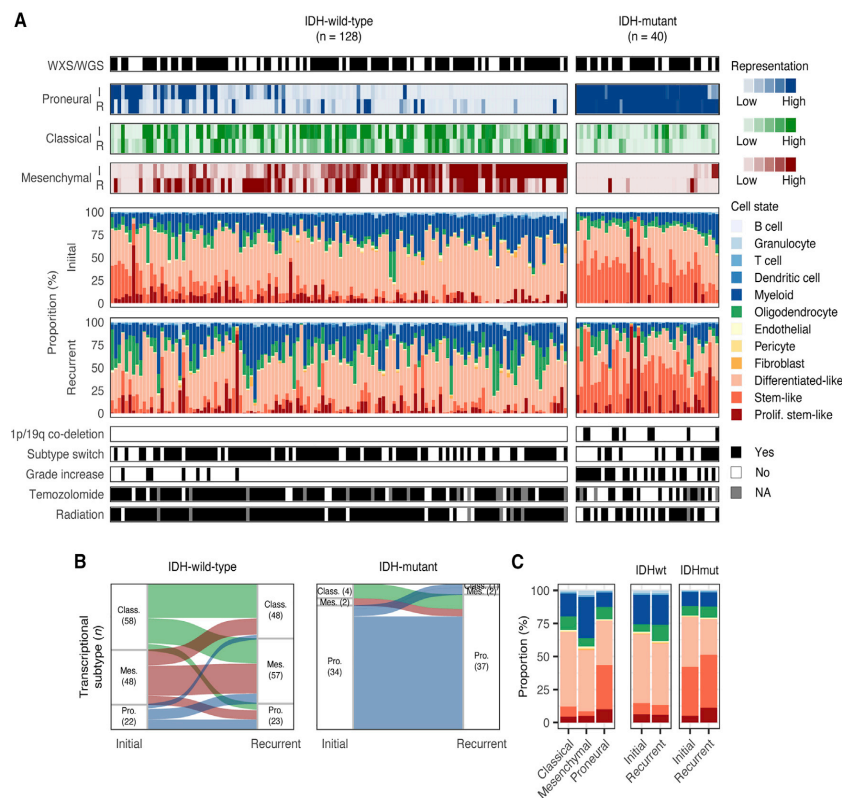
We first assessed the representation of the classical, mesenchymal, and proneural bulk transcriptional subtypes across the GLASS cohort. IDHwt tumors exhibited primarily classical and mesenchymal characteristics compared with IDHmut tumors, which were largely proneural (Figure 1A). Longitudinally, the dominant subtype in IDHwt tumors switched in 49% of patients, with classical to mesenchymal being the most common transition. In contrast, 78% of IDHmut tumors remained proneural at both time points (Figure 1B). Next, we deconvoluted the GLASS gene expression dataset by applying CIBERSORTx (Newman et al., 2019) using reference cell-state signatures derived from 55,284 single-cell transcriptomes from 11 adult

# Mutational signatures of genotoxic cancer therapies



CellPress

Cell  
Resource



**Figure 1. Longitudinal cellular heterogeneity in glioma**

(A) Each column represents an initial (I) and recurrent (R) tumor pair. Pairs are arranged based on the combined representation of the proneural and mesenchymal subtypes in their initial tumors. The first track indicates whole-exome (WXS) or whole-genome sequencing (WGS) data availability. The next three tracks indicate bulk subtype signature representation. Stacked bar plots indicate cell-state composition based on the single-cell-based deconvolution method, CIBERSORTx. (B) Sankey plot indicating whether the highest-scoring transcriptional subtype changed at recurrence. Numbers in parentheses indicate the number of samples of each subtype: proneural (Pro), classical (Class), and mesenchymal (Mes). (C) Average cell-state composition of transcriptional subtypes (left) and initial and recurrent tumors by IDH status (right).

See also Figure S1 and Tables S1, S2, and S3.

patients spanning glioma subtypes and time points (Johnson et al., 2021) (Table S3). Unsupervised analyses of this data identified 12 cell states representing the glial, stromal, immune, and neoplastic cells commonly observed in glioma. The neoplastic population expressed a shared set of markers (e.g., SOX2), and it was split across three pan-glioma cell states, differentiated-like, stem-like, and proliferating stem-like, that together captured the developmental, lineage commitment, and proliferation states observed across numerous glioma single-cell studies (Bhaduri et al., 2020; Castellan et al., 2021; Couturier et al., 2020; Garofano et al., 2021; Nefitel et al., 2019; Richards et al., 2021; Tirosh et al., 2016; Venteicher et al., 2017; Wang et al., 2019; Yuan et al., 2018). Specifically, the differentiated-like state encompassed neoplastic cells exhibiting oligodendrocyte-like, astrocyte-like (EGFR+), and mesenchymal-like (CD44+) processes, whereas the stem-like states could be

segregated by cell-cycle activity (Ki67+) and resembled undifferentiated and progenitor-like neoplastic cells (OLIG2+) (Nefitel et al., 2019; Venteicher et al., 2017). To validate this approach, we applied the method to bulk glioma RNA-seq profiles that had ground truth cellular proportions determined by (1) synthetic mixing of single-cell profiles, (2) single-cell RNA-seq, and (3) histo-cytometry of whole-slide multiplex immunofluorescence scans (Figures S1A–S1C). Together, these orthogonal analyses supported the validity of the CIBERSORTx deconvolution approach in glioma.

When deconvoluting the GLASS dataset, we observed variations in cellular composition across each subtype consistent with prior literature (Nefitel et al., 2019; Wang et al., 2017). The classical and mesenchymal subtypes were dominated by differentiated-like neoplastic cells, with mesenchymal samples also having high levels of stromal and immune cells. The proneural

subtype, in contrast, consisted mostly of proliferating stem-like and stem-like neoplastic cells (Figure 1C). Subtype compositions were similar regardless of whether the tumor was initial or recurrent, and we observed similar associations in the Cancer Genome Atlas (TCGA) glioma cohort (Figures S1D and S1E). Longitudinally, IDHwt tumors had significantly higher levels of oligodendrocytes and significantly lower levels of differentiated-like neoplastic cells at recurrence ( $p = 5e-6$  and  $4e-3$ , paired t test). These changes remained significant even after accounting for the extent of surgical resection, suggesting a greater admixture of neoplastic cells and oligodendrocytes (Figure S1F). We observed similar changes in cellular composition when using an independent IDHwt glioma cell-state classification model (Nefel et al., 2019), including a significant decrease at recurrence in the astrocyte-like neoplastic cell state that is dominant in classical tumors ( $p = 7e-3$ , paired t test; Figure S1G). Recurrent IDHmut tumors exhibited significantly higher levels of proliferating stem-like neoplastic cells and a significantly lower differentiated-like neoplastic cell fraction ( $p = 1e-3$  and  $2e-6$ , paired t test; Figure 1C). Stratifying this group by 1p/19q co-deletion status revealed that the increase in proliferating stem-like cells was only significant in the IDHmut-noncodel subtype, whereas IDHmut-codel tumors exhibited a significant increase in stem-like cells ( $p = 0.04$ , paired t test; Figure S1H). Overall, these results indicated that IDHwt and IDHmut tumors undergo distinct cell-state changes at recurrence.

### Histological features underlie subtype switching and cell-state changes at recurrence

Intratumoral heterogeneity is a hallmark of glioma and is abundant in hematoxylin and eosin-stained (H&E) tissue slides, where features such as microvascular proliferation and necrosis are used for diagnosis and grading by pathologists (Kristensen et al., 2019). The Ivy glioblastoma atlas project (Ivy GAP) has defined and microdissected five “anatomic” features on the basis of reference histology, including two features at the tumor periphery (leading-edge and infiltrating tumor) and three features in the tumor core (cellular tumor, pseudopalisading cells around necrosis, and microvascular proliferation) (Puchalski et al., 2018). Each feature exhibits a distinct transcriptional profile, suggesting that the cell-state composition changes we observed at recurrence may be related to changes in a tumor’s underlying physical structure. To better understand this relationship, we assessed the cellular composition of each feature through transcriptional deconvolution and multiplex immunofluorescence (Figures 2A and S2A). Leading-edge features, which have been shown to exhibit proneural subtype and neural tissue characteristics (Gill et al., 2014; Puchalski et al., 2018), were rich in oligodendrocytes and stem-like neoplastic cells. Pseudopalisading cells around necrosis features, which are areas of hypoxia, exhibited high levels of differentiated-like neoplastic cells. Conversely, microvascular proliferation features were enriched in proliferating stem-like neoplastic cells, supporting the role of oxygen in influencing cell state. Finally, the cellular tumor feature exhibited sample-specific variation, with high levels of differentiated-like neoplastic cells in IDHwt samples and high levels of stem-like cells in IDHmut samples. Overall, each cell state’s distribution was more significantly associated with the histological

feature than with the patient from which it was derived (two-way ANOVA; Figure S2B).

Next, we examined how feature representation varied over time by deconvoluting the GLASS dataset with Ivy GAP feature-specific gene signatures. To assess the performance, we compared the resulting proportions with pathologist estimates of related features in a subset of samples with matched H&E slides (Table S4). This revealed that the method accurately distinguished periphery- and tumor core-associated features (Figure S2C) and identified expected correlations between the pseudopalisading cells around necrosis feature and pathologist estimates of the slide area occupied by necrosis (Figure S2D). However, in recurrent samples, the deconvolution performance of some features was influenced by the presence of recurrence-specific histological features that were not profiled by Ivy GAP (Figure S2E). In GLASS, differences in each bulk transcriptional subtype’s anatomy were consistent with their underlying cell-state composition (Figure 2B). Longitudinally, IDHwt tumors had significantly higher leading-edge content at recurrence ( $p = 4e-5$ , paired t test; Figure 2B), consistent with the oligodendrocyte increase in this subtype (Figure 1C). In most cases this increase was independent of whether a tumor underwent a transcriptional subtype transition, suggesting that it was a general feature at recurrence (Figure 2C). At the cell-state level, we found that changes in the abundance of different neoplastic cell states over time associated with changes in different histological features in a subtype-dependent manner (Figure 2D). Specifically, in IDHwt tumors, changes in the abundance of differentiated-like, stem-like, and proliferating stem-like cells positively associated with changes in pseudopalisading cells around necrosis, leading edge, and microvascular proliferation features, respectively.

Given these associations, we hypothesized that some subtype switches in IDHwt tumors are attributable to changes in histological feature composition over time. To test this, we inferred the cell-state composition of each sample’s tumor core by adjusting for the presence of non-neoplastic cells and leading-edge content. Although several subtype switches were associated with changes in at least one neoplastic cell fraction pre-adjustment, the strongest difference observed post-adjustment was a decrease in stem-like cells in tumors that underwent a proneural-to-mesenchymal transition ( $p = 3e-4$ , paired t test; Figures S2F and S2G). This association remained significant after adjusting for the remaining non-cellular tumor features, suggesting that tumors undergoing this switch exhibit a loss of stem-like cells independent of histological feature composition (Figures 2E and S2F). Collectively, the results indicate that most IDHwt subtype switches relate to changes in a tumor’s underlying physical structure and microenvironment. However, changes observed in the proneural-to-mesenchymal transition may be indicative of tumor-wide changes that reflect neoplastic cell-intrinsic processes at recurrence.

### Acquired somatic alterations at recurrence associate with changes in cellular composition

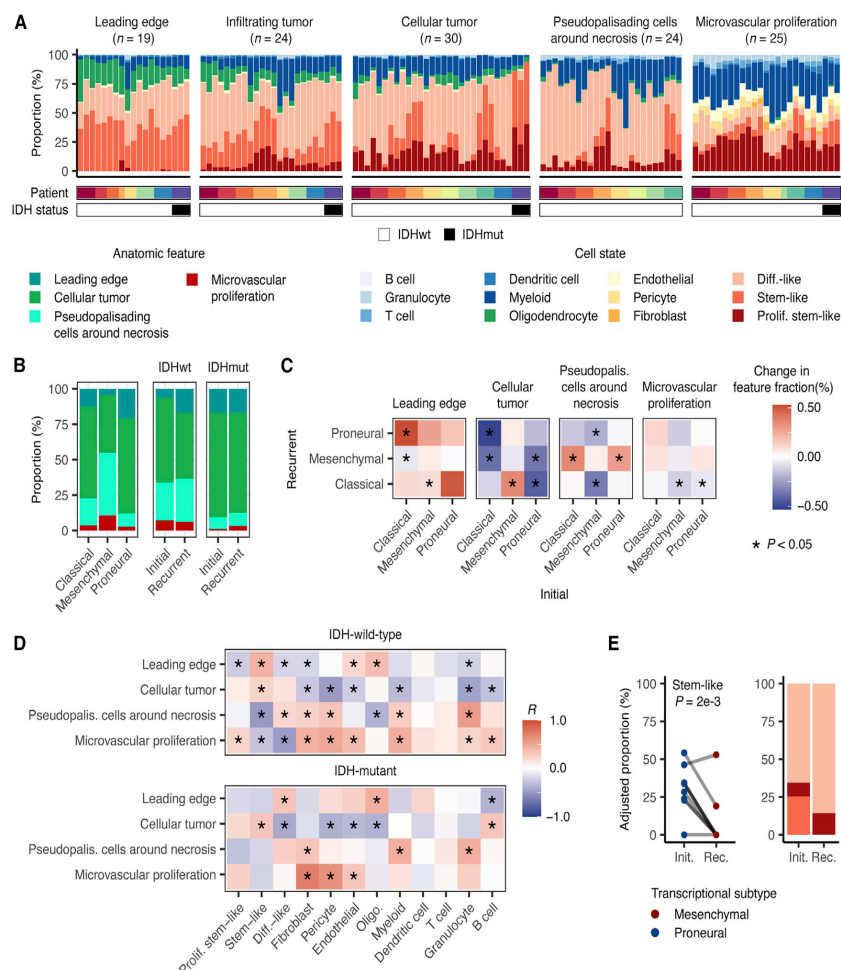
Although most tumors exhibited changes in cell state and associated histological feature composition, the factors underlying these changes remained unclear. Somatic genetic alterations

# Mutational signatures of genotoxic cancer therapies



CellPress

Cell  
Resource



**Figure 2. Histological features underlie subtype switching and cell-state changes at recurrence**

(A) Cell-state composition of each of the reference histology-defined Ivy GAP features across 10 patients. Each patient is indicated by a different color in the patient track.

(B) Average histological feature composition of transcriptional subtypes (left) and initial and recurrent tumors by IDH status (right) in GLASS.

(C) Heatmap depicting the changes in each histological feature between initial and recurrent tumors undergoing the indicated subtype transition. The initial subtype is indicated in the columns, and the recurrent subtype is indicated in the rows. Colors represent the change in the fraction of the indicated features.

(D) Heatmap depicting the Pearson correlation coefficients measuring the association between the change in a histological feature and the change in a cell state when going from an initial tumor to recurrence. In (C) and (D) \* indicates a significant correlation ( $p < 0.05$ ).

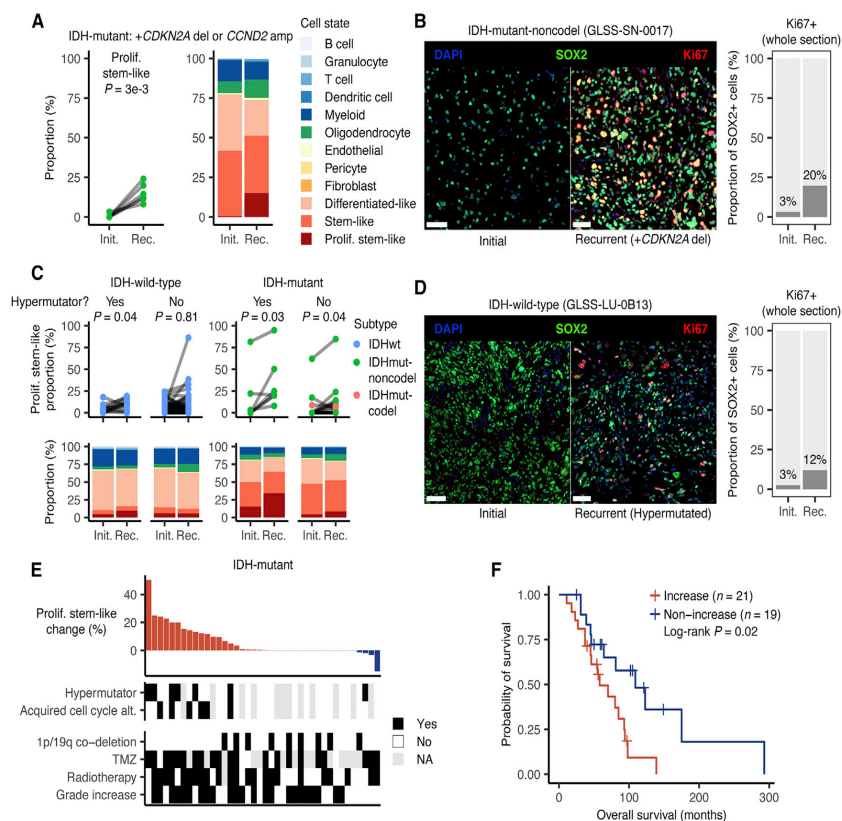
(E) Left: ladder plot depicting the change in the adjusted stem-like cell proportion between paired initial and recurrent tumors undergoing a proneural-to-mesenchymal transition. Right: the average adjusted neoplastic cell proportions for the tumor pairs outlined on the left. Neoplastic cell proportions were adjusted for the presence of non-neoplastic cells as well as non-cellular tumor content. p value from paired t test.

See also [Figure S2](#) and [Table S4](#).

have been shown to associate with cell-state distribution in IDHwt and IDHmut glioma (Johnson et al., 2021; Neftel et al., 2019; Tirosch et al., 2016; Verhaak et al., 2010). Thus, we hypothesized that genetic changes at recurrence impact cellular composition. To test this, we compared neoplastic cell-state dis-

tribution between samples that acquired or lost driver mutations. Within IDHmut tumors, this revealed acquired alterations in cell-cycle regulators, including deletions in *CDKN2A* and amplifications in *CCND2*, to be associated with increases in proliferating stem-like cells ( $p = 3e-3$ , paired t test,  $n = 6$ ; [Figure 3A](#)).

# Mutational signatures of genotoxic cancer therapies



**Figure 3. Acquired somatic alterations at recurrence associate with changes in cellular composition**

(A) Left: ladder plot depicting the change in the proliferating stem-like cell proportion between paired initial and recurrent IDHmut tumors that acquired *CDKN2A* deletions or *CCND2* amplifications. *p* value from paired *t* test. Right: stacked bar plot depicting the average proportions of each cell state for the tumor pairs in the ladder plots.

(B) Left: representative multiplex immunofluorescence images from a matched initial and recurrent IDHmut tumor pair that acquired a *CDKN2A* deletion at recurrence. Right: stacked bar plot depicting the proportion of SOX2+/Ki67+ cells among all SOX2+ cells across the entire tissue section for each sample.

(C) Top: ladder plots depicting the change in the proliferating stem-like cell proportion between paired initial and recurrent tumors, stratified by hypermutation status. Paired *t* test *p* values are indicated. Bottom: average proportions of each cell state for the tumor pairs outlined above.

(D) Left: representative multiplex immunofluorescence images from a matched initial and recurrent IDHwt tumor pair that was hypermutated at recurrence. Right: stacked bar plot depicting the proportion of SOX2+/Ki67+ cells among all SOX2+ cells across the entire tissue section for each sample.

(E) Change in proliferating stem-like cell fraction between initial and recurrent tumors from IDHmut tumor pairs.

(F) Kaplan-Meier plot depicting the survival distributions of patients that exhibited an increase or non-increase in proliferating stem-like cells at recurrence. *p* value from log-rank test.

In (B) and (D), scale bars represent 50  $\mu$ m. See also Figure S3.

Whole-slide multiplex immunofluorescence scans of a recurrent IDHmut tumor with an acquired *CDKN2A* deletion showed a significantly higher number of SOX2+/Ki67+ cells in comparison with the matched primary, confirming the association ( $p < 1e-5$ , Fisher's exact test; Figure 3B). We did not observe this association in IDHwt tumors, which typically harbor *CDKN2A* deletions at initial presentation (Figure S3A).

Next, we examined how neoplastic cell states associated with treatment-induced hypermutation, which frequently occurs after the administration of alkylating agents such as temozolomide

(Barthel et al., 2019). In both IDHwt and IDHmut glioma, hypermutation was also associated with increased proliferating stem-like neoplastic cells ( $p < 0.05$ , paired *t* test,  $n = 13$  and 7, respectively; Figure 3C). Multiplex immunofluorescence scans of an IDHwt tumor pair with a hypermutated recurrence confirmed this association, with the recurrence having a significantly higher number of SOX2+/Ki67+ cells ( $p < 1e-5$ , Fisher's exact test; Figure 3D). In IDHmut tumors, hypermutation largely occurred independent of acquired copy-number changes in *CDKN2A* and *CCND2*, suggesting that there are multiple genetic





routes to increasing proliferating stem-like neoplastic cells at recurrence (Figure 3E). Notably, neither of these alterations associated with changes in microvascular proliferation, suggesting that increases in proliferating stem-like neoplastic cells were a result of genetics and not microenvironmental interactions (Figure S3B). Survival analyses revealed that increases in proliferating stem-like neoplastic cells in IDHmut tumors were significantly associated with reduced overall survival ( $p = 0.02$ , log-rank test; Figure 3F) and remained so after adjusting for age, grade, and 1p/19q co-deletion status ( $p = 0.02$ , Wald test; Figure S3C). Collectively, these results indicate that genetic evolution at recurrence can alter neoplastic glioma cells toward a more proliferative phenotype that associates with poor prognosis.

In addition to neoplastic cells, genetic alterations have been associated with changes in the microenvironmental composition of tumors (Wellenstein and de Visser, 2018). Thus, we examined how each non-neoplastic cell state differed across tumor pairs that acquired or lost selected driver mutations at recurrence. In IDHwt tumors, non-hypermethylated recurrences that acquired *NF1* mutations all underwent a mesenchymal transition and exhibited increases in granulocytes and myeloid cells, with the granulocyte association being significant ( $p = 0.03$ , paired t test,  $n = 3$ ; Figure S3D). Additionally, several copy-number alterations, such as loss of *EGFR* or *PDGFRA* amplification, associated with increased non-neoplastic cell content ( $p < 0.05$ , paired t test,  $n = 11$  and  $n = 4$ , respectively) and a transition to the mesenchymal subtype ( $p = 0.05$ , Fisher's exact test; Figures S3E and S3F). We did not observe any significant changes in the fractions of non-neoplastic cells when comparing hypermethylated recurrences with their corresponding non-hypermethylated initial tumors, including T cells (Figure S3G). These results together indicate that although genetic evolution is a major driver of changes in neoplastic cell-state composition at recurrence, it has less of an effect on a tumor's microenvironment during this time.

### Neuronal signaling activity is increased in recurrent IDHwt tumors

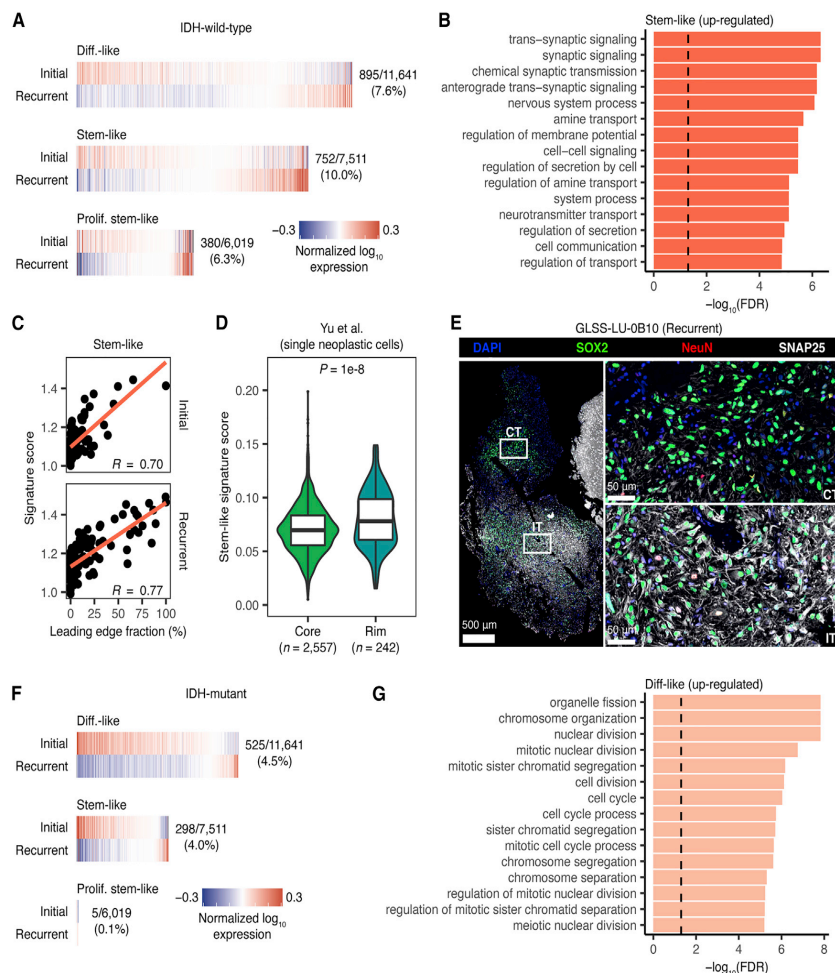
Only a subset of tumors demonstrated a genetic-associated increase in proliferating stem cell content at recurrence. We hypothesized that in remaining gliomas, treatment-induced cell-state changes may not manifest as a ubiquitous shift in cellular composition. To test this, we utilized our pan-glioma single-cell RNA-seq dataset (Johnson et al., 2021) as a reference to deconvolute GLASS bulk gene expression profiles into their component differentiated-like, stem-like, proliferating stem-like, and myeloid gene expression profiles (Figure S4A). To validate these profiles, we compared them with those derived from fluorescence-activated cell sorting (FACS)-purified glioma-specific CD45<sup>-</sup> and myeloid populations (Klemm et al., 2020). This revealed strong concordance between the corresponding profiles of each cell state (Figures S4B and S4C). Next, we compared the cell-state-specific gene expression profiles between the initial and recurrent tumor for each pair receiving temozolomide and/or radiotherapy. In IDHwt tumors, 10.0% of the 7,511 genes that could be inferred in stem-like cells were significantly differentially expressed at recurrence (false discovery rate [FDR]  $< 0.05$ , Wilcoxon signed-rank test). This number was 7.6% of the 11,641

differentiated-like state genes and 6.3% of the 6,019 proliferating stem-like state genes (Figure 4A; Table S5). Based on these results, we defined recurrence-specific signatures as the genes that were significantly up-regulated at recurrence in each cell state. Within our pan-glioma single-cell dataset, we confirmed the recurrence-specific nature of each signature by comparing their expression between single neoplastic cells from unmatched recurrent and initial tumors (Figure S4D). To understand the functions these cell states up-regulate at recurrence, we then performed gene ontology (GO) enrichment analysis on each signature. This revealed that the stem-like signature was significantly enriched in terms relating to neuronal signaling, whereas the differentiated-like and proliferating stem-like signatures exhibited similar, but weaker, associations (Figures 4B and S4E).

Given our observation that recurrent IDHwt tumors show an increase in oligodendrocytes and leading-edge content, we hypothesized that neuronal signaling in stem-like neoplastic cells may result from tumor-neuron interactions. To test this hypothesis, we examined how the stem-like neoplastic cell recurrence signature associated with histological feature content in the GLASS cohort. This revealed a positive association between neuronal signaling in the stem-like neoplastic cell-specific expression and leading-edge content (Figure 4C). Notably, we observed this result at both time points, suggesting that neuronal signaling in stem-like neoplastic cells may be driven more by tumor-neuron interactions themselves than neoplastic cell-intrinsic changes that take place at recurrence. We next evaluated neuronal signaling by comparing how the expression of the stem-like neoplastic cell recurrence signature differed between single neoplastic cells collected from the invasive rim, where there are high levels of neurons, versus those collected from the tumor core (Yu et al., 2020). This revealed significantly higher signature expression at the invasive rim, further supporting the association between neuronal signaling and tumor-neuron interactions (Figure 4D). Finally, we performed multiplex immunofluorescence to examine how neoplastic cell expression of neuronal markers differed between pathologist-annotated histological features in recurrent glioma (Figure S4F). Within the infiltrating tumor region, we found neurons (NeuN+) and a high number of neoplastic cells (SOX2+) staining positively for SNAP25, a neuronal marker that was part of our stem-like neoplastic cell recurrence signature. In contrast, there were few neurons and no SNAP25+ cells in the cellular tumor region (Figure 4E). Collectively, these results suggest that increased normal cell content at recurrence associates with higher signaling between neoplastic cells and neighboring neural cells. Although we did not observe an association between increased neuronal signaling in stem-like neoplastic cells and overall survival ( $p > 0.05$ , Wald test), neuron-to-glioma synapses have previously been implicated in increased tumor growth and invasion (Venkataramani et al., 2019; Venkatesh et al., 2015, 2017, 2019). Our results are consistent with these findings and support a model of greater tumor invasion into the normal brain at recurrence that could be facilitated by increased interactions between neurons and neoplastic cells.

The neoplastic cell-state signatures in recurrent IDHmut tumors that received treatment were distinct from those in IDHwt tumors, with the largest proportion of differentially

# Mutational signatures of genotoxic cancer therapies



**Figure 4. Neuronal signaling activity is increased in recurrent IDHwt tumors**

(A) Heatmaps depicting the average normalized  $\log_{10}$  expression level of genes that were differentially expressed between neoplastic cell states from initial and recurrent IDHwt tumors that received treatment. Fractions indicate the number of differentially expressed genes out of the number of genes inferred for that cell state's profile.

(B) Bar plot depicting the  $-\log_{10}(\text{FDR})$  from a GO enrichment analysis of the genes significantly up-regulated at recurrence in stem-like neoplastic cell-specific gene expression profiles from IDHwt tumors.

(C) Scatterplot depicting the association between leading-edge fraction and the average expression of the stem-like neoplastic cell recurrence signature for samples in the GLASS dataset.

(D) Violin plot depicting the average expression of the stem-like neoplastic cell recurrence signature in neoplastic single cells collected from the invasive rim and tumor core of 9 grade 4 gliomas (Yu et al., 2020). p value from Wilcoxon rank-sum test.

(E) Multiplex immunofluorescence images of the interface between the cellular tumor (top right; CT) and infiltrating tumor (bottom right; IT) histological features in a recurrent IDHwt tumor. Histological features were defined by a neuropathologist using the H&E image in Figure S4F.

(F) Heatmaps depicting the average normalized  $\log_{10}$  expression level of genes that were differentially expressed between neoplastic cell states from initial and recurrent IDHmut tumors that received treatment. Fractions are as outlined in (A).

(G) Bar plot depicting the  $-\log_{10}(\text{FDR})$  from a GO enrichment analysis of the genes significantly up-regulated at recurrence in differentiated-like neoplastic cell-specific gene expression profiles from IDHmut tumors.

In (B) and (G), the dashed line corresponds to  $\text{FDR} < 0.05$ . See also Figure S4 and Table S5.



expressed genes found in the differentiated-like instead of the stem-like state (FDR < 0.05, Wilcoxon signed-rank test; Figure 4F; Table S5). The genes in the differentiated-like state that were up-regulated at recurrence were significantly enriched in terms relating to the cell cycle and mitosis (FDR < 0.05; Figure 4G), whereas the stem-like signature exhibited similar associations at a relaxed significance threshold (FDR < 0.1; Figure S4G). These signatures were consistent with those found in higher-grade tumors, and accordingly, we observed that these changes were strongest in the tumor pairs that recurred at a higher grade (Figure S4H). Furthermore, when we compared signature expression in single cells of the same cell state, we found that the signatures were differentially expressed in cells derived from grade 3 versus grade 2 tumors (Figure S4I). Taken together, these results indicate that IDHwt and IDHmut tumors recur in distinct manners that reflect distinct microenvironmental and genetic contributions.

### Mesenchymal tumors exhibit a distinct myeloid cell phenotype

The mesenchymal subtype has been associated with high levels of myeloid cells, treatment resistance, and poor patient survival (Bhat et al., 2013; Kim et al., 2021; Wang et al., 2017). Thus, we sought to understand the factors driving tumors toward this subtype over time. Confirming previous findings, IDHwt tumors with a mesenchymal recurrence exhibited a significantly shorter surgical interval compared with those with non-mesenchymal recurrences ( $p = 0.03$ , log-rank test; Figure S5A) (Wang et al., 2017). However, this association was weaker in a multi-variate model (Figure S5B). Single-cell studies have previously shown that samples of this subtype exhibit high levels of neoplastic cells that express a distinct mesenchymal-like expression signature (Nefitel et al., 2019). Analysis of the neoplastic cell-state-specific expression profiles in samples undergoing a mesenchymal transition revealed that differentiated-like cells, but not stem-like cells, up-regulated this signature at recurrence (Figure S5C).

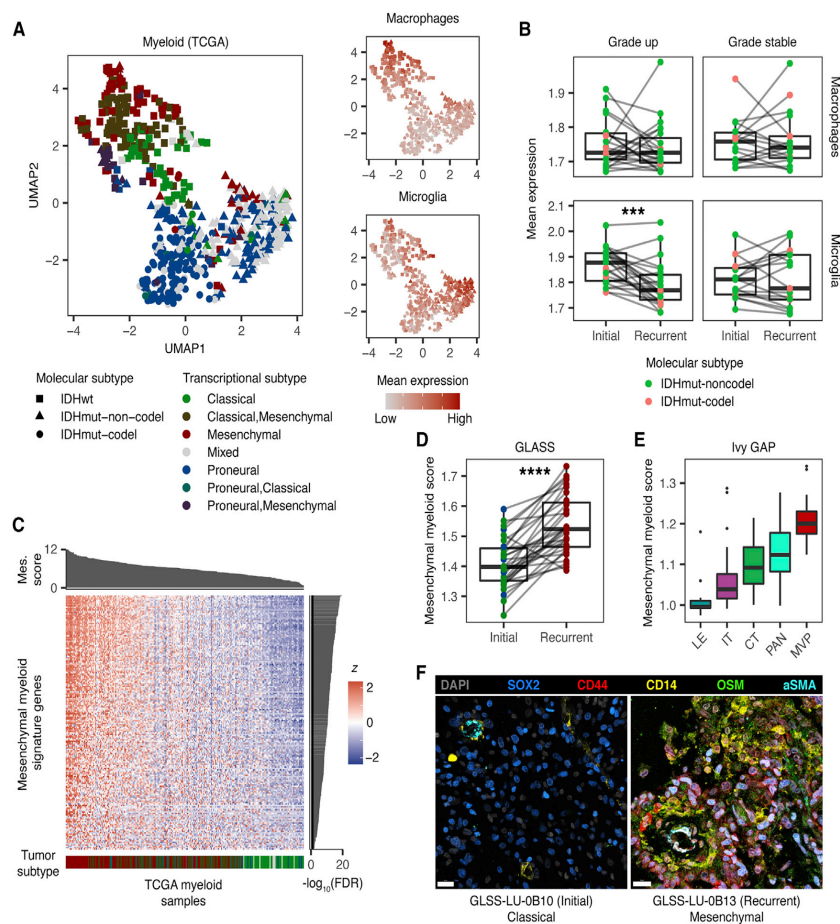
Given the changes in cellular composition associated with a mesenchymal transition, we hypothesized a role for interactions between tumor-infiltrating myeloid cells and neoplastic cells. To test this, we compared the myeloid compartment between glioma subtypes by deconvoluting the myeloid-specific gene expression profiles from 687 TCGA glioma transcriptomes. The myeloid compartment in IDHwt tumors was characterized by high blood-derived macrophage signature activity (Muller et al., 2017), whereas the myeloid cells in the IDHmut-noncode tumors exhibited a high expression of genes associated with brain-resident microglia (Figure 5A). Stratifying this cohort by transcriptional subtype revealed that the blood-derived macrophage signature was directly correlated with mesenchymal subtype representation, whereas microglial gene expression was highest among tumors of the mixed subtype classification that is seen most frequently in IDHmut-noncode glioma (Figure S5D). Consistent with these results, principal component analysis of tumor and normal brain myeloid cell expression profiles revealed that those from proneural tumors most closely resembled those from normal brain tissue, whereas mesenchymal myeloid profiles were more distinct (Figure S5E). Histological feature associations in IDHwt tumors revealed that the blood-derived macro-

phage signature was positively correlated with the abundance of microvascular proliferation and pseudopalisading cells around necrosis features, whereas the microglia signature associated with leading-edge content. In contrast, the blood-derived macrophage signature was negatively associated with leading-edge content in IDHmut tumors, whereas the microglia signature did not exhibit any clear associations (Figure S5F). Longitudinally, when holding the transcriptional subtype constant, we observed very few differentially expressed genes in the myeloid cell profiles from matched initial and recurrent tumors (Figure S5G). However, the myeloid profiles in IDHmut tumors that increased grade at recurrence exhibited a significant decrease in microglia signature expression, suggesting a shift in myeloid cell states away from brain-resident microglia ( $p = 4e-4$ , Wilcoxon signed-rank test; Figure 5B).

Macrophages are highly plastic and capable of changing their transcriptional programs in response to different stimuli (Xue et al., 2014). Thus, we hypothesized that different glioma transcriptional subtypes would exhibit distinct myeloid expression programs. To test this, we compared myeloid-specific expression profiles from each transcriptional subtype to those from normal brain tissue (Figure S5H; Table S6). Myeloid cells from the classical and mesenchymal subtypes exhibited an immunosuppressive phenotype, with each signature including genes from the blood-derived macrophage signature as well as the immune checkpoint genes, *PDCD1LG2* and *IDO1*. In addition to this shared signature, myeloid cells from mesenchymal glioma specifically up-regulated another 300 genes, suggesting distinct biology. To better understand the processes taking place in this subtype, we directly compared the myeloid gene expression profiles between mesenchymal and non-mesenchymal IDHwt tumors in TCGA. This analysis revealed a 186-gene signature that was significantly up-regulated in mesenchymal tumors (FDR < 0.05, fold-change > 1.5; Figure 5C; Table S7) and enriched in chemokine signaling and lymphocyte chemotaxis functions (Figure S5I). Expression of this signature in single myeloid cells from our single-cell dataset was strongly associated with bulk RNA-seq mesenchymal glioma subtype score from the same patient ( $R = 0.89$ ,  $p = 3e-3$ ; Figure S5J). Longitudinally, IDHwt tumors undergoing a mesenchymal transition exhibited myeloid-specific expression profiles with significantly higher expression of this signature at recurrence ( $p = 8e-8$ , Wilcoxon signed-rank test; Figure 5D).

Overall, these analyses revealed a mesenchymal-specific myeloid cell state that associated with dynamic changes in neoplastic cell expression over time. We hypothesized that these cells represent a subset of blood-derived macrophages that interact directly with mesenchymal neoplastic cells. We used the Ivy GAP dataset to determine where this myeloid cell state resided in the tumor. This revealed that the mesenchymal myeloid signature expression was strongest in the pseudopalisading cells around necrosis and microvascular proliferation features that also harbor high levels of blood-derived macrophages (Figure 5E). Correlating the myeloid-specific expression of this signature with histological feature proportions in TCGA revealed similar results (Figure S5K). Next, we performed a ligand-receptor interaction analysis to identify candidate ligand-receptor pairs that associate with mesenchymal transitions over time. To probe these interactions, we downloaded a set of 1,894





**Figure 5. Mesenchymal tumors exhibit a distinct myeloid cell phenotype**

(A) Left: uniform manifold approximation and projection (UMAP) dimensionality reduction plot of the CIBERSORTx-inferred myeloid profiles from TCGA. Right: UMAP plot colored based on the relative mean expression of macrophage and microglia signatures.

(B) Box and ladder plots depicting the difference in the mean expression of the indicated signatures between initial and recurrent IDHmut tumors from GLASS that do and do not recur at higher grades. \*\*\* indicates Wilcoxon signed-rank test p value < 1e-3.

(C) Heatmap depicting the normalized expression Z score of genes that were differentially expressed between myeloid cells from mesenchymal and non-mesenchymal TCGA tumors. The top sidebar indicates the bulk mesenchymal score of each sample divided by 1,000. The right sidebar indicates the  $-\log_{10}$  Wilcoxon rank-sum test FDR of the association for each gene. The bottom sidebar indicates the transcriptional subtype of each sample per (A).

(D) Box and ladder plots depicting the difference in the mean expression of the mesenchymal myeloid signature between initial and recurrent IDHwt tumors undergoing a mesenchymal transition in GLASS. \*\*\*\* indicates Wilcoxon signed-rank test p < 1e-5.

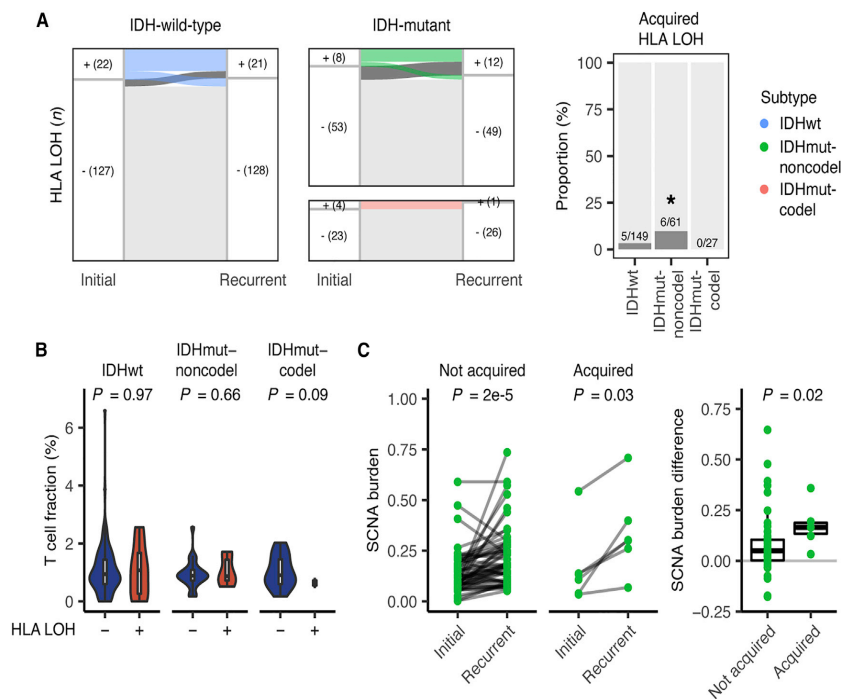
(E) Boxplot depicting the mean mesenchymal myeloid signature expression for CIBERSORTx-inferred myeloid profiles from different histological features in the Ivy GAP dataset: leading edge (LE), infiltrating tumor (IT), cellular tumor (CT), pseudopalisading cells around necrosis (PAN), and microvascular proliferation (MVP).

(F) Representative multiplex immunofluorescence images of myeloid cells near blood vessels from classical (left) and mesenchymal (right) IDHwt tumors. Scale bars represent 20  $\mu$ m.

See also Figure S5 and Tables S6, S7, and S8.

literature-supported ligand-receptor pairs (Ramilowski et al., 2015) and identified all pairs that had one component expressed in a tumor's deconvoluted myeloid profile and the other expressed in the differentiated-like neoplastic cell profile. We

then compared how the longitudinal expression change of each component associated with the change in each tumor pair's mesenchymal subtype score. This identified 105 putative ligand-receptor pairs where each component exhibited a



**Figure 6. Antigen presentation is disrupted at recurrence in IDHmut-noncode glioma**

(A) Left: Sankey plots indicating whether a tumor pair acquires or loses HLA LOH at recurrence. The colored lines indicate HLA LOH in the initial tumor, and the dark gray lines indicate acquired HLA LOH. Right: stacked bar plot indicating the proportion of samples for each glioma subtype that acquired HLA LOH at recurrence. \*Fisher's exact test, p value < 0.05.

(B) Violin plot depicting the difference in T cell proportion in samples with and without HLA LOH. p values from the t test.

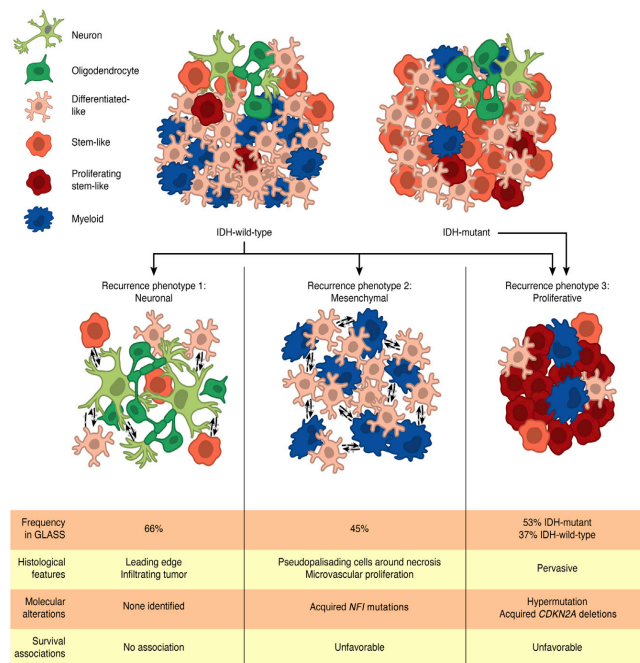
(C) Left: ladder plots depicting the change in SCNA burden between paired initial and recurrent IDHmut-noncode tumors that did and did not acquire HLA LOH. p values from Wilcoxon signed-rank test. Right: boxplot depicting the difference in the change in SCNA burden between IDHmut-noncode tumor pairs that did and did not acquire HLA LOH. p value from Wilcoxon rank-sum test.

See also Figure S6.

positive association ( $R > 0$ ,  $FDR < 0.05$ ). Of these pairs, 49 also exhibited these associations in our single-cell dataset (Table S8). This analysis revealed that the expression of oncostatin M (*OSM*) and oncostatin M receptor (*OSMR*) by myeloid cells and differentiated-like neoplastic cells, respectively, was one of the strongest correlates of the mesenchymal subtype, consistent with studies showing that this signaling associates with mesenchymal-like expression programs both *in vitro* and *in vivo* (Hara et al., 2021; Junk et al., 2017). To determine whether spatial convergence of *OSM*-expressing myeloid cells (CD14+) and mesenchymal-like neoplastic cells (CD44+/SOX2+) takes place in human tissue samples, we examined their distribution using multiplex immunofluorescence. In mesenchymal IDHwt glioma, we observed high *OSM* expression in myeloid cells near blood vessels and mesenchymal neoplastic cells, while these expression patterns were not observed in classical glioma (Figure 5F). These analyses together identify a candidate ligand-receptor interaction that can potentially be targeted to change a tumor's trajectory following treatment.

### Antigen presentation is disrupted at recurrence in IDHmut-noncode glioma

The interactions between myeloid cells and mesenchymal neoplastic cells suggest a role for the immune system in shaping glioma evolution. T cells may drive cancer evolution through the elimination of neoantigen-presenting tumor subclones (Grasso et al., 2018; McGranahan et al., 2017; Rooney et al., 2015; Rosenthal et al., 2019; Zhang et al., 2018). Although rare in glioma, these cells have been shown to select for epigenetic changes and specific genetic alterations (Gangoso et al., 2021; Kane et al., 2020) and converge with rare, recorded responses to immune checkpoint inhibition (Cloughesy et al., 2019; Zhao et al., 2019). We hypothesized that if T cell selection was taking place in glioma, we would observe high rates of loss of heterozygosity (LOH) in the human leukocyte antigen (HLA) genes that are central to the presentation of neoantigens. We observed HLA LOH in at least one time point in 19% of GLASS patients (Figure 6A). Within IDHwt and IDHmut-codel tumors, HLA LOH was found at similar rates between initial and recurrent tumors,



**Figure 7. Recurrent diffuse gliomas can be grouped into three recurrence phenotypes**  
Analysis of the GLASS dataset reveals that recurrent IDHwt and IDHmut tumors can be grouped into three recurrence phenotypes: neuronal, mesenchymal, and proliferative. Each of these phenotypes is associated with specific cellular and histological features and molecular alterations, with some also associating with poor patient survival. Some tumors can exhibit multiple phenotypes at once. Frequencies of the neuronal, mesenchymal, and proliferative phenotypes in the GLASS cohort were determined based on the number of recurrent samples that exhibited increased oligodendrocyte content, the classification as the mesenchymal transcriptional subtype, and increased proliferating stem-like content, respectively.

mors as well ( $p < 0.05$ , Wilcoxon rank-sum test; Figure S6C). Taken together, these results suggest that disruption of antigen presentation in glioma is likely a byproduct of genome-wide copy-number changes rather than being a result of selection by cytolytic T cells.

## DISCUSSION

Here, we combined longitudinal, single-cell, and spatially resolved datasets to comprehensively define the transcriptional and compositional changes that gliomas sustain at recurrence. Supplementing the treatment-associated genetic alterations we previously described (Barthel et al., 2019; Kocakavuk et al., 2021), we have identified the following three distinct recurrence-specific phenotypes: neuronal, mesenchymal, and proliferative. Through integrative profiling, we found that each phenotype converges with cellular, genetic, and histological features that emerge at recurrence (Figure 7). Consequently, they associate with less favorable clinical trajectories and are present in IDHwt and IDHmut tumors at different rates, with IDHwt tumors exhibiting all three phenotypes at recurrence and a subset of IDHmut tumors exhibiting the proliferative phenotype. Notably, these are not mutually exclusive, with some IDHwt tumors simultaneously exhibiting features associated with multiple phenotypes. Overall, this grouping offers a framework to better understand progression in diffuse glioma, and it can help guide clinical decision-making for recurrent disease.

Through single-cell- and histology-based deconvolution approaches, we observed that IDHwt tumors exhibited significant increases in oligodendrocytes and leading-edge features that were consistent with increased infiltration into the brain parenchyma at recurrence. Stem-like neoplastic cells of these tumors, especially cells at the invasive tumor margin, upregulated genes related to neuronal signaling. Overall, these changes occurred in 66% of the IDHwt tumor pairs in the GLASS cohort, suggesting that this neuronal phenotype is a frequent feature at recurrence.

with most affected pairs exhibiting this alteration at both time points. In contrast, significantly more IDHmut-noncode tumors acquired HLA LOH at recurrence ( $p = 0.04$ , Fisher's exact test), suggesting this alteration could be under positive selection. To test this, we used a simulation approach (McGranahan et al., 2017) that determined whether focal losses of the HLA genes occurred at a higher rate than expected by chance, given a sample's overall somatic copy-number alteration (SCNA) burden. In both IDHwt and IDHmut recurrences, we did not observe evidence of positive selection using this approach ( $p > 0.05$ ). Furthermore, we did not observe an association between HLA LOH status and T cell-mediated selection metrics, including the fraction of infiltrating T cells in each tumor (Figure 6B), the rates of neoantigen depletion (Figure S6A), and the number of neoantigens binding to the kept versus lost alleles (Figure S6B).

Overall, our results suggested that HLA LOH in glioma was not selected for, contrasting it with other cancer types. We hypothesized instead it was a passenger event and thus would be more likely to occur in samples with high SCNA burdens. In support of this, we found that IDHmut-noncode tumors that acquired HLA LOH at recurrence exhibited significantly larger increases in SCNA burden than those that did not ( $p = 0.02$ , Wilcoxon rank-sum test; Figure 6C). In IDHwt tumors, we did not observe these longitudinal associations. However, at both the initial and recurrent time points, IDHwt tumors with HLA LOH exhibited significantly higher SCNA burdens than those with both HLA alleles, supporting that HLA LOH is a passenger event in these tu-



This phenotype likely derives from exposure of neoplastic cells to neuronal cells that may exist at diagnosis but is more frequent at recurrence when more tumor cells have invaded the neighboring brain tissue. Our findings build upon a growing appreciation of the role that neuron-glioma interactions play in glioma invasion and progression (Venkataramani et al., 2019; Venkatesh et al., 2015, 2019).

Other cohort-wide changes in the microenvironmental composition of recurrent IDHwt tumors were limited. However, in agreement with other studies, we found that myeloid cell phenotypes varied in relation to tumor subtype and neoplastic cell state (Chen et al., 2021; Klemm et al., 2020; Ochocka et al., 2021; Pombo Antunes et al., 2021; Sa et al., 2020). Notably, myeloid cells in mesenchymal tumors exhibited a distinct transcriptional program compared with other subtypes. This signature was apparent in tumors undergoing a mesenchymal transition, suggesting a distinct myeloid cell state is involved in driving tumors toward the mesenchymal subtype in response to treatment. In support, ligand-receptor analyses revealed several candidate signaling pairs associated with mesenchymal transitions, including the previously identified OSM-OSMR interaction (Hara et al., 2021; Junk et al., 2017). Resolving the directionality of these tumor-myeloid interactions, and determining whether additional factors mediate them, will help predict which tumors will exhibit a mesenchymal phenotype at recurrence.

A subset of IDHwt and IDHmut gliomas exhibited a proliferative phenotype that was characterized by increased proliferating stem-like neoplastic cells and shorter overall survival. Analysis of the acquired somatic alterations in these tumors revealed that this phenotype was associated with temozolomide-driven hypermutation and acquired alterations of the cell-cycle regulators *CDKN2A* and *CCND2*, which have been shown to occur exclusively in post-radiation IDHmut gliomas (Kocakavuk et al., 2021). These findings suggest that tumors undergo co-occurring genetic and cell phenotype evolutionary processes in response to chemotherapy and radiotherapy. Our data highlight an unmet clinical need for tools that predict treatment-induced genetic changes and identify patients that may benefit from subsequent chemoradiation-sensitizing therapies.

Therapy resistance remains a significant obstacle for patients with diffuse glioma and must be overcome to improve quality of life and survival. Overall, our results reveal that gliomas undergo changes in cell states that associate with changes in genetics and the microenvironment, providing a baseline for building predictive models of treatment response. Future efforts by the GLASS consortium include expansion of the cohort, integration of digitized tissue sections, and associating clinical and genomic datasets with radiographic imaging data (Bakas et al., 2020). Going forward, the transcriptional and compositional changes we have identified can be integrated with imaging-based results to further assess the molecular and microenvironmental heterogeneity of glioma and identify clinically targetable factors to aid in shaping a patient's disease trajectory.

### Limitations of the study

The GLASS patient cohort consists of relatively younger individuals who had a more favorable clinical outcome resulting from the requirement they can sustain multiple surgical procedures

across their disease trajectory. Findings from our study may be less applicable to patients experiencing severe disease. Regarding the methods used in this study, bulk RNA-seq profile deconvolution approaches are limited in their ability to detect rare cellular subpopulations and can only attribute cell-state-specific expression activity to the cell states defined in their input signature matrix. The signature matrices that we used lacked neurons and astrocytes, as well as recurrence-specific histological features, limiting the direct assessment of their respective contributions to mediating treatment resistance. Due to these limitations, our analyses were mainly directed toward understanding the broad differences between longitudinal samples and transcriptional subtypes where we were well powered to make comparisons.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human subjects
- METHOD DETAILS
  - GLASS datasets
  - Public datasets
  - Whole-exome and whole-genome analysis
  - RNA preprocessing
  - Quality control
  - Bulk transcriptional subtype classification
  - Joint single-cell and bulk RNAseq dataset
  - Deconvolution analyses
  - Immunofluorescence staining and image acquisition
  - Histo-cytometry
  - Validation of cell-state proportions
  - Annotation and validation of histological features
  - Histological feature adjustment
  - Validation of cell state gene expression profiles
  - Cell-state gene expression profile analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2022.04.038>.

### ACKNOWLEDGMENTS

We thank the patients and their families for their generous donations to biomedical research. We gratefully acknowledge the genome technology, single-cell biology, and microscopy services at the Jackson Laboratory for their expert assistance and Z. Reifsnnyder (Jackson Laboratory) for the graphic design. We thank R. Puchalski for his helpful suggestions when designing the histopathology validation studies. This work is supported by the National Institutes of Health grants, R01CA237208, R21NS114873, R21CA256575, U2CCA252979 (R.G.W.V.), P30CA034196 (R.G.W.V. and K.P.), R01CA222146 (H.N., L.M.P.,

and I.D.), P30CA016672, P50CA127001 (J.T.H.), P30CA13148 (E.G.V.M.), U01CA242871, R01NS042645, and U24CA189523 (S.B.); the University of Texas M D Anderson Cancer Center Glioblastoma Moon Shots Program (J.T.H. and D.R.O.); the São Paulo Research Foundation grants 2018/00583-0, 2019/14928-1 (T.M.M.); NCI-FCRDC contract 28XS100 (E.G.V.M.); the Leeds Hospitals Charity grant 9R11/14-11 and UKRI Fellowship MR/T020504/1 (L.F.S.); the Dutch Cancer Society project 11026 (P.W.); European Union's Horizon 2020 GLIOTRAIN initiative (#766069) (A.T.B. and K.W.); the Department of Defense grants CA170278 (H.N. and T.S.S.) and W81XWH1910246 (R.G.W.V.). This work was also supported by generous gifts from the Dabbiere family (R.G.W.V.). Tissue banking within Brain Tumour Northwest is supported by the Sydney Driscoll Neuroscience Foundation and within Leeds by the Yorkshire's Brain Tumour Charity and the OSCAR's Pediatric Brain Tumour Charity. F.P.B. was supported by the JAX Scholar program and the National Cancer Institute (K99CA226387). K.C.J. was the recipient of an American Cancer Society Fellowship (130984-PF-17-141-01-DMC). F.S.V. is supported by the JAX Scholar program and a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund for Medical Research.

## AUTHOR CONTRIBUTIONS

Conceptualization, F.S.V., H.N., A. Iavarone, and R.G.W.V.; methodology, F.S.V.; validation, F.S.V., J.M., J.T.H., M.P.N., P.W., L.A.D.C., D.B., P.V.G., A.W., K.A., A. Ismail, S.K.S., D.N.-B., and K.P.; formal analysis, F.S.V., K.C.J., E.K., N.A., and K.W.; investigation, F.S.V., J.M., S.K.S., G.-L., and C.G.; resources, F.S.V., J.M., J.T.H., P.W., L.A.D.C., T.M.M., A.W., A. Ismail, H.-E.M., S.P., L.G., K.J.A., J.S.B.-S., S.B., H.K.G., D.R.O., S.H.P., E.G.V.M., A.M.E.W., M.W., K.P., L.F.S., L.M.P., and R.G.W.V.; data curation, F.S.V., K.C.J., T.M.M., T.E.W., T.S.S., F.P.B., H.K., N.A., I.D., C.W., D.R.O., L.F.S., and L.M.P.; writing—original draft, F.S.V., R.G.W.V.; writing—review & editing, F.S.V., K.C.J., J.M., J.T.H., P.W., T.M.M., P.V.G., A.W., K.A., F.P.B., E.K., I.D., J.S.B.-S., S.B., H.K.G., M.K., D.R.O., E.G.V.M., A.M.E.W., C.W., T.W., M.W., L.F.S., L.M.P., H.N., A. Iavarone, and R.G.W.V.; supervision, H.N., A. Iavarone, R.G.W.V.; all co-authors and contributors discussed the results and commented on the manuscript.

## DECLARATION OF INTERESTS

R.G.W.V. is a co-founder of Boundless Bio and a consultant for Stellanova Therapeutics. M.K. has received research funding from AbbVie and Bristol Myers Squibb, and he is on the advisory board for Janssen; he has received honoraria from the Jackson Laboratory. D.R.O. has received funding from Integra and Agios. F.P.B. has performed consulting for Bristol Myers Squibb. M.W. has received research grants from AbbVie, Adastr, Apogenix, Merck, Sharp & Dohme, Novocure, and Quercis and honoraria for lectures or advisory board participation or consulting from AbbVie, Adastr, Basilea, Bristol Meyer Squibb, Celgene, Medac, Merck, Sharp & Dohme, Merck, Nerviano Medical Sciences, Novartis, Orbus, Philogen, Roche, Tocagen, and yMabs. A.M.E.W. reported receiving institutional financial support for an advisory role from Polyphor, IPSEN, Karyopharm, and Novartis; unrestricted research grants from IPSEN and Novartis; and study budgets from AbbVie, BMS, Genzyme, Karyopharm Therapeutics, and Roche, all outside the submitted work. H.K.G. has performed consulting for AbbVie, and he is a member of the speaker bureau for AbbVie and Ignyta. K.P. is a scientific advisory board member and owns stock in Cue BioPharma.

Received: April 18, 2021  
Revised: January 22, 2022  
Accepted: April 28, 2022  
Published: May 31, 2022

## REFERENCES

Bakas, S., Ormond, D.R., Alfaro-Munoz, K.D., Smits, M., Cooper, L.A.D., Verhaak, R., and Poisson, L.M. (2020). iGLASS: imaging integration into the Glioma Longitudinal Analysis Consortium. *Neuro. Oncol* 22, 1545–1546.

Barthel, F.P., Johnson, K.C., Varn, F.S., Moskalik, A.D., Tanner, G., Kokavuk, E., Anderson, K.J., Abiola, O., Aldape, K., Alfaro, K.D., et al. (2019). Longitudinal molecular trajectories of diffuse glioma in adults. *Nature* 576, 112–120.

Bhaduri, A., Di Lullo, E., Jung, D., Müller, S., Crouch, E.E., Espinosa, C.S., Ozawa, T., Alvarado, B., Spatazza, J., Cadwell, C.R., et al. (2020). Outer radial glia-like cancer stem cells contribute to heterogeneity of glioblastoma. *Cell Stem Cell* 26, 48–63.e6.

Bhat, K.P.L., Balasubramanian, V., Vaillant, B., Ezhilarsan, R., Hummelink, K., Hollingsworth, F., Wani, K., Heathcock, L., James, J.D., Goodman, L.D., et al. (2013). Mesenchymal differentiation mediated by NF- $\kappa$ B promotes radiation resistance in glioblastoma. *Cancer Cell* 24, 331–346.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527.

Brennan, C.W., Verhaak, R.G., McKenna, A., Campos, B., Nounshmeir, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al. (2013). The somatic genomic landscape of glioblastoma. *Cell* 155, 462–477.

Cancer Genome Atlas Research Network, Brat, D.J., Verhaak, R.G., Aldape, K.D., Yung, W.K., Salama, S.R., Cooper, L.A., Rheinbay, E., Miller, C.R., Vitucci, M., et al. (2015). Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N. Engl. J. Med.* 372, 2481–2498.

Castellan, M., Guarnieri, A., Fujimura, A., Zanonato, F., Battilana, G., Panciera, T., Sladitschek, H.L., Contessotto, P., Citron, A., Grilli, A., et al. (2021). Single-cell analyses reveal YAP/TAZ as regulators of stemness and cell plasticity in Glioblastoma. *Nat. Cancer* 2, 174–188.

Ceccarelli, M., Barthel, F.P., Malta, T.M., Sabetot, T.S., Salama, S.R., Murray, B.A., Morozova, O., Newton, Y., Radenbaugh, A., Pagnotta, S.M., et al. (2016). Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* 164, 550–563.

Chen, A.X., Gartrell, R.D., Zhao, J., Upadhyayula, P.S., Zhao, W., Yuan, J., Minns, H.E., Dovas, A., Bruce, J.N., Lasorella, A., et al. (2021). Single-cell characterization of macrophages in glioblastoma reveals MARCO as a mesenchymal pro-tumor marker. *Genome Med.* 13, 88.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890.

Cloughesy, T.F., Mochizuki, A.Y., Orpilla, J.R., Hugo, W., Lee, A.H., Davidson, T.B., Wang, A.C., Ellingson, B.M., Rytlewski, J.A., Sanders, C.M., et al. (2019). Neoadjuvant anti-PD-1 immunotherapy promotes a survival benefit with intratumoral and systemic immune responses in recurrent glioblastoma. *Nat. Med.* 25, 477–486.

Cooper, L.A., Gutman, D.A., Chisolm, C., Appin, C., Kong, J., Rong, Y., Kurc, T., Van Meir, E.G., Saltz, J.H., Moreno, C.S., and Brat, D.J. (2012). The tumor microenvironment strongly impacts master transcriptional regulators and gene expression class of glioblastoma. *Am. J. Pathol.* 180, 2108–2119.

Couturier, C.P., Ayyadury, S., Le, P.U., Nadaf, J., Monlong, J., Riva, G., Allache, R., Baig, S., Yan, X., Bourgey, M., et al. (2020). Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nat. Commun.* 11, 3406.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.

Droop, A., Bruns, A., Tanner, G., Rippas, N., Morton, R., Harrison, S., King, H., Ashton, K., Syed, K., Jenkinson, M.D., et al. (2018). How to analyse the spatiotemporal tumour samples needed to investigate cancer evolution: a case study using paired primary and recurrent glioblastoma. *Int. J. Cancer* 142, 1620–1626.

Gangoso, E., Southgate, B., Bradley, L., Rus, S., Galvez-Cancino, F., McGivern, N., Güç, E., Kapourani, C.A., Byron, A., Ferguson, K.M., et al. (2021). Glioblastomas acquire myeloid-affiliated transcriptional programs via epigenetic immunoeffiting to elicit immune evasion. *Cell* 184, 2454–2470.e26.

Garofano, L., Migliozi, S., Oh, Y.T., D'Angelo, F., Najac, R.D., Ko, A., Frangaj, B., Caruso, F.P., Yu, K., Yuan, J., et al. (2021). Pathway-based classification of





- glioblastoma uncovers a mitochondrial subtype with therapeutic vulnerabilities. *Nat. Cancer* 2, 141–156.
- Gener, M.Y., Kastenmuller, W., Ifrim, I., Kabat, J., and Germain, R.N. (2012). Histo-cytometry: a method for highly multiplex quantitative tissue imaging analysis applied to dendritic cell subset microanatomy in lymph nodes. *Immunity* 37, 364–376.
- Gill, B.J., Pisapia, D.J., Malone, H.R., Goldstein, H., Lei, L., Sonabend, A., Yun, J., Samanamud, J., Sims, J.S., Banu, M., et al. (2014). MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma. *Proc. Natl. Acad. Sci. USA* 111, 12550–12555.
- GLASS Consortium (2018). Glioma through the looking glass: molecular evolution of diffuse gliomas and the Glioma Longitudinal Analysis Consortium. *Neuro. Oncol* 20, 873–884.
- Goldman, M.J., Craft, B., Hastie, M., Repčeka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* 38, 675–678.
- Grasso, C.S., Giannakis, M., Wells, D.K., Hamada, T., Mu, X.J., Quist, M., Nowak, J.A., Nishihara, R., Qian, Z.R., Inamura, K., et al. (2018). Genetic mechanisms of immune evasion in colorectal cancer. *Cancer Discov.* 8, 730–749.
- Ha, G., Roth, A., Khattra, J., Ho, J., Yap, D., Prentice, L.M., Melnyk, N., McPherson, A., Bashashati, A., Laks, E., et al. (2014). TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome Res.* 24, 1881–1893.
- Hara, T., Chanoch-Myers, R., Mathewson, N.D., Myskiw, C., Atta, L., Bussema, L., Eichhorn, S.W., Greenwald, A.C., Kinker, G.S., Rodman, C., et al. (2021). Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma. *Cancer Cell* 39, 779–792.e11.
- Hundal, J., Carreno, B.M., Petti, A.A., Linette, G.P., Griffith, O.L., Mardis, E.R., and Griffith, M. (2016). pVAC-Seq: a genome-guided in silico approach to identifying tumor neoantigens. *Genome Med.* 8, 11.
- Johnson, K.C., Anderson, K.J., Courtois, E.T., Gujar, A.D., Barthel, F.P., Varn, F.S., Luo, D., Seignon, M., Yi, E., Kim, H., et al. (2021). Single-cell multimodal glioma analyses identify epigenetic regulators of cellular plasticity and environmental stress response. *Nat. Genet.* 53, 1456–1468.
- Junk, D.J., Bryson, B.L., Smigielski, J.M., Parameswaran, N., Bartel, C.A., and Jackson, M.W. (2017). Oncostatin M promotes cancer cell plasticity through cooperative STAT3-SMAD3 signaling. *Oncogene* 36, 4001–4013.
- Kane, J.R., Zhao, J., Tsujuchi, T., Laffleur, B., Arrieta, V.A., Mahajan, A., Rao, G., Mela, A., Dmello, C., Chen, L., et al. (2020). CD8<sup>+</sup> T-cell-mediated immunoeffecting influences genomic evolution and immune evasion in murine gliomas. *Clin. Cancer Res.* 26, 4390–4401.
- Kim, H., Zheng, S., Amini, S.S., Virk, S.M., Mikkelsen, T., Brat, D.J., Grimsby, J., Sougnez, C., Muller, F., Hu, J., et al. (2015a). Whole-genome and mitosec- tor exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Res.* 25, 316–327.
- Kim, J., Lee, I.-H., Cho, H.J., Park, C.-K., Jung, Y.-S., Kim, Y., Nam, S.H., Kim, B.S., Johnson, M.D., Kong, D.-S., et al. (2015b). Spatiotemporal evolution of the primary glioblastoma genome. *Cancer Cell* 28, 318–328.
- Kim, Y., Varn, F.S., Park, S.H., Yoon, B.W., Park, H.R., Lee, C., Verhaak, R.G.W., and Paek, S.H. (2021). Perspective of mesenchymal transformation in glioblastoma. *Acta Neuropathol. Commun.* 9, 50.
- Klemm, F., Maas, R.R., Bowman, R.L., Kornete, M., Soukup, K., Nassiri, S., Brouland, J.-P., Iacobuzio-Donahue, C.A., Brennan, C., Tabar, V., et al. (2020). Interrogation of the microenvironmental landscape in brain tumors reveals disease-specific alterations of immune cells. *Cell* 181, 1643–1660.e17.
- Kocakavuk, E., Anderson, K.J., Varn, F.S., Johnson, K.C., Amin, S.B., Sulman, E.P., Lolkema, M.P., Barthel, F.P., and Verhaak, R.G.W. (2021). Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer. *Nat. Genet.* 53, 1088–1096.
- Korber, V., Yang, J., Barah, P., Wu, Y., Stichel, D., Gu, Z., Fletcher, M.N.C., Jones, D., Hentschel, B., Lamszus, K., et al. (2019). Evolutionary trajectories of IDH(WT) glioblastomas reveal a common path of early tumorigenesis instigated years ahead of initial diagnosis. *Cancer Cell* 35, 692–704.e12.
- Kristensen, B.W., Priesterbach-Ackley, L.P., Petersen, J.K., and Wesseling, P. (2019). Molecular pathology of tumors of the central nervous system. *Ann. Oncol.* 30, 1265–1278.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Louis, D.N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., and Ellison, D.W. (2016). The 2016 World Health Organization Classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* 131, 803–820.
- Mazor, T., Pankov, A., Johnson, B.E., Hong, C., Hamilton, E.G., Bell, R.J.A., Smirnov, I.V., Reis, G.F., Phillips, J.J., Barnes, M.J., et al. (2015). DNA methylation and somatic mutations converge on the cell cycle and define similar evolutionary histories in brain tumors. *Cancer Cell* 28, 307–317.
- McGranahan, N., Rosenthal, R., Hiley, C.T., Rowan, A.J., Watkins, T.B.K., Wilson, G.A., Birbak, N.J., Veeriah, S., Van Loo, P., Herrero, J., et al. (2017). Allele-specific HLA loss and immune escape in lung cancer evolution. *Cell* 171, 1259–1271.e11.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- Muller, S., Kohanbash, G., Liu, S.J., Alvarado, B., Carrera, D., Bhaduri, A., Watchmaker, P.B., Yagnik, G., Di Lullo, E., Malatesta, M., et al. (2017). Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment. *Genome Biol.* 18, 234.
- Nefel, C., Laffy, J., Filbin, M.G., Hara, T., Shore, M.E., Rahme, G.J., Richman, A.R., Silverbush, D., Shaw, M.L., Hebert, C.M., et al. (2019). An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* 178, 835–849.e21.
- Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* 37, 773–782.
- Ochocka, N., Segit, P., Walentynowicz, K.A., Wojnicki, K., Cyranowski, S., Swatler, J., Mieczkowski, J., and Kaminska, B. (2021). Single-cell RNA sequencing reveals functional heterogeneity of glioma-associated brain macrophages. *Nat. Commun.* 12, 1151.
- Polański, K., Young, M.D., Miao, Z., Meyer, K.B., Teichmann, S.A., and Park, J.E. (2020). BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* 36, 964–965.
- Pombo Antunes, A.R., Scheytjens, I., Lodi, F., Messiaen, J., Antoranz, A., Duerinck, J., Kancheva, D., Martens, L., De Vlamincq, K., Van Hove, H., et al. (2021). Single-cell profiling of myeloid cells in glioblastoma across species and disease stage reveals macrophage competition and specialization. *Nat. Neurosci.* 24, 595–610.
- Puchalski, R.B., Shah, N., Miller, J., Dalley, R., Nomura, S.R., Yoon, J.-G., Smith, K.A., Lankovitch, M., Bertagnoli, D., Bickley, K., et al. (2018). An anatomic transcriptional atlas of human glioblastoma. *Science* 360, 660–663.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Ramilowski, J.A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Sataogom, V.P., Itoh, M., Kawaji, H., Carninci, P., Rost, B., and Forrest, A.R.R. (2015). A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* 6, 7866.
- Richards, L.M., Whitley, O.K.N., MacLeod, G., Cavalli, F.M.G., Coutinho, F.J., Jaramillo, J.E., Svergun, N., Riverin, M., Croucher, D.C., Kushida, M., et al. (2021). Gradient of developmental and injury response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity. *Nat. Cancer* 2, 157–173.

- Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61.
- Rosenthal, R., Cadieux, E.L., Salgado, R., Bakir, M.A., Moore, D.A., Hiley, C.T., Lund, T., Tanić, M., Reading, J.L., Joshi, K., et al. (2019). Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479–485.
- Sa, J.K., Chang, N., Lee, H.W., Cho, H.J., Ceccarelli, M., Cerulo, L., Yin, J., Kim, S.S., Caruso, F.P., Lee, M., et al. (2020). Transcriptional regulatory networks of tumor-associated macrophages that drive malignancy in mesenchymal glioblastoma. *Genome Biol.* **21**, 216.
- Steen, C.B., Liu, C.L., Alizadeh, A.A., and Newman, A.M. (2020). Profiling cell type abundance and expression in bulk tissues with CIBERSORTx. In *Stem Cell Transcriptional Networks: Methods and Protocols*, B.L. Kidder, ed. (Springer), pp. 135–157.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21.
- Szolek, A., Schubert, B., Mohr, C., Sturm, M., Feldhahn, M., and Kohlbacher, O. (2014). OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* **30**, 3310–3316.
- Tirosh, I., Venteicher, A.S., Hebert, C., Escalante, L.E., Patel, A.P., Yizhak, K., Fisher, J.M., Rodman, C., Mount, C., Filbin, M.G., et al. (2016). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309–313.
- Venkataramani, V., Tanev, D.I., Strahle, C., Studier-Fischer, A., Fankhauser, L., Kessler, T., Körber, C., Kardorf, M., Ratliff, M., Xie, R., et al. (2019). Glutamate synaptic input to glioma cells drives brain tumour progression. *Nature* **573**, 532–538.
- Venkatesh, H.S., Johung, T.B., Caretti, V., Noll, A., Tang, Y., Nagaraja, S., Gibson, E.M., Mount, C.W., Polepalli, J., Mitra, S.S., et al. (2015). Neuronal activity promotes glioma growth through Neurotrophin-3 secretion. *Cell* **161**, 803–816.
- Venkatesh, H.S., Morishita, W., Geraghty, A.C., Silverbush, D., Gillespie, S.M., Arzt, M., Tam, L.T., Espenel, C., Ponnuswami, A., Ni, L., et al. (2019). Electrical and synaptic integration of glioma into neural circuits. *Nature* **573**, 539–545.
- Venkatesh, H.S., Tam, L.T., Woo, P.J., Lennon, J., Nagaraja, S., Gillespie, S.M., Ni, J., Duveau, D.Y., Morris, P.J., Zhao, J.J., et al. (2017). Targeting neuronal activity-regulated neurotrophin-3 dependency in high-grade glioma. *Nature* **549**, 533–537.
- Venteicher, A.S., Tirosh, I., Hebert, C., Yizhak, K., Neftci, C., Filbin, M.G., Hovestadt, V., Escalante, L.E., Shaw, M.L., Rodman, C., et al. (2017). Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**, eaai8478.
- Verhaak, R.G., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110.
- Wang, J., Cazzato, E., Ladewig, E., Frattini, V., Rosenbloom, D.I., Zairis, S., Abate, F., Liu, Z., Elliott, O., Shin, Y.-J., et al. (2016). Clonal evolution of glioblastoma under therapy. *Nat. Genet.* **48**, 768–776.
- Wang, L., Babikir, H., Müller, S., Yagnik, G., Shamardani, K., Catalan, F., Kohanbash, G., Alvarado, B., Di Lullo, E., Kriegstein, A., et al. (2019). The phenotypes of proliferating glioblastoma cells reside on a single axis of variation. *Cancer Discov.* **9**, 1708–1719.
- Wang, M., Yao, L.-C., Cheng, M., Cai, D., Martinek, J., Pan, C.-X., Shi, W., Ma, A.-H., De Vere White, R.W., Airhart, S., et al. (2018). Humanized mice in studying efficacy and mechanisms of PD-1-targeted cancer immunotherapy. *FASEB J.* **32**, 1537–1549.
- Wang, Q., Hu, B., Hu, X., Kim, H., Squatrito, M., Scarpace, L., deCarvalho, A.C., Lyu, S., Li, P., Li, Y., et al. (2017). Tumor evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with immunological changes in the microenvironment. *Cancer Cell* **32**, 42–56.e6.
- Wellenstein, M.D., and de Visser, K.E. (2018). Cancer-cell-intrinsic mechanisms shaping the tumor immune landscape. *Immunity* **48**, 399–416.
- Wen, P.Y., Weller, M., Lee, E.Q., Alexander, B.M., Barnholtz-Sloan, J.S., Barthel, F.P., Batchelor, T.T., Brindley, R.S., Chang, S.M., Chiocca, E.A., et al. (2020). 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. Oncol.* **22**, 1073–1113.
- Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15.
- Wu, T.-C., Xu, K., Martinek, J., Young, R.R., Banchev, R., George, J., Turner, J., Kim, K.I., Zurawski, S., Wang, X., et al. (2018). IL1 receptor antagonist controls transcriptional signature of inflammation in patients with metastatic breast cancer. *Cancer Res.* **78**, 5243–5258.
- Xue, J., Schmidt, S.V., Sander, J., Draffehn, A., Krebs, W., Quester, I., De Nardo, D., Gohel, T.D., Emde, M., Schmidleithner, L., et al. (2014). Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity* **40**, 274–288.
- Yu, K., Hu, Y., Wu, F., Guo, Q., Qian, Z., Hu, W., Chen, J., Wang, K., Fan, X., Wu, X., et al. (2020). Surveying brain tumor heterogeneity by single-cell RNA-sequencing of multi-sector biopsies. *Natl. Sci. Rev.* **7**, 1306–1318.
- Yuan, J., Levitin, H.M., Frattini, V., Bush, E.C., Boyett, D.M., Samanamud, J., Ceccarelli, M., Dovas, A., Zanazzi, G., Canoll, P., et al. (2018). Single-cell transcriptome analysis of lineage diversity in high-grade glioma. *Genome Med.* **10**, 57.
- Zhang, A.W., McPherson, A., Milne, K., Kroeger, D.R., Hamilton, P.T., Miranda, A., Funnell, T., Little, N., de Souza, C.P.E., Laan, S., et al. (2018). Interfaces of malignant and immunologic clonal dynamics in ovarian cancer. *Cell* **173**, 1755–1769.e22.
- Zhao, J., Chen, A.X., Gartrell, R.D., Silverman, A.M., Aparicio, L., Chu, T., Bordbar, D., Shan, D., Samanamud, J., Mahajan, A., et al. (2019). Immune and genomic correlates of response to anti-PD-1 immunotherapy in glioblastoma. *Nat. Med.* **25**, 462–469.

# Mutational signatures of genotoxic cancer therapies



CellPress

Cell  
Resource

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
BV421 anti KI67	BD Biosciences	RRID:AB_2686897
AF488 anti SOX2	Thermo Fisher Scientific	RRID:AB_2574479
AF555 anti EGFR	Cell Signaling Technology	Cat#5108S
Rabbit anti CD14	Abcam	RRID:AB_2889158
AF647 anti Olig2	Abcam	Cat#ab225100
AF700 anti CD44	Novus Biologicals	Cat#NBP1-41266AF700
AF568 Goat anti Rabbit Highly cross absorbed secondary antibody	Thermo Fisher Scientific	RRID:AB_10563566
AF594 anti SNAP25	Novus Biologicals	Cat#NBP2-74245AF594
AF700 anti NeuN	Novus Biologicals	Cat#NBP1-92693AF700
AF647 anti alpha SMA	Novus Biologicals	Cat#NBP2-34522AF647
JF549 anti Oncostatin M	Novus Biologicals	Cat#NB120-10842JF549
<b>Biological samples</b>		
Glioma tissue and matched normal blood	Henry Ford Health System	N/A
Glioma tissue and matched normal blood	Seoul National University	N/A
<b>Chemicals</b>		
Histo-Clear	National Diagnostics	Cat#HS-202
Antigen Retrieval Buffer (Citrate, pH6)	Abcam	Cat#ab93678
Fc Receptor Blocker	Innovex	Cat#NB309
Background Buster	Innovex	Cat#NB306
Fluoromount G	SouthernBiotech	Cat#0100-01
Cover Glass	Thermo Scientific	Cat#152450
Slides	Denville Scientific	Cat#M1021
Saponin	Sigma	Cat#S7900-100G
Triton X-100	Sigma	Cat#T8787
Bovine serum albumin IgG free	Jackson Immuno Research	RRID:AB_2336946
Normal rabbit serum	Jackson Immuno Research	RRID:AB_2337123
Sytox blue	Thermo Fisher	Cat#S11348
DAPI	Thermo Fisher	Cat#D1306
<b>Critical commercial assays</b>		
AllPrep DNA/RNA Mini Kit	Qiagen	Cat#80204
SureSelectXT Low-Input Reagent Kit	Agilent	Cat#5191-4080
SureSelectXT Human All Exon V6 +COSMIC	Agilent	Cat#5190-9307
QIAamp DNA Mini Kit	Qiagen	Cat#51104
KAPA Hyper Prep Kit (Illumina)	Roche	Cat#7962363001
KAPA mRNA Hyperprep Kit	Roche	Cat#8098123702
Tempus xO Assay	Tempus	N/A
KAPA stranded mRNAseq Kit	Roche	Cat#7962207001
NuGEN Ovation RNAseq System	Tecan	Cat#7102-A01
<b>Deposited data</b>		
Processed DNA somatic alteration data	This paper	<a href="https://www.synapse.org/glass">https://www.synapse.org/glass</a>
RNAseq pseudocount and TPM data	This paper	<a href="https://www.synapse.org/glass">https://www.synapse.org/glass</a>

(Continued on next page)



# Mutational signatures of genotoxic cancer therapies

<i>Continued</i>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Digitized H&E images	This paper	<a href="https://styx.neurology.emory.edu/girder/#collection/625dda70622f966e826a0446">https://styx.neurology.emory.edu/girder/#collection/625dda70622f966e826a0446</a>
Custom pipeline and analysis code	This paper	<a href="https://github.com/fsvarn/GLASSx/">https://github.com/fsvarn/GLASSx/</a>
Longitudinal glioma RNAseq fastq files	European Genome Phenome Archive	EGAS00001001041
Longitudinal glioma RNAseq fastq files	European Genome Phenome Archive	EGAS00001001880
Longitudinal glioma RNAseq bam files	European Genome Phenome Archive	EGAS00001001033
Longitudinal glioma RNAseq bam files	European Genome Phenome Archive	EGAS00001001255
Longitudinal glioma RNAseq fastq files	European Genome Phenome Archive	EGAS00001003790
Longitudinal glioma RNAseq fastq files	Sequencing Read Archive	BioProject#PRJNA320312
Longitudinal glioma whole exome and RNAseq fastq files	Sequencing Read Archive	BioProject# PRJNA482620
Longitudinal TCGA GBM LGG RNAseq fastq files	Genomic Data Commons	<a href="https://portal.gdc.cancer.gov/">https://portal.gdc.cancer.gov/</a>
TOIL TCGA TARGET GTEx RNAseq TPM data	University of California Santa Cruz Xenabrowser	<a href="https://xenabrowser.net/datapages/">https://xenabrowser.net/datapages/</a>
Ivy Glioblastoma Atlas Project RNAseq FPKM data	Ivy GAP	<a href="https://glioblastoma.alleninstitute.org/static/download.html">https://glioblastoma.alleninstitute.org/static/download.html</a>
Processed glioblastoma single-cell data	Broad Single Cell Portal	Study: Single cell RNA-seq of adult and pediatric glioblastoma
Multi-sector single-cell glioma RNAseq count data	Gene Expression Omnibus	GSE117891
RNAseq count data from FACS-sorted glioma cell populations	BrainTIME Portal	<a href="https://joycelab.shinyapps.io/braintime/">https://joycelab.shinyapps.io/braintime/</a>
b37 reference genome (human_g1k_v37_decoy)	GATK Resource Bundle	<a href="https://gatk.broadinstitute.org/hc/en-us/articles/360035890811-Resource-bundle">https://gatk.broadinstitute.org/hc/en-us/articles/360035890811-Resource-bundle</a>
Pan-glioma single-cell RNAseq data	Synapse	<a href="https://www.synapse.org/#/Synapse:syn26375549">https://www.synapse.org/#/Synapse:syn26375549</a>
<i>Software and algorithms</i>		
bedtools	Quinlan and Hall, 2010	<a href="https://bedtools.readthedocs.io/en/latest/">https://bedtools.readthedocs.io/en/latest/</a>
Seuratv3.2.3	Stuart et al., 2019	<a href="https://satijalab.org/seurat/">https://satijalab.org/seurat/</a>
BWA MEM 0.7.17	Li and Durbin, 2009	<a href="http://bio-bwa.sourceforge.net/">http://bio-bwa.sourceforge.net/</a>
GATK 4.0.10.1	McKenna et al., 2010	<a href="https://gatk.broadinstitute.org/hc/en-us">https://gatk.broadinstitute.org/hc/en-us</a>
TITAN	Ha et al., 2014	<a href="https://github.com/gavinha/TitanCNA">https://github.com/gavinha/TitanCNA</a>
OptiType v1.3.2	Szolek et al., 2014	<a href="https://github.com/FRED-2/OptiType">https://github.com/FRED-2/OptiType</a>
pVACseq v4.0.10	Hundal et al., 2016	<a href="https://pvac-seq.readthedocs.io/en/latest/">https://pvac-seq.readthedocs.io/en/latest/</a>
LOHHLA	McGranahan et al., 2017	<a href="https://bitbucket.org/mcgranahanlab/lohlla/">https://bitbucket.org/mcgranahanlab/lohlla/</a>
STARv2.7.5	Dobin et al., 2013	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
fastp v0.20.0	Chen et al., 2018	<a href="https://github.com/OpenGene/fastp">https://github.com/OpenGene/fastp</a>
kallisto v0.46.0	Bray et al., 2016	<a href="https://pachterlab.github.io/kallisto/">https://pachterlab.github.io/kallisto/</a>
ssgsea.GBM.classification	Wang et al., 2017	N/A
CIBERSORTx webservice	Newman et al., 2019	<a href="https://cibersortx.stanford.edu/">https://cibersortx.stanford.edu/</a>
CIBERSORTx docker	Newman et al., 2019	<a href="https://hub.docker.com/r/cibersortx/hires">https://hub.docker.com/r/cibersortx/hires</a>
Imaris 9.0.2 and 9.4	Bitplane	<a href="http://www.bitplane.com/Imaris/Imaris">http://www.bitplane.com/Imaris/Imaris</a>
Flowjo v10	Flowjo LLC	<a href="https://www.flowjo.com/solutions/flowjo">https://www.flowjo.com/solutions/flowjo</a>
R v3.6.1	The R Project for Statistical Computing	<a href="https://www.r-project.org/">https://www.r-project.org/</a>
topGO v2.38.1	Bioconductor	<a href="https://bioconductor.org/packages/release/bioc/html/topGO.html">https://bioconductor.org/packages/release/bioc/html/topGO.html</a>



## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources should be directed to the lead contact, Roel Verhaak ([Roel.Verhaak@jax.org](mailto:Roel.Verhaak@jax.org)).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

- Estimated count and transcript per million gene expression matrices as well as mutation calls, copy number calls, and all downstream tables generated for this study can be downloaded on Synapse (<https://www.synapse.org/glass>). Digitized H&E images are available on the Digital Slide Archive (<https://styx.neurology.emory.edu/girder/#collection/625dda70622f966e826a0446>).
- All custom scripts, pipelines, and code used in data processing and figure creation is available on the project's Github repository (<https://github.com/fsvrn/GLASSx>).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Human subjects

Human tissue collection was performed with written informed consent from patients. The protocol used to collect and sequence specimens, or collect and analyze data from patients, was approved by the institutional review board (IRB) of the Jackson Laboratory (17-007 DUA MDA 17349-JGM, 17-008 DUA MDA 17425-JGM, 18-003 DUA Kyoto-JGM, 18-004 DUA Samsung-JGM, 18-005 DUA Vienna-JGM, 18-006 DUA Dana-JGM, 18-008 DUA LEEDS-JGM, 19-007 DUA DKFZ-JGM, 19-008 DUA Case-JGM, 20-005 DUA HenryF-JGM, 20-015 DUA MDA-JGM, 20-26541 dbGaP-JGM, 2019-057-JGM, 2019-084-JGM). Patients were males and females. Clinical characteristics of the cohort are summarized in [Table S2](#).

## METHOD DETAILS

### GLASS datasets

Datasets added to GLASS came from both published and unpublished sources ([Table S1](#)). Collectively, the newly added data consisted of DNA sequencing data from 109 glioma samples (53 patients) and RNA sequencing data from 392 samples (206 patients).

Newly generated DNA and RNA sequencing data was collected for a cohort of frozen samples from Henry Ford Health System and Seoul National University. For the Henry Ford Health System cohort, DNA and RNA were simultaneously extracted from each sample using the AllPrep DNA/RNA Mini Kit from Qiagen (#80204). Exon capture was then performed using the Agilent SureSelectXT Low-Input Reagent Kit and the V6 + COSMIC capture library and the resulting reads were subjected to 150 base pair paired-end sequencing at the University of Southern California using an Illumina NovaSeq 6000. RNA from these tissues was processed and sequenced at Psomagen. For the Seoul National University cohort, DNA and RNA were simultaneously extracted from each tumor sample at The Jackson Laboratory for Genomic Medicine using the AllPrep DNA/RNA Mini Kit from Qiagen (#80204). For blood samples, DNA was extracted using the QIAamp DNA Mini and Blood Mini Kit from Qiagen (#51104). 200ng of DNA was sheared to 400bp using a LE220 focused-ultrasonicator (Covaris) and size selected using Ampure XP beads (Beckman Coulter). The fragments were treated with end-repair, A-tailing, and ligation of Illumina unique adapters (Illumina) using the KAPA Hyper Prep Kit (Illumina) from Roche (#7962363001). Whole genome libraries were then subjected to 150 base pair paired-end sequencing on the Illumina NovaSeq 6000 to achieve 25X coverage for normal samples and >35-40X coverage for tumor samples. RNAseq libraries were prepared with the KAPA mRNA Hyperprep kit from Roche (#8098123702) and then sequenced using an Illumina NovaSeq 6000 platform generating paired-end reads of 150 base pairs.

New RNAseq data was also generated for cohorts coming from Case Western Reserve University, the Chinese University of Hong Kong, the Luxembourg Institute of Health, and MD Anderson Cancer Center. For Case Western Reserve University, RNA from frozen tissues was processed at Tempus (Chicago, IL) using the Tempus xO assay and then sequencing using an Illumina HiSeq 4000 platform. For the Chinese University of Hong Kong cohort, RNAseq libraries were prepared with the KAPA Stranded mRNAseq kit (Roche) and then sequenced at The Jackson Laboratory for Genomic Medicine using an Illumina HiSeq4000 platform generating paired-end reads of 75 base pairs. For the Luxembourg Institute of Health cohort, RNAseq libraries were prepared with the KAPA mRNA Hyperprep kit (Roche) and then sequenced at The Jackson Laboratory for Genomic Medicine using an Illumina NovaSeq 6000 platform generating paired-end reads of 150 base pairs. For the MD Anderson cohort, purified double-stranded cDNA generated from 150 ng of FFPE sample-derived RNA was prepared using the NuGEN Ovation RNAseq System and subjected to paired-end sequencing using an Illumina HiSeq 2000 or HiSeq 2500 platform.

The remaining datasets were generated as described in their respective publications (cited below). For most of these cohorts, whole exome and/or whole genome sequencing data were downloaded and processed as described during creation of the initial

GLASS dataset (Barthel et al., 2019). RNAseq fastq files from the Samsung Medical Center (SM) cohort were delivered via hard disk and are available to download from the European Genome-Phenome Archive (EGA) under accession numbers EGAS00001001041 and EGAS00001001880 (Kim et al., 2015b; Wang et al., 2016). RNAseq bam files for the original Henry Ford Health System (HF) and the University of California San Francisco (SF) cohorts were downloaded from EGA under accession numbers EGAS00001001033 and EGAS00001001255, respectively, and converted to fastq files for subsequent processing using bedtools (Kim et al., 2015a; Mazor et al., 2015; Quinlan and Hall, 2010). RNAseq fastq files for the University of Leeds (LU) cohort were downloaded from EGA under accession number EGAS00001003790 (Droop et al., 2018). For the first Columbia cohort (CU-R), which consisted of samples originally collected from the Istituto Neurologico C. Besta, RNAfastq files were delivered via hard disk and are available to download at the Sequencing Read Archive (SRA) under BioProject number PRJNA320312 (Wang et al., 2016). For the second Columbia cohort (CU-P), which featured samples that had been treated with immune checkpoint inhibitors, raw fastq reads for whole exome and RNA-seq were obtained from SRA under BioProject number PRJNA482620 (Zhao et al., 2019). RNAseq fastq files from the Low Grade Glioma (LGG) and Glioblastoma Multiforme (GBM) projects in TCGA were obtained from the Genomic Data Commons legacy archive (<https://portal.gdc.cancer.gov/legacy-archive/>) (Brennan et al., 2013; Cancer Genome Atlas Research Network et al., 2015).

For each dataset, clinical data based on patient medical records was provided by the participating institution or, when not available, was obtained from the dataset's respective publication. To create a shared clinical dataset to be used throughout the study, clinical data sheets were combined and organized into a common set of variables as previously described in the supplemental information of the original GLASS study (Barthel et al., 2019).

### Public datasets

Processed and batch-corrected RNAseq data from the TCGA and Genotype-Tissue Expression (GTEx) projects were obtained from the University of California Santa Cruz Xenabrowser (cohort: TCGA TARGET GTEx, dataset ID: TcgaTargetGtex\_rsem\_gene\_tpm, author: UCSC TOIL RNA-seq recompute) (Goldman et al., 2020), and then subset to only include TCGA glioma (GBM/LGG), GTEx Brain Frontal Cortex, and GTEx Cortex samples. Normalized gene-level fragments per kilobase million (FPKM) for the Ivy Glioblastoma Atlas Project (Ivy GAP) dataset were obtained from the Ivy GAP website (<https://glioblastoma.alleninstitute.org/static/download.html>) (Puchalski et al., 2018). Processed single-cell data and associated metadata for a set of 28 IDHwt glioblastomas processed using SmartSeq2 was obtained from the Broad Single Cell Portal (Study: Single cell RNA-seq of adult and pediatric glioblastoma; [https://singlecell.broadinstitute.org/single\\_cell/study/SCP393/single-cell-rna-seq-of-adult-and-pediatric-glioblastoma](https://singlecell.broadinstitute.org/single_cell/study/SCP393/single-cell-rna-seq-of-adult-and-pediatric-glioblastoma)) (Nefitel et al., 2019). Raw count data and metadata from a multi-sector single-cell RNAseq dataset (Yu et al., 2020) was obtained from the Gene Expression Omnibus (GSE117891) and processed using the "NormalizeData" function in the R package "Seurat" v3.2.3 (Stuart et al., 2019). Neoplastic cells in this dataset were determined as described previously (Garofano et al., 2021). Raw count data and clinical annotation data from a set of glioma-derived cell populations purified using fluorescence activated cell sorting (FACS) were obtained from the Brain Tumor Immune Micro Environment (BrainTIME) portal and converted to counts per million (CPM) for downstream analysis (<https://joycelab.shinyapps.io/braintime/>) (Klemm et al., 2020).

### Whole-exome and whole-genome analysis

Whole exome and genome alignment, fingerprinting, variant detection, variant post-processing, mutation burden calculation, copy number segmentation, copy number calling, copy number-based purity, ploidy, HLA typing, and neoantigen calling were all performed using previously described pipelines that were developed during the initial GLASS data release (Barthel et al., 2019). Briefly, whole exome and whole genome reads were aligned to the b37 genome (human\_g1k\_v37\_decoy) using Burrows-Wheeler Aligner (BWA) MEM 0.7.17 (Li and Durbin, 2009) and pre-processed according to GATK Best Practices with GATK 4.0.10.1 (McKenna et al., 2010). Fingerprinting on the resulting files was performed using 'CrosscheckFingerprints' to confirm all readgroups from a given sample and all samples from a given patient match, with all mismatches being labelled and dropped from downstream analysis. Somatic mutations were called using GATK4.1 MuTect2. Hypermutation was defined for all recurrent tumors that had more than 10 mutations per megabase sequenced, as described previously (Barthel et al., 2019). Copy number segmentation and calling was performed according to GATK Best Practices as previously described. Copy number-based tumor purity and ploidy were determined using TITAN (Ha et al., 2014). Four-digit HLA class I types were determined from the normal bam files for each sample using OptiType v1.3.2 (Szolek et al., 2014). Neoantigens were called from each patient's somatic mutations and HLA types using pVACseq v4.0.10 (Hundal et al., 2016). Neoantigen depletion was calculated as described previously (Barthel et al., 2019). Loss of heterozygosity (LOH) for each sample's HLA type was called from their respective matched tumor and normal bam files using LOHHLA run with default parameters and a coverage filter of 10 (McGranahan et al., 2017). HLA LOH was called if the estimated copy number for an allele using binning and B-allele frequency was < 0.5 and the *P*-value for allelic imbalance was < 0.05 (paired t-test). Positive selection of focal HLA LOH events was determined using a previously described simulation approach (McGranahan et al., 2017) where a tumor's null probability of deletion LOH for a given genomic region was determined based on the overall proportion of its genome exhibiting a deletion LOH event according to TITAN.

### RNA preprocessing

To ensure each RNAseq file matched to the DNA and RNAseq files from their respective sample and patient, RNAseq fastq files were aligned to the b37 genome using STARv2.7.5 (Dobin et al., 2013) and the resulting bams were then preprocessed using the same



pipelines described for DNA sequencing (Barthel et al., 2019). Fingerprinting was then performed on each bam at the readgroup and patient levels using 'CrosscheckFingerprints.' For each patient-level comparison, each RNA bam was compared to all other RNA and DNA bams coming from the same patient. All mismatches were labelled and dropped from downstream analysis.

RNAseq fastq files were pre-processed with fastp v0.20.0 (Chen et al., 2018). Transcripts per million (TPM) values were then calculated from each sample's pre-processed files using kallisto v0.46.0 (Bray et al., 2016) inputted with an index file built from the Ensemblv75 reference transcriptome. Strand-specific library preparation information was obtained for each sample from the source provider or using STARv2.7.5 quantMode set with the 'GeneCounts' parameter. The resulting TPM values for each sample were combined into a transcript expression matrix for downstream analysis. To create a gene expression matrix, transcript TPM values were collapsed and summed by their respective gene symbols.

### Quality control

To be included in longitudinal downstream analyses, all samples from the same patient had to be collected at least one month apart, as described previously (Barthel et al., 2019). For DNA samples to be included in longitudinal downstream analyses, two samples from a given patient had to pass a previously described quality control process based on fingerprinting, coverage, and copy number variation (Barthel et al., 2019). For RNA samples to be included in longitudinal downstream analyses, two samples from a given patient had to pass a patient-level fingerprinting filter that ensured that the RNA samples matched each other and the patient's respective DNA samples if available. Tumor pairs that had DNA and RNA that passed filtering at each timepoint were used in all downstream analyses that integrated DNA and RNA data.

### Bulk transcriptional subtype classification

Bulk transcriptional subtyping was performed on each GLASS or TCGA sample's processed RNAseq profile using the "ssgsea.GBM.classification" R package (Wang et al., 2017). This method outputs an enrichment score quantifying the representation of each of the three bulk glioma subtypes in a sample as well as a *P*-value indicating the significance of this representation. For analyses that required a single subtype to be assigned to each sample, the subtype with the lowest *P*-value was chosen. In cases where there were ties between subtypes, the subtype with the highest enrichment score was chosen. For analyses that did not require a single subtype designation, all subtypes with *P*-value < 0.05 were assigned to the sample, with "mixed" subtype designations used when all subtypes were equally represented.

### Joint single-cell and bulk RNAseq dataset

Single-cell and bulk RNA sequencing data were generated and processed as previously described (Johnson et al., 2021) and are available for download on Synapse (<https://www.synapse.org/#Synapse:syn2225778>). Briefly, tumor surgical specimens were freshly collected, minced, and partitioned into single-cell and bulk fractions from the same tumor aliquot. The tissues aliquoted for single cell analyses were then mechanically and enzymatically dissociated using the Brain Tumor Dissociation Kit (P) (Miltenyi Cat. No. 130-095-942). FACS was performed to select for viable single cells (Propidium Iodide-, Calcein+ singlets) and enrich for tumor cells by limiting the proportion of non-tumor cells (e.g., immune (CD45+) and endothelial (CD31+) cells). Sorted cells were then loaded on a 10X Chromium chip using the single-cell 3' mRNA kit (10X Genomics). The Cell Ranger pipeline (v3.0.2) was used to convert Illumina base call files to fastq files and align fastqs to hg19 10X reference genome (version 1.2.0) to be compatible with our bulk sequencing data. Data preprocessing and analysis was performed using the Scanpy package (1.3.7) (Wolf et al., 2018) with batch correction performed using BBKNN (Polanski et al., 2020). RNA was extracted for tissues with sufficient tissue and bulk RNAseq libraries were prepared with KAPA mRNA HyperPrep kit (Roche). Bulk RNA sequencing data was processed with the same pipeline as the GLASS samples.

### Deconvolution analyses

Cellular proportions and cell state-specific gene expression matrices were inferred from bulk RNAseq gene expression matrices using CIBERSORTx (Newman et al., 2019). Reference scRNAseq signature matrices were created from our internal 10x-derived scRNAseq dataset (Johnson et al., 2021) and a publicly available SmartSeq2-derived scRNAseq dataset (Nefel et al., 2019) using the 'Create Signature Matrix' module on the CIBERSORTx webserver (<https://cibersortx.stanford.edu/>) with default parameters and quantile normalization disabled. The Ivy GAP signature matrix was downloaded from a prior publication (Puchalski et al., 2018). The CIBERSORTx webserver currently recommends users input no more than 5,000 different single-cell profiles when creating their signature matrix (Steen et al., 2020). To meet this recommendation, our internal scRNAseq dataset, which is made up of 55,284 single cells, was randomly downsampled to 5,000 cells using the 'sample' command in R with the seed set to 11. The cells not included in signature matrix formation were then set aside for validation analyses.

Single-cell-derived cellular proportions and cell state-specific gene expression profiles were inferred from bulk RNAseq datasets using the CIBERSORTx High-Resolution docker container (<https://hub.docker.com/r/cibersortx/hi-res>) following CIBERSORTx instructions. For all runs, the bulk RNAseq dataset was input as the 'mixture' file and the respective signature matrix was input as the 'sigmatrix' file. For runs using our 10x-derived internal scRNAseq signatures, batch correction was done in 'S-mode' by setting the 'rmbatchSmode' parameter to TRUE, while for runs using SmartSeq2-derived scRNAseq signatures batch correction was done in 'B-mode' by setting the 'rmbatchBmode' parameter to TRUE. For each run, the input signature matrix's respective

CIBERSORTx-created “source gene expression profile” was input for batch correction. For all runs, the ‘subsetgenes’ parameter was set to a file containing the intersection of the gene symbols between the run’s respective source gene expression profile and the bulk RNAseq matrix that was being deconvoluted. For the run applying our internal scRNAseq dataset to the bulk GLASS RNAseq matrix, the ‘groundtruth’ parameter was set to a ground truth FACS-purified dataset that was generated as described below. Cellular proportions representing pre-created Ivy GAP signatures were inferred using the ‘Impute Cell Fractions’ module on the CIBERSORTx webserver set to relative mode with quantile normalization and batch correction disabled and 100 permutations for significance analysis.

### Immunofluorescence staining and image acquisition

Tissue samples used in multiplex immunofluorescence microscopy were formalin-fixed, paraffin-embedded and sectioned to a thickness of 5  $\mu\text{m}$  unless otherwise stated. Tissue sections were heated at 58°C for 10 minutes, dewaxed in HistoClear (National Diagnostics) for 20 min and rehydrated in a graded series of alcohol (ethanol:deionized water 100:0, 90:10, 70:30, 50:50, 0:100; 5 min each). Heat-induced epitope retrieval (95°C) was conducted in citrate buffer (pH 6) for 15 min using a BioSB TinoRetriever. After antigen retrieval, tissue sections were permeabilized with PBS 0.1% Triton-X100, washed with PBS and consecutively treated with Fc Receptor Block (Innovex bioscience) for 40 min + Background Buster (Innovex bioscience) for an additional 30 min. The sections were then stained with primary antibodies, diluted in PBS + 5% BSA overnight at 4°C, and then washed and stained with the secondary antibodies at room temperature for 30 minutes. Afterwards, tissues were washed and secondary antibodies were saturated with rabbit normal serum diluted at 1/20 in PBS for 15 minutes at room temperature. Tissues were then stained with directly conjugated antibody mix for 1 hour at room temperature and washed. Nuclei were counterstained with 4',6-diamidino-2-phenylindole (1  $\mu\text{g}/\text{mL}$ ) or SytoxBlue 1/3000 for 2 minutes. Tissues were then mounted in Fluoromount-G mounting media.

Images were acquired on a Leica SP8 confocal microscope equipped with an automated motorized stage. Spectral unmixing was achieved with combination of white light laser tuned laser line for each specific fluorophore, tunable detection window for each marker and sequential acquisition. Whole-slide scans were acquired with a dry 20x objective, while partial slide scans for the panels that included OSM and SNAP25 were acquired with a 40x oil immersion objective. Tiles were stitched and max projected using Leica LAS X software.

### Histo-cytometry

Quantification of single-cell protein expression from immunofluorescence scans was performed using histo-cytometry as previously described (Germer et al., 2012; Wang et al., 2018; Wu et al., 2018). Briefly, each whole slide tissue scan was segmented using Imaris software (version 9.0.2). Using the “spot” function in Imaris, images were segmented using individual cells with a nucleus equal or larger than 5  $\mu\text{m}$  as a seeding point to extend each cells’ surface. The accuracy of the segmentation was manually verified for each sample and adjusted if needed. For each generated spot, x and y coordinates and mean intensity values for all channels were combined and exported into Flowjo v10 to select regions of interest, if needed. Final coordinates and intensity values were exported into a csv file for further analysis in R.

### Validation of cell-state proportions

Cell state proportions derived from our internal scRNAseq dataset were validated using three approaches. In the first approach, synthetic mixtures were made using the single-cell gene expression profiles that had been left out of signature creation. Each synthetic mixture represented the average expression profile of 5,000 single cells where the number of cells of one cell state were manually set and the remaining cells were randomly sampled. Each cell state had its level manually set in 11 mixtures, where it represented 0% of the cells in the first mixture and then increased in 10% increments until reaching 100% in the final mixture. In cases where there were fewer than 5,000 single cells of a given cell state, making 100% representation not possible, the preset proportion instead represented the percent of available cells of that cell state rather than the percent of cells in the mixture. Each synthetic mixture had its true proportions recorded and the resulting mixtures were input into CIBERSORTx for deconvolution. Comparisons of the true and inferred proportions were then performed through correlation analysis.

In the second approach, the cell state proportions inferred from bulk RNAseq data were compared to the cell state proportions quantified by scRNAseq for each sample in our internal scRNAseq dataset. Some samples in the scRNAseq dataset were enriched for CD45<sup>+</sup> cells via FACS and therefore precluded true cell state abundance when considering both neoplastic and non-neoplastic cells. To address this, comparisons were restricted to the relative proportions of each neoplastic cell state. Non-neoplastic cell proportions were removed, and neoplastic cells proportions were then renormalized so that the sum of each neoplastic cell state proportion in each sample added up to 1.

In the third approach, cell state proportions inferred from bulk RNAseq data were compared to the cell state proportions quantified through multiplex immunofluorescence and histo-cytometry analyses performed on whole tissue scans for a subset of samples in the GLASS cohort. To determine the identity of each cell in a tissue scan, expression thresholds were set for each marker based on the marker’s expression distribution across the dataset. For bimodal distributions the threshold was set to the local minima between the two maxima, while for normal distributions the threshold was set to the global maximum. Cells that were negative for all markers



were excluded from further analysis. To facilitate comparisons between expression and immunofluorescence-based estimates, analyses were restricted only to the cell states identified in both platforms, and the resulting fractions were renormalized so that the sum of each proportion added up to 1.

### Annotation and validation of histological features

Digitized images of H&E slides were obtained for a subset of GLASS samples and stored centrally on the Digital Slide Archive (<https://styx.neurology.emory.edu/girder/>). In a subset of samples for which FFPE slides were available for multiplex immunofluorescence staining, representative histological features were digitally outlined by a board-certified neuropathologist.

Transcriptomic histological deconvolution was validated by comparing expression-based and neuropathologist-based estimates of feature abundance. To accomplish this, a team of six neuropathologists were instructed to estimate the proportion of the slide area occupied by different histological features for 10 GLASS samples (5 primary-recurrent tumor pairs) where the H&E slide was directly adjacent to the tumor region sent for RNA-sequencing. Neuropathologists were blinded to the type of glioma in each slide and did not have knowledge of the expression-based scores prior to scoring. To standardize feature evaluation across neuropathologists, common definitions for each feature were established. Definitions for features expected to be observed in both primary and recurrent tumors were loosely based on those used by Ivy GAP, while recurrence-specific features were collaboratively defined by the neuropathologist team. During the evaluation process, each evaluator received a template with these feature definitions and was instructed to annotate the entire slide so that the total estimates for each sample summed to 100% (Table S4). Consensus pathology estimates for each slide were then calculated as the mean neuropathologist estimate of a given feature and were used for all downstream analyses. Results for the necrosis feature samples were additionally reproduced using publicly available neuropathologist estimates from TCGA H&E slides (Cooper et al., 2012).

### Histological feature adjustment

For analyses examining how histological features influenced subtype switching, a tumor sample's cell state composition profile was adjusted to remove cell states that could be attributed to a specific histological feature. To do this, the tumor sample's proportion of a given histological feature was multiplied by the average proportion of each cell state from all samples of that feature in Ivy GAP. These numbers were then subtracted from their respective cell state's proportion in the tumor sample and the resulting profile was then renormalized so that all proportions summed to 1. In cases where the new cell state proportion was less than 0, the value was set to 0 before renormalization.

### Validation of cell state gene expression profiles

Concordance between CIBERSORTx-inferred cell state-specific gene expression profiles and a ground truth set of FACS-purified gene expression profiles was assessed using the 'groundtruth' parameter in CIBERSORTx. The ground truth dataset used in this step was generated from a previously released glioma dataset (Klemm et al., 2020) by collapsing all glioma-derived CD45<sup>+</sup> profiles into an average CD45<sup>+</sup> profile and all glioma-derived macrophage/microglia profiles into an average myeloid cell profile. This dataset was input into CIBERSORTx using the 'groundtruth' parameter during the run applying our internal scRNAseq signature matrix to the GLASS bulk RNAseq dataset. The resulting quality control files output during this run, primarily "SM\_GEPs\_HeatMap.txt", were then used to perform correlation analyses assessing the similarity between the inferred neoplastic cell and myeloid profiles and the ground truth profiles.

### Cell-state gene expression profile analysis

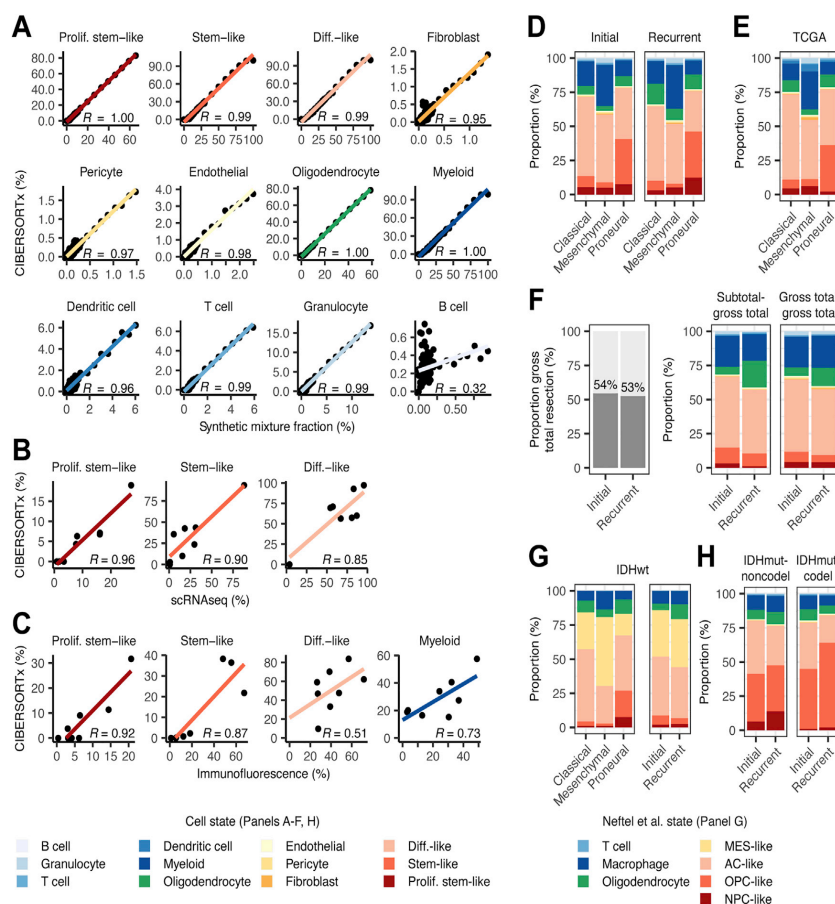
To facilitate downstream analyses on each CIBERSORTx-inferred cell state-specific gene expression profile, each of the resulting expression matrices were log-transformed and all genes that could not be imputed or had a variance of 0 across the dataset were removed. For each cell state-specific gene expression matrix, Wilcoxon signed-rank tests were used to determine the differentially expressed genes between initial and recurrent tumors and the resulting *P*-values were corrected for multiple testing using the Benjamini-Hochberg procedure. Signature scores in cell state-specific gene expression profiles and single-cell RNAseq profiles were defined as the average expression of the genes in the signature. In cases where the expression of some of the genes in the signature could not be determined, the intersection of the signature and the available genes was taken when calculating the signature score. For GO enrichment analyses on signatures derived from cell state-specific gene expression profiles, the background gene set only included the genes CIBERSORTx was able to impute for the cell state from which the signature was derived.

## QUANTIFICATION AND STATISTICAL ANALYSIS

All data analyses were conducted in R v3.6.1 and PostgreSQL 10.6. GO enrichment analyses were performed using the "classic" algorithm in the R package "topGO" v2.38.1. Survival analyses were performed using the R "survival" package. When comparing variables between groups, t-tests were used for cell state proportions while non-parametric tests were used for all other variables (i.e., gene expression, signature score, neoantigen number).



## Supplemental figures



**Figure S1. Validation of deconvolution results, related to Figure 1**

(A) Scatterplots depicting the association between the true proportion and the CIBERSORTx-inferred proportion for each cell state in gene expression profiles from synthetic mixtures composed of different combinations of single cells.

(B) Scatterplots depicting the association between the proportion of each neoplastic cell state determined from single-cell RNA-seq and the non-neoplastic cell-adjusted neoplastic cell-state proportion inferred from CIBERSORTx applied to each sample's respective bulk tumor RNA-seq profile.

(C) Scatterplots depicting the associations between the relative proportions of proliferating stem-like neoplastic cells (SOX2+ and Ki67+), stem-like neoplastic cells (SOX2+, OLIG2+, and Ki67-), differentiated-like neoplastic cells (SOX2-, OLIG2-, Ki67-, EGFR+ or CD44+), and myeloid cells (CD14+ and SOX2-) inferred using multiplex immunofluorescence and the corresponding cell-state proportions inferred from expression data using CIBERSORTx. The myeloid cell CIBERSORTx fraction represents the sum of the myeloid, granulocyte, and dendritic cell fractions. In all plots, Pearson correlation coefficients are indicated.

(D) The average cell-state composition of each bulk transcriptional subtype across initial and recurrent GLASS samples.

(E) The average cell-state composition of each bulk transcriptional subtype across all TCGA samples.

(F) Left: stacked bar plot indicating the proportion of IDHWT tumors that underwent a gross total resection at each time point. Right: the average proportions of each cell state for tumors that underwent a subtotal resection at initial time point and a gross total resection at recurrence (subtotal-gross total) and tumors that underwent a gross total resection at both time points (gross total-gross total).

(G) Left: the average Nefitel et al. cell-state composition of each bulk transcriptional subtype for all IDHWT GLASS tumors. Right: the average Nefitel et al. cell-state composition of initial and recurrent IDHWT tumors.

(H) The average cell-state composition of initial and recurrent IDHmut tumors stratified by 1p/19q co-deletion status. Colors for all panels are indicated at the bottom of the figure.

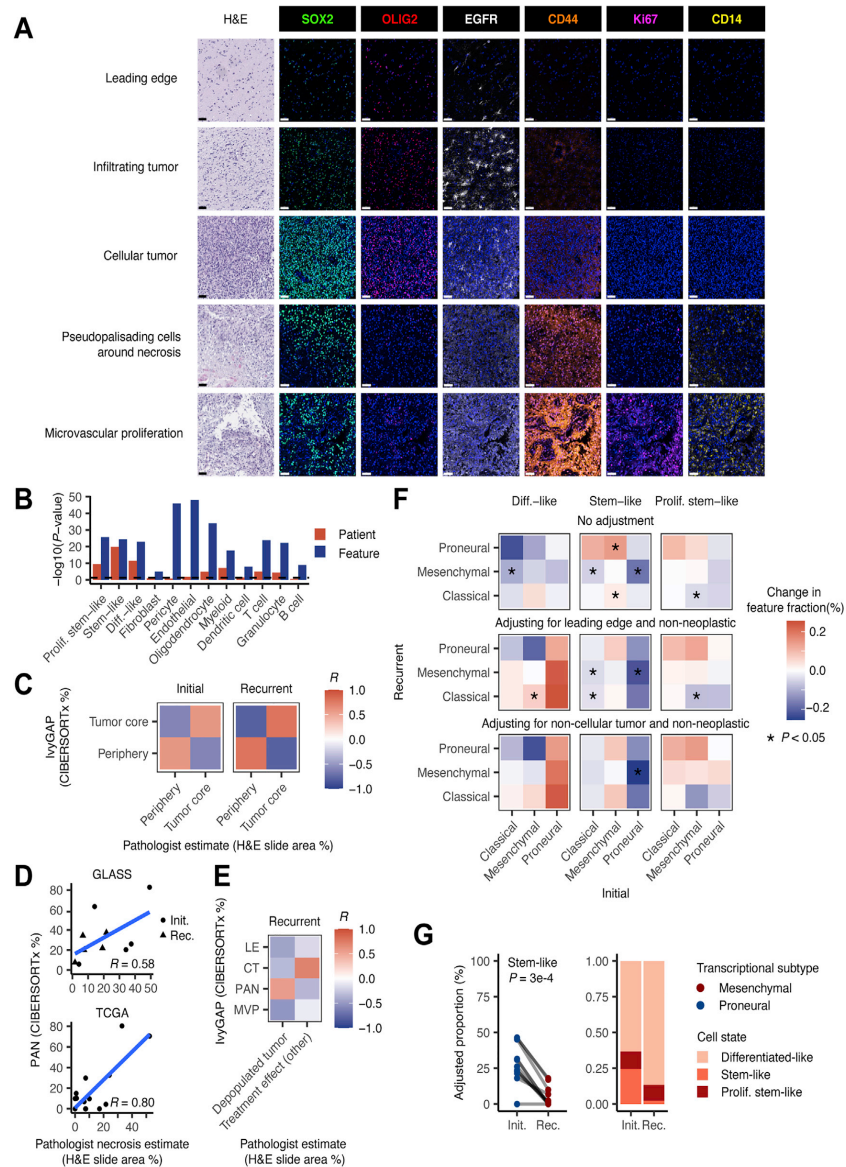
See also Tables S1, S2, and S3.

# Mutational signatures of genotoxic cancer therapies



CellPress

Cell  
Resource



**Figure S2. Relationship between cell state and histological feature composition, related to Figure 2**

(A) Representative H&E and multiplex immunofluorescence images for each IvYGAP histological feature. Features were identified by a neuropathologist based on the H&E images on the left. The leading edge, infiltrating tumor, and cellular tumor features are from GLSS-LU-0B10 (primary), whereas the pseudopalisading cells around necrosis and microvascular proliferation features are from GLSS-LU-0B9 (primary). Scale bars represent 50  $\mu$ m.

(B) Bar plot depicting the  $-\log_{10}$  p value from a two-way ANOVA measuring whether the fractions of each cell state in a sample associate with the patient the sample was derived from (red bar) and the feature the sample represents (blue bar). The dashed line corresponds to  $p = 0.05$ .

(C) Heatmap depicting the Pearson correlation coefficients measuring the association between pathologist and CIBERSORTx estimates of nucleated tumor core- and periphery-related histological features. Evaluations were performed across 5 initial and 5 recurrent samples.

(legend continued on next page)



# Mutational signatures of genotoxic cancer therapies



---

(D) Scatterplots depicting the association between pathologist estimates of necrosis and CIBERSORTx estimates of the IvyGAP pseudopalisading cells around necrosis feature in the GLASS and TCGA datasets. Shapes indicate initial and recurrence status.

(E) Heatmap depicting the Pearson correlation coefficients measuring the association between pathologist estimates of recurrence-specific nucleated histological features and CIBERSORTx estimates of IvyGAP features. Evaluations were performed across 5 recurrent samples. In (C)–(E), pathologist estimates were based on the relative percent of the H&E slide area occupied by a given feature while CIBERSORTx estimates were based on RNA-seq. Abbreviations: leading edge (LE), infiltrating tumor (IT), cellular tumor (CT), pseudopalisading cells around necrosis (PAN), and microvascular proliferation (MVP).

(F) Heatmap depicting the changes in each neoplastic cell state between initial and recurrent tumors undergoing the indicated subtype transition. The initial subtype is indicated in the columns, and the recurrent subtype is indicated in the rows. Each row of heatmaps reflects a different histological feature adjustment. Colors represent the change in the fraction of the indicated features between initial and recurrent tumors, while \* indicates a paired t test p value < 0.05.

(G) Left: ladder plot depicting the change in the adjusted stem-like cell proportion between paired initial and recurrent tumors undergoing a proneural-to-mesenchymal transition. Right: the average adjusted proportions for neoplastic cells for the tumor pairs outlined on the left. Neoplastic cell proportions were adjusted for the presence of non-neoplastic cells and leading-edge content. p value from paired t test.

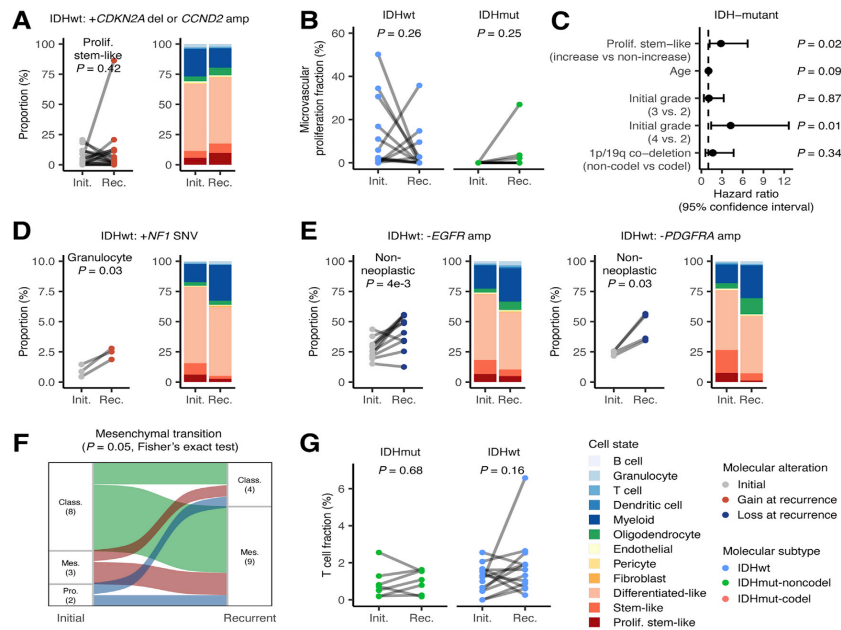
See also [Table S4](#).

# Mutational signatures of genotoxic cancer therapies



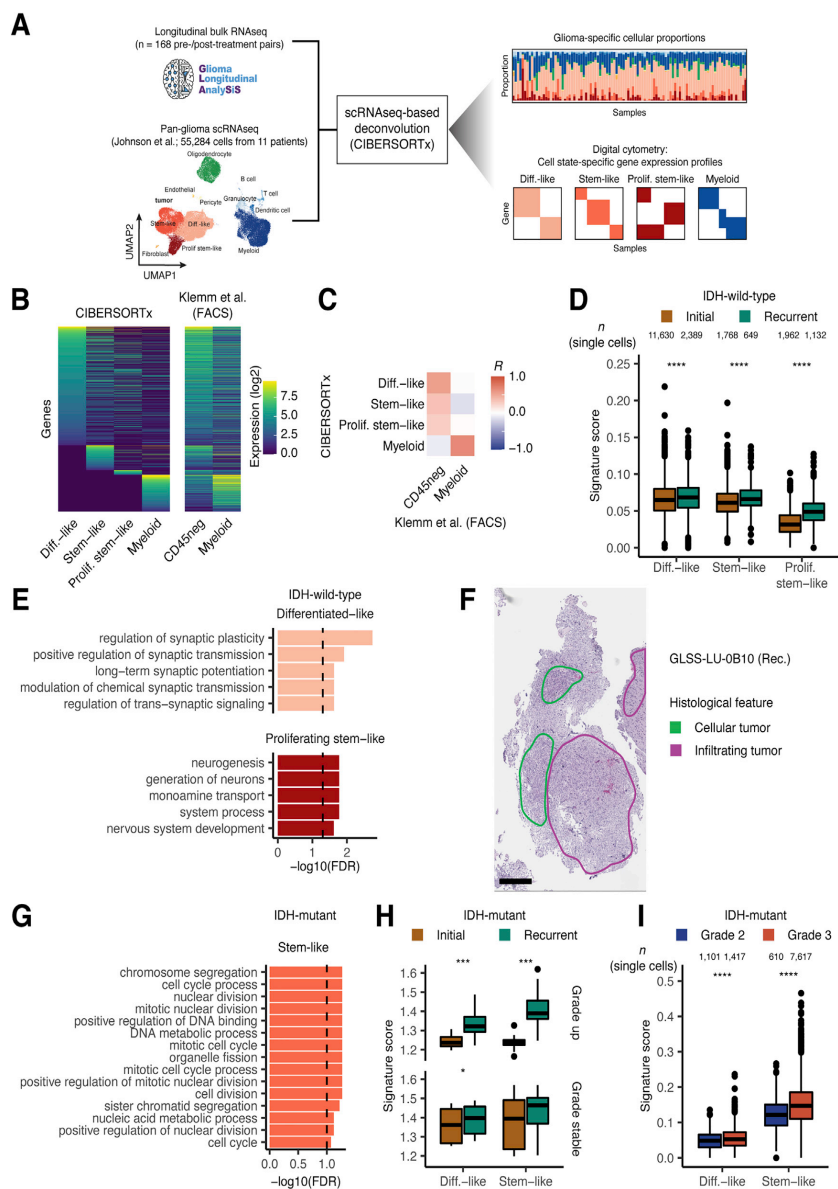
CellPress

Cell  
Resource



**Figure S3. Cell-state changes in samples that have acquired or lost somatic alterations, related to Figure 3**  
 (A) Left: ladder plot depicting the change in the proliferating stem-like cell proportion between paired initial and recurrent IDHWT tumors that acquired *CDKN2A* deletions or *CCND2* amplifications. Right: stacked bar plot depicting the average proportions of each cell state for the tumor pairs in the ladder plots.  
 (B) Ladder plots depicting the difference in microvascular proliferation fraction in IDHWT tumors that acquired hypermutation and IDHmut tumors that acquired a cell-cycle alteration or hypermutation at recurrence.  
 (C) Forest plot depicting the results of a multivariable Cox proportional hazards model that included proliferating stem-like cell increase, age, initial grade, and 1p/19q co-deletion status as variables. Points represent the hazard ratio, and lines represent the 95% confidence interval.  $p$  values were calculated using the Wald test.  
 (D) Left: ladder plots depicting the change in granulocyte proportion in IDHWT tumors that acquired mutations in *NF1* at recurrence. Right: the average proportions of each cell state for the tumor pairs in the ladder plots.  
 (E) Non-neoplastic cell-state differences in IDHWT tumors that lost *EGFR* or *PDGFRA* amplifications at recurrence. (E) is split by alteration. Ladder plots depict the change in the non-neoplastic cell-state proportion between paired initial and recurrent tumors, whereas the stacked bar plots depict the average proportions of each cell state for these tumors.  
 (F) Sankey plot indicating whether the highest-scoring transcriptional subtype changed at recurrence for the tumors depicted in (E). Each color reflects the transcriptional subtype in the initial tumors. Numbers in parentheses indicate the number of samples. Subtype abbreviations: proneural (Pro.), classical (Class.), and mesenchymal (Mes.).  
 (G) Ladder plots depicting the difference in T cell fraction in tumors that underwent hypermutation at recurrence.  
 In all figures,  $p$  values were calculated using a paired  $t$  test unless otherwise noted.

# Mutational signatures of genotoxic cancer therapies



**Figure S4. Validation and analysis of cell-state-specific gene expression profiles, related to Figure 4**

(A) Schema for single-cell RNA-seq-based deconvolution of cell-state-specific gene expression profiles using the pan-glioma single-cell RNA-seq dataset (Johnson et al., 2021).

(B) Heatmap depicting the relationship between the CIBERSORTx-inferred gene expression profiles and gene expression profiles from analogous cell types from a FACS-purified ground-truth dataset (Klemm et al., 2020). In the CD45neg column in the Klemm et al. heatmap, which represents a composite gene expression profile from the non-immune cells purified from a collection of glioma tumors, gene expression patterns from all three neoplastic cell states can be observed.

(C) Heatmap depicting the correlation coefficients between each CIBERSORTx-inferred cell-state-specific gene expression profile and the gene expression profiles from the FACS-purified ground-truth dataset.

(legend continued on next page)

# Mutational signatures of genotoxic cancer therapies



CellPress

Cell  
Resource

---

(D) Boxplot depicting the average signature expression in single cells of the indicated neoplastic cell states from unmatched initial and recurrent IDHWT tumors.

(E) Bar plot depicting the  $-\log_{10}(\text{FDR})$  from a GO enrichment analysis of the genes significantly up-regulated at recurrence in the differentiated-like and proliferating stem-like neoplastic cell-specific gene expression profiles from IDHWT tumors. The top 5 GO terms for each cell state are included. The dashed line corresponds to  $\text{FDR} < 0.05$ .

(F) H&E image is used to define the histological features used for multiplex immunofluorescence staining in [Figure S4E](#). Neuropathologist annotations of cellular tumor and infiltrating tumor features are highlighted in the indicated colors. The scale bar represents 500  $\mu\text{m}$ .

(G) Bar plot depicting the  $-\log_{10}(\text{FDR})$  from a GO enrichment analysis of the genes significantly up-regulated at recurrence in the differentiated-like neoplastic cell-specific gene expression profiles from IDHmut tumors. The dashed line corresponds to  $\text{FDR} < 0.1$ .

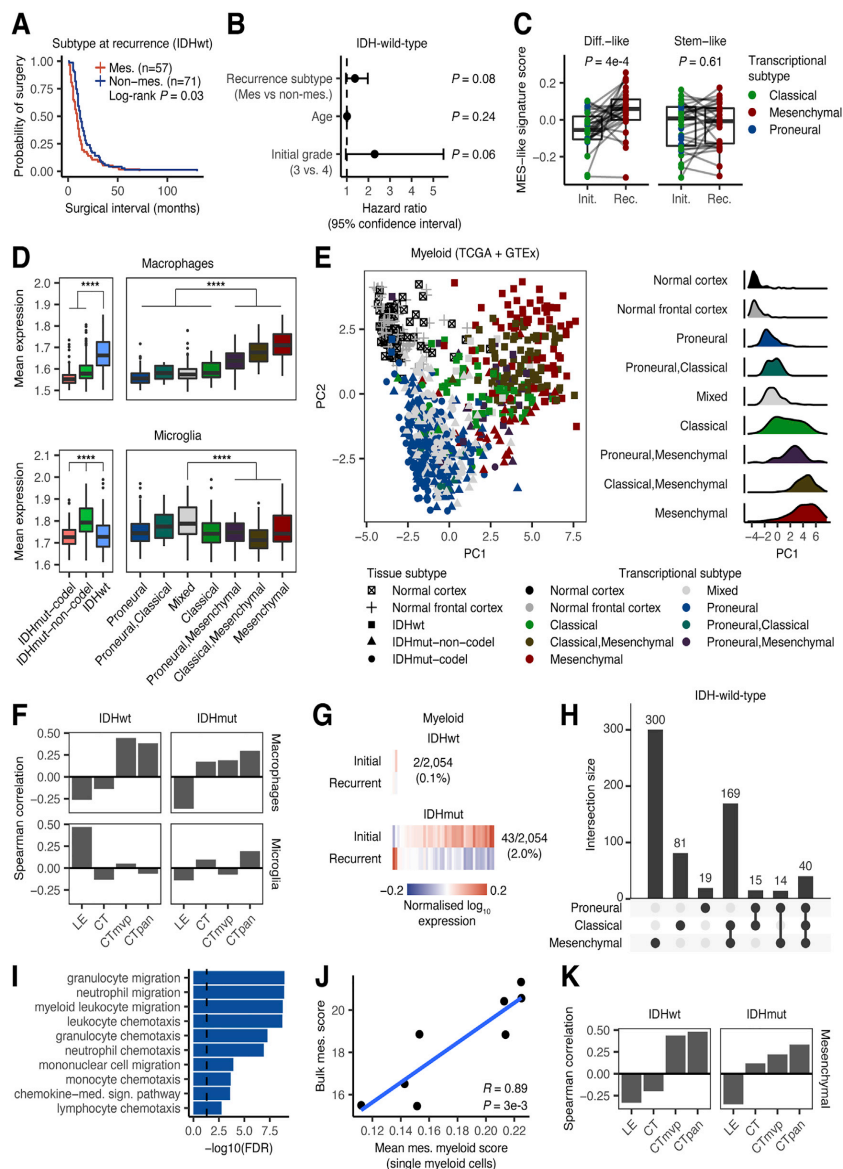
(H) Boxplot depicting the average signature expression in neoplastic cell-state-specific gene expression profiles for each IDHmut tumor pair in GLASS. Comparisons are stratified based on whether the tumor pair was grade stable or exhibited a grade increase at recurrence.

(I) Boxplot depicting the average signature expression in single cells of the indicated neoplastic cell states from grade 2 and grade 3. In (D) and (I), single cells were derived from the joint single-cell and bulk RNA-seq dataset ([Johnson et al., 2021](#)).

Across all panels, \*\*\*\* indicates  $p < 1e-5$ , \*\*\* indicates  $p < 1e-3$  and \* indicates  $p < 0.05$ . p values in [Figures S4D](#) and [S4H](#) were calculated using the Wilcoxon signed-rank test, while p values in [Figure S4I](#) were calculated using the Wilcoxon rank-sum test.

See also [Table S5](#).

# Mutational signatures of genotoxic cancer therapies



**Figure S5. Characterization of the mesenchymal myeloid signature, related to Figure 5**

(A) Kaplan-Meier plot depicting the surgical interval distributions of patients with tumors that were and were not mesenchymal at recurrence. p value was calculated using the log-rank test.  
 (B) Forest plot depicting the results of a multivariable Cox proportional hazards model that included recurrent tumor subtype, age, and initial grade as variables. Points represent the hazard ratio, and lines represent the 95% confidence interval. p values were calculated using the Wald test.  
 (C) Box and ladder plots depicting the difference in the median-normalized mean expression of the Nefel et al. MES-like signature between initial and recurrent IDHWT tumors from GLASS undergoing a mesenchymal transition. Point colors indicate transcriptional subtype. p values were calculated using the Wilcoxon signed-rank test.

(legend continued on next page)

# Mutational signatures of genotoxic cancer therapies



CellPress

Cell  
Resource

(D) Boxplots depicting the average macrophage and microglia gene expression signatures in CIBERSORTx-inferred myeloid-specific gene expression profiles from TCGA. Samples are stratified by IDH and 1p/19q co-deletion status (left) and bulk transcriptional subtype (right). \*\*\*\* indicates Wilcoxon rank-sum test p value < 1e-5.

(E) Left: principal component analysis plot of the CIBERSORTx-inferred myeloid profiles from TCGA and GTEx. Colors indicate bulk transcriptional subtype; shapes indicate tissue subtype. Right: density plot depicting the distribution of principal component 1 (PC1) among each transcriptional subtype.

(F) Bar plots depicting the Spearman correlation coefficients measuring the association between the myeloid-specific expression scores for the macrophage and microglia signatures versus the presence of the four Ivy GAP histological features in TCGA. The features measured were leading edge (LE), cellular tumor (CT), pseudopalisading cells around necrosis (CTpan), and microvascular proliferation (CTmvp).

(G) Heatmaps depicting the average normalized  $\log_{10}$  expression level of genes that were differentially expressed between myeloid cell states from initial and recurrent IDHWT and IDHmut tumors in GLASS that did not undergo a subtype switch. Fractions indicate the number of differentially expressed genes out of the number of genes inferred for that cell state's profile.

(H) Upset plot depicting the intersection of significantly up-regulated genes in the myeloid-specific gene expression profiles from each transcriptional subtype relative to the normal brain cortex. Intersections between signatures are shown in the combination matrix. The number of genes uniquely found in each set is indicated above each bar.

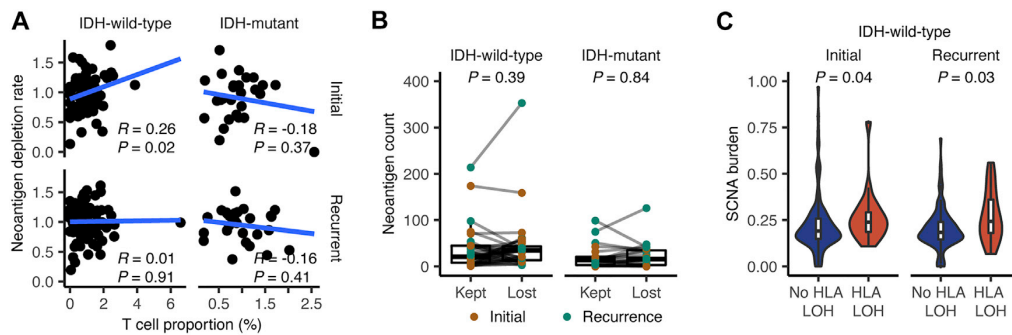
(I) Bar plot depicting the  $-\log_{10}(\text{FDR})$  from a GO enrichment analysis for the genes in the mesenchymal myeloid signature.

(J) Scatterplot depicting the association between the mean mesenchymal myeloid signature expression in single myeloid cells and the mesenchymal subtype score calculated from bulk RNA-seq for each patient.

(K) Bar plots depicting the Spearman correlation coefficients measuring the association between the myeloid-specific expression scores for the mesenchymal myeloid signature versus the presence of the four Ivy GAP histological features in TCGA, as in (F).

See also [Tables S6](#) and [S7](#).

# Mutational signatures of genotoxic cancer therapies



**Figure S6. Analysis of neoantigen-mediated T cell selection in glioma, related to Figure 6**

(A) Scatterplots depicting the association between the T cell proportion and the neoantigen depletion rate in initial and recurrent GLASS samples.

(B) Box and ladder plots depicting the difference in the number of neoantigens binding to the kept and lost allele. Points are colored based on whether the sample was an initial or recurrent tumor. p values were calculated using the Wilcoxon signed-rank test.

(C) Violin plots depicting the distribution of the somatic copy-number alteration burden in initial and recurrent IDHWT GLASS samples that did and did not exhibit HLA LOH. p values were calculated using the Wilcoxon rank-sum test.

## Chapter 6

# Comparative molecular life history of spontaneous canine and human gliomas

This chapter is based on the following publication<sup>65</sup>:

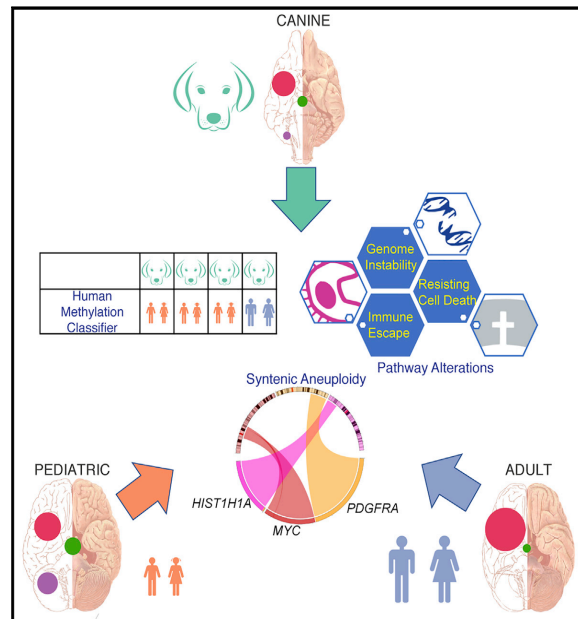
Amin, S.A., Anderson, K.J., Boudreau, C.E., Martinez-Ledesma, E., **Kocakavuk, E.**, *et al.* Comparative Molecular Life History of Spontaneous Canine and Human Gliomas. *Cancer Cell* **37**, 243-257.e7 (2020).  
<https://doi.org/10.1016/j.ccell.2020.01.004>



# Cancer Cell

## Comparative Molecular Life History of Spontaneous Canine and Human Gliomas

### Graphical Abstract



### Authors

Samirkumar B. Amin,  
Kevin J. Anderson,  
C. Elizabeth Boudreau, ...,  
Amy B. Heimberger,  
Jonathan M. Levine, Roel G.W. Verhaak

### Correspondence

roel.verhaak@jax.org

### In Brief

Amin et al. characterize the molecular landscape of canine gliomas and compare it with that of human pediatric and adult gliomas, revealing high similarity between human pediatric and canine gliomas. The cross-species analysis identifies conserved glioma drivers and aneuploidy as a hallmark of high-grade disease.

### Highlights

- Genomic, epigenomic, and transcriptomic characterization of sporadic glioma in dogs
- Somatic alterations in canine glioma converge with human glioma drivers
- Canine glioma resemble pediatric human glioma by mutation rate and DNA methylation
- Microenvironment similarity between canine and human pediatric and adult glioma



Amin et al., 2020, *Cancer Cell* 37, 243–257  
February 10, 2020 © 2020 Elsevier Inc.  
<https://doi.org/10.1016/j.ccell.2020.01.004>

CellPress

## Comparative Molecular Life History of Spontaneous Canine and Human Gliomas

Samirkumar B. Amin,<sup>1</sup> Kevin J. Anderson,<sup>1,23</sup> C. Elizabeth Boudreau,<sup>2,23</sup> Emmanuel Martinez-Ledesma,<sup>3,4,23</sup> Emre Kocakavuk,<sup>1,5</sup> Kevin C. Johnson,<sup>1</sup> Floris P. Barthel,<sup>1</sup> Frederick S. Varn,<sup>1</sup> Cynthia Kassab,<sup>6</sup> Xiaoyang Ling,<sup>6</sup> Hoon Kim,<sup>1</sup> Mary Barter,<sup>7</sup> Ching C. Lau,<sup>1,8,9</sup> Chew Yee Ngan,<sup>1</sup> Margaret Chapman,<sup>1</sup> Jennifer W. Koehler,<sup>10</sup> James P. Long,<sup>6,11</sup> Andrew D. Miller,<sup>12</sup> C. Ryan Miller,<sup>13,22</sup> Brian F. Porter,<sup>14</sup> Daniel R. Rissi,<sup>15</sup>

(Author list continued on next page)

<sup>1</sup>The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA

<sup>2</sup>Department of Small Animal Clinical Sciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843, USA

<sup>3</sup>Tecnologico de Monterrey, Escuela de Medicina y Ciencias de la Salud, Avenue Morones Prieto 3000, Monterrey, Nuevo Leon 64710, Mexico

<sup>4</sup>Department of Neuro-Oncology, the University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

<sup>5</sup>DKFZ Division of Translational Neurooncology at the West German Cancer Center (WTZ), German Cancer Consortium (DKTK) Partner Site & Department of Neurosurgery, University Hospital Essen, Essen, Germany

<sup>6</sup>Department of Neurosurgery, the University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

<sup>7</sup>The Jackson Laboratory, Bar Harbor, ME 04609, USA

<sup>8</sup>Connecticut Children's Medical Center, Hartford, CT 06106, USA

<sup>9</sup>University of Connecticut School of Medicine, Farmington, CT 06032, USA

<sup>10</sup>Department of Pathobiology, College of Veterinary Medicine, Auburn University, Auburn, AL, USA

<sup>11</sup>Department of Biostatistics, the University of Texas MD Anderson Cancer Center, Houston, TX, USA

<sup>12</sup>Department of Biomedical Sciences, Section of Anatomic Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA

<sup>13</sup>Departments of Pathology and Laboratory Medicine, Neurology, and Pharmacology, Lineberger Comprehensive Cancer Center and Neuroscience Center, University of North Carolina School of Medicine, Chapel Hill, NC, USA

(Affiliations continued on next page)

### SUMMARY

Sporadic gliomas in companion dogs provide a window on the interaction between tumorigenic mechanisms and host environment. We compared the molecular profiles of canine gliomas with those of human pediatric and adult gliomas to characterize evolutionarily conserved mammalian mutational processes in gliomagenesis. Employing whole-genome, exome, transcriptome, and methylation sequencing of 83 canine gliomas, we found alterations shared between canine and human gliomas such as the receptor tyrosine kinases, *TP53* and cell-cycle pathways, and *IDH1* R132. Canine gliomas showed high similarity with human pediatric gliomas per robust aneuploidy, mutational rates, relative timing of mutations, and DNA-methylation patterns. Our cross-species comparative genomic analysis provides unique insights into glioma etiology and the chronology of glioma-causing somatic alterations.

### INTRODUCTION

The natural history of cancer is marked by temporal acquisition of diverse genetic and epigenetic aberrations. The inevi-

table intratumoral and interpatient heterogeneity among evolving cancer cells poses a major obstacle in our understanding of cancer evolution and designing effective treatment strategies (Alizadeh et al., 2015). Recent developments in

#### Significance

Diffuse gliomas are the most common malignant brain tumors, with high-grade tumors carrying a dismal prognosis. Preclinical models have proven themselves as poor predictors of clinical efficacy. Spontaneous glioma in dogs provides an attractive alternative model because of their comparable tumor microenvironment and tumor life history. We determined the similarities and differences between human and canine gliomas through genomic profiling, and leveraged our datasets to identify conserved somatic drivers, mutational processes, and temporal ordering of somatic glioma events across species. Canine gliomas resemble human gliomas at (epi-)genetic levels and are more reminiscent of pediatric than adult disease, thus rationalizing sporadic canine glioma as a preclinical model tailored to measuring treatment efficacies in patients with canine or human glioma.



Christina Mazcko,<sup>16</sup> Amy K. LeBlanc,<sup>16</sup> Peter J. Dickinson,<sup>17</sup> Rebecca A. Packer,<sup>18</sup> Amanda R. Taylor,<sup>19,21</sup> John H. Rossmel Jr.,<sup>20</sup> Kevin D. Woolard,<sup>17</sup> Amy B. Heimberger,<sup>6,24</sup> Jonathan M. Levine,<sup>2,24</sup> and Roel G.W. Verhaak<sup>1,24,25,\*</sup>

<sup>14</sup>Department of Veterinary Pathobiology, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX, USA

<sup>15</sup>Department of Pathology and Athens Veterinary Diagnostic Laboratory, College of Veterinary Medicine, University of Georgia, Athens, GA, USA

<sup>16</sup>Comparative Oncology Program, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

<sup>17</sup>Department of Surgical and Radiological Sciences, UC Davis School of Veterinary Medicine, Davis, CA, USA

<sup>18</sup>Department of Clinical Sciences, College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO, USA

<sup>19</sup>Auburn University College of Veterinary Medicine, Auburn, AL, USA

<sup>20</sup>VA-MD College of Veterinary Medicine, Blacksburg, VA, USA

<sup>21</sup>Present address: MedVet Medical and Cancer Center for Pets, Columbus, OH, USA

<sup>22</sup>Present address: Department of Pathology, Division of Neuropathology, and O'Neal Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, AL, USA

<sup>23</sup>These authors contributed equally

<sup>24</sup>Co-senior authors

<sup>25</sup>Lead Contact

\*Correspondence: [roel.verhaak@jax.org](mailto:roel.verhaak@jax.org)

<https://doi.org/10.1016/j.ccell.2020.01.004>

high-throughput lineage tracing, organoid cultures, and patient-derived xenografts have provided better resolution of heterogeneity and driver events. Nonetheless, in the absence of natural host response, preclinical *in vitro* and rodent models are unable to fully recapitulate a spontaneously evolving tumor's life history. This limitation challenges the accuracy of predicting therapeutic responses in these preclinical models, especially response to immunotherapies (Buque and Galluzzi, 2018).

Somatic evolution of cancers may follow convergent patterns across mammalian species by selecting cells that carry beneficial mutations in highly conserved regions, i.e., genes and their regulatory non-coding regions enabling one or more cancer hallmarks (Hanahan and Weinberg, 2011). Unlike induced cancer models, comparative genomics of spontaneous tumors across species provides a unique advantage to identify defects in such shared, evolutionarily constrained regions (Lindblad-Toh et al., 2011) and to evaluate the importance of host context in the tumor's evolution. In addition to their natural tumorigenesis, spontaneous cancers in dogs are marked by the presence of a fully functional tumor micro-environment (Khanna et al., 2006; LeBlanc et al., 2016). Cancer cells are subject to clonal selection and drift, and the resulting tumor is molded by selection pressure from the tissue context (DeGregori, 2017; Fortunato et al., 2017). This Darwinian adaptation may select for somatic alterations in evolutionarily conserved regions in both dogs and humans that are relevant to tumorigenesis.

Sporadic gliomas occur in companion dogs at frequencies similar to those in humans (Snyder et al., 2006; Song et al., 2013). Genomic characterization of canine glioma has a distinct merit, in that dogs are diagnosed in the adult stage of life but with an age distribution that is comparable with human pediatric disease. This seeming conundrum in fact creates an opportunity to compare somatic drivers and their relative timing in canine glioma with those in human glioma. Studies involving comparative genomics of spontaneous canine cancers have already enabled identification of breed-specific, disease-risk loci under strong evolutionary constraints and with known roles in human cancer, e.g., germline

*FGF4* retrogene expression in chondrodysplasia (Parker et al., 2009), somatic *BRAF* V600E mutation in canine invasive transitional cell carcinoma of the bladder (Decker et al., 2015b), recurrent somatic *SETD2* mutations in canine osteosarcoma (Sakthikumar et al., 2018), and *TP53* pathway alterations in canine melanoma (Hendricks et al., 2018; Wong et al., 2019). Earlier studies in canine gliomas have characterized somatic copy-number alterations syntenic with those in human adult gliomas (Dickinson et al., 2016) and have identified genetic susceptibility factors near genes such as *CAMKK2*, *P2RX7*, and *DVL2* (Mansour et al., 2018; Truve et al., 2016).

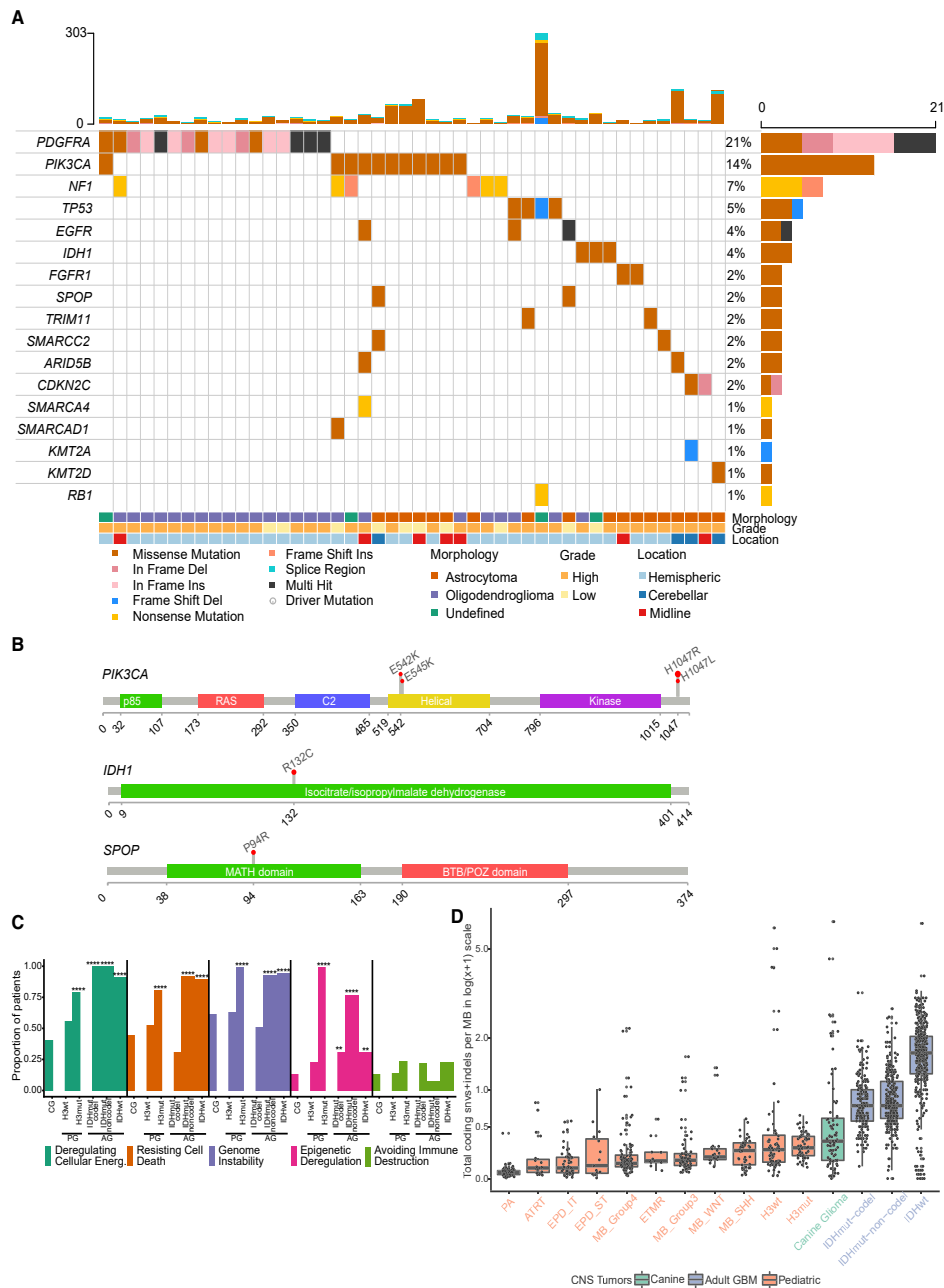
Here, we have performed comparative genomic, transcriptomic, and epigenetic profiling across three population structures, canine glioma, human adult glioma (Ceccarelli et al., 2016), and human pediatric glioma (Gröbner et al., 2018; Ma et al., 2018), to study somatic evolutionary traits of glioma across two species and in different age groups. We leveraged genomic profiles to infer molecular life history in order to understand cross-species convergent evolution of glioma (Aktipis et al., 2013; Stearns, 1992).

## RESULTS

### Human Glioma Driver Events Are Frequently Found in Canine Disease

We performed whole-genome, exome, transcriptome, and methylation sequencing (373 libraries) on canine gliomas ( $n = 83$ ) and germline ( $n = 67$ ) samples from 83 dogs (NCBI SRA accession: PRJNA579792), with all samples obtained via necropsy. Using the recently updated criteria for diagnostic histopathological classification (Koehler et al., 2018), 46 cases were classified as oligodendroglioma, 31 cases as astrocytoma, and 6 cases as undefined glioma (Table S1). We defined a common set of 81 cases for which whole-genome and exome data were available with minimum of 30× coverage in exome regions (Table S1 and Figure S1A; STAR Methods). From mutation calls derived from all 81 cases, we detected somatic mutational driver events using dNdS (Martincorena et al., 2017), MuSiC2 (Dees et al., 2012), and a semi-supervised comparison with known

# Mutational signatures of genotoxic cancer therapies



**Figure 1. Comparative Somatic Landscape of Canine and Human Gliomas**

(A) Somatic variants in canine gliomas. Top bar plot shows patient-wise frequency of somatic variants ( $n = 46$  of 81 canine patients) and right-side bar plot shows gene-wise frequency of somatic variant types. Bottom annotations show relevant patient-specific annotations.

(legend continued on next page)



cancer drivers in human adult and human pediatric cancers (Bailey et al., 2018; Gröbner et al., 2018; Ma et al., 2018) (Figure 1A and Table S2; STAR Methods). We detected mutations in genes associated with human pediatric (Mackay et al., 2017) and adult glioma (Brennan et al., 2013; Ceccarelli et al., 2016) such as the *TP53*, *PDGFRA*, *PIK3CA*, and *EGFR* (Figure S1B), as well as recurrent hotspot and mutually exclusive mutations with high oncogenic impact according to the Catalog of Somatic Mutations in Cancer (COSMIC) database (Tate et al., 2019) in *PIK3CA* H1047R/L (n = 8), *PDGFRA* K385I/M (n = 6), *IDH1* R132C (n = 3), and *SPOP* P94R (n = 1; 1 shared with *PIK3CA* H1047R) (Figure 1B and Table S3). These mutations were also identified as being under positive selection or as significantly mutated genes using the dNdS (Martincorena et al., 2017) approach (Table S2) and thus indicating driver mutations of canine gliomas. Mutations affecting the *IDH1* R132 codon are a defining characteristic of low-grade adult gliomas (Cancer Genome Atlas Research Network et al., 2015) and were detected infrequently in pediatric and canine gliomas (n = 3/81). Overall, 36/81 (44%) of canine gliomas carried at least one significantly mutated gene. This proportion was comparable with published findings in human pediatric gliomas (114/217, 52%, chi-square p value 0.54) (Gröbner et al., 2018) but contrasted with the frequency at which adult gliomas contain at least one significantly mutated gene alteration (753/812, 93%, chi-square p value 0.0004). To demonstrate similarity between canine gliomas and human gliomas, we summarized levels of somatic coding mutations, high-level copy amplifications and deep deletions in gene sets reflecting previously reported cancer hallmarks (Table S4). We tallied weighted pathway contributions per cohort (canine, adult, pediatric) by the number of coding mutations within each cohort and genes per pathway. Adult glioma is commonly separated into subtypes on the basis of IDH mutation as well as chromosome arm 1p and 19q deletion, resulting in three subtypes: (1) IDH wild type; (2) IDH mutant with codeletion (IDHmut-codel); and (3) IDH mutant without codeletion (IDHmut-noncodel) (Louis et al., 2016). Pediatric high-grade gliomas are separated based on histone H3 mutation status into two subtypes: histone H3 gene mutant (H3 mutant) versus wild type (H3 wild type) (Louis et al., 2016). We did not include low-grade pediatric gliomas in our comparison due to the paucity of somatic alterations in these glioma types (Jones et al., 2013; Pollack et al., 2019; Zhang et al., 2013). We found that canine gliomas were most similar to pediatric H3 wild-type gliomas at the pathway alteration level, i.e., comparable hallmark enrichment with no significant difference between groups. Pediatric H3 mutant, adult IDH wild type, and

IDHmut-noncodel gliomas showed increased frequency of gene mutations in cancer hallmarks such as deregulating cellular energetics, genomic instability, and resisting cell death (Figures 1C and S1C). Among all 11 cancer hallmarks tested, “avoiding immune destruction” scored low across both canine and human gliomas (Table S4), potentially owing to the immune-cold nature of gliomas (Boussiotis and Charest, 2018; Brown et al., 2018).

We compared mutation burden between canine and a variety of human pediatric and adult cohorts using coding mutation rates from 4,761 human patients (Bailey et al., 2018; Ceccarelli et al., 2016; Gröbner et al., 2018; Ma et al., 2018) (STAR Methods). The somatic mutation rate of canine glioma (0.34 coding mutations per megabase; 95% confidence interval [CI]: 0.15–0.6) was similar to that of human pediatric gliomas (Figures 1D and S1D). High-grade canine gliomas (n = 63/81) had mutation rates comparable with those of pediatric H3-mutant and H3 wild-type subtypes (0.34, 0.27, 0.25 coding mutations per megabase, respectively; Wilcoxon p value 0.18 and 0.1; Figure S1E), but significantly lower than in human adult IDH-mutant and IDH wild-type gliomas (0.77 and 1.67 coding mutations per megabase, respectively; Wilcoxon p values of  $8 \times 10^{-9}$  or less). Low mutation burden has been linked to fewer mutations in cancer-driving genes (Martincorena et al., 2017) and may explain the relative paucity of significantly mutated genes observed in canine gliomas, including weaker positive selection ( $q > 0.1$ ) for known and mutated cancer genes (n = 50; Figure S1F). These results demonstrate that the landscape of somatic single-nucleotide variants is similar to that of human glioma, and suggests that canine glioma aligns more closely with human pediatric glioma than with adult disease.

### Aneuploidy Is a Major Driver of Canine and Pediatric High-Grade Glioma

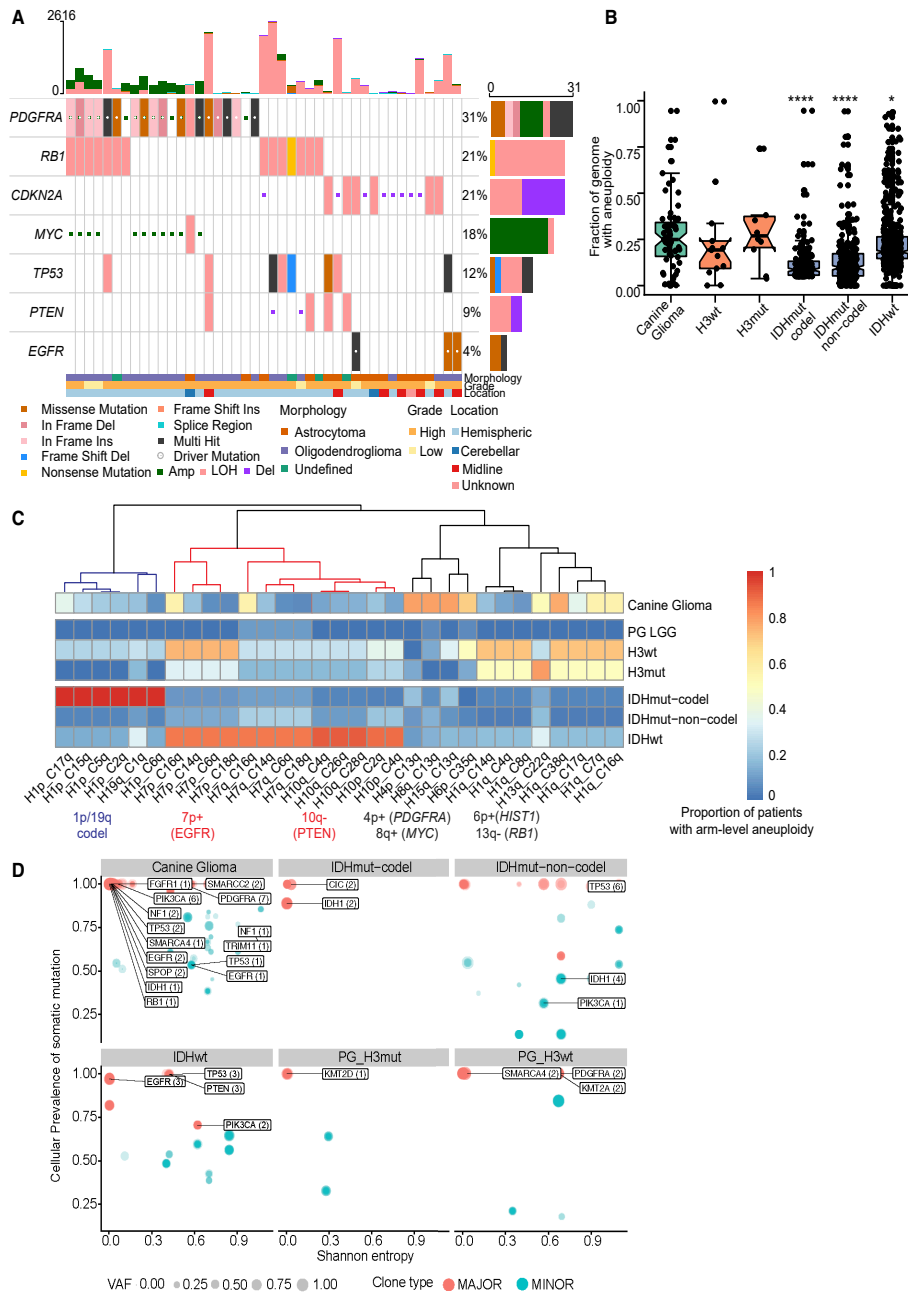
We compared the DNA copy-number landscape of glioma across species with a focus on the >50% of canine gliomas (45/81) without evidence of significantly mutated genes. No focal copy-number amplifications were detected among canine gliomas. Human glioma tumor-suppressor gene *CDKN2A/B* was homozygously deleted in 8/67 (12%, all astrocytomas), and *PTEN* in 2/67 (3%) of canine glioma genomes (Figures 2A and S2A). Together, 67/81 (83%) patients with canine glioma contained somatic mutations and/or focal copy alterations in known human glioma drivers (Figure S1E). Contrasting with the limited presence of focal DNA copy-number alterations was the high frequency of arm-level copy gains (canine chromosomes 7q, 13q, 16q, 20q, 34q, 35q, and 38q) and arm-level losses (canine

(B) Gene lollipop plots showing recurrent hotspot mutations for three genes: *PIK3CA*, *IDH1*, and *SPOP*. All hotspot mutations are ortholog to validated COSMIC mutations in human cancers.

(C) Hallmark enrichment of somatic cancer drivers (mutations and copy-number alterations) across canine glioma (CG) and WHO molecular subtypes of human adult (IDH wild-type, IDHmut-codel, IDHmut-noncodel) and pediatric (H3-mutant and H3 wild-type) high-grade glioma. y axis represents proportion of patients in the respective cohort harboring mutations in selected five hallmarks. Two-sided Fisher’s exact test was used for comparison of proportions between cohorts. p values less than the threshold ( $p < 0.05$ ) are shown (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ).

(D) Somatic mutation rate across canine and human brain tumors: Box plot showing somatic mutation rates as coding mutations per megabase in  $\log_{10}$  or  $\log_{10}(x+1)$  scale. x axis shows 11 types of pediatric brain tumors (Gröbner et al., 2018), canine glioma (n = 81), adult pediatric high-grade gliomas separated by H3 mutant and H3 wild type, and adult gliomas separated by IDH mutation and 1p/19q codeletion status (far right). Each box spans the first and third quartiles with the median in the center. The lower and upper whiskers extend up to 1.5 times interquartile range, and values outside whiskers are outliers. PA, pilocytic astrocytoma; ATRT, atypical teratoid rhabdoid tumor; EPD\_ST, ependymoma supratentorial; ETMR, embryonal tumors with multilayered rosettes; MB, medulloblastoma. Tumors are sorted in ascending order by increasing mutation rate. See also Figure S1 and Tables S1, S2, S3, and S4.

# Mutational signatures of genotoxic cancer therapies



(legend on next page)

chromosomes 1q, 5q, 12q, 22q, and 26q) (Figure S2B). The most frequent arm-level alteration comprised the shared syntenic regions of glioma drivers *PDGFRA*, *KIT*, and *MYC* (Figure S2C) and typically resulted in more than four copies of these genes (canine 13q+; 11/67 cases, 16%). Other common arm-level alterations included *PIK3CA* (canine 34q+) and the *HIST1* cluster (canine 35q+) as well as hemizygous loss of heterozygosity of tumor-suppressor genes *TP53*, *RB1*, and *PTEN* (Figure 2A).

We quantified the prevalence of aneuploidy across the canine, human pediatric, and adult glioma populations (Taylor et al., 2018). For copy-number estimation, matched tumor-normal whole-genome sequencing profiles from canine (n = 67) and pediatric gliomas (n = 50) (Ma et al., 2018), and Affymetrix SNP6 profiles for adult gliomas (n = 969) (Ceccarelli et al., 2016) were analyzed (STAR Methods). We calculated aneuploidy as the proportion of the copy-number segmented genome that was non-diploid (STAR Methods). Canine glioma independent of tumor grade had a median of 25% genome aneuploidy, which was significantly higher than that in adult IDH-mutant tumors (8%–9% of genome) and marginally higher than in adult IDH wild-type glioma (18% of genome) (Figure 2B). In contrast, pediatric H3 wild type (19% of genome) and H3 mutant (26% of genome) showed rates of aneuploidy comparable with that of canine glioma. We then searched for aneuploidy within syntenic regions, which may be subject to selection pressure during gliomagenesis. We mapped canine chromosome arms to their human counterparts and used unsupervised hierarchical clustering of the most variable syntenic aneuploid regions to identify regions of shared aneuploidy (Figure 2C). The analysis revealed three aneuploidy clusters. The first cluster (blue dendrogram) consisted of human 1p/19q codeletions seen commonly in adult IDH-mutant gliomas but observed in 20%–36% of canine (across four canine chromosomes of cases) and 25% of H3 wild-type and H3-mutant human pediatric gliomas. The second cluster (red dendrogram) consisted of arm-level aneuploidy of human 7p (*EGFR*) and 10q (*PTEN*) arms characteristic of human adult IDH wild-type (86% and 92% of patients, respectively) and pediatric H3-mutant and H3 wild-type gliomas (33% and 75% of patients) for which 5% and 14% of canine gliomas showed

arm-level aneuploidy in the *EGFR* and *PTEN* regions, respectively. None of three *IDH1* mutant canine gliomas shared these syntenic aberrations, suggesting a mutually exclusive pattern as observed in human gliomas. The third cluster (black dendrogram) consisted of human 4p/8q and syntenic canine 13q arm, which contains the genes *PDGFRA* and *MYC*, amplified in 78% of canine gliomas. The *ACVR1* and the *HIST1* genes are frequently mutated in pediatric high-grade gliomas and in particular H3.1K27M diffuse intrinsic pontine glioma (Mackay et al., 2017). We observed loss of the syntenic human 2q/canine 36q region (containing *ACVR1*) within 37%, 28%, and 17% of canine, pediatric H3 wild-type, and H3-mutant gliomas, respectively. In contrast, this alteration was not observed in human pediatric or adult IDH-mutant glioma and was present in 6% of IDH wild-type adult gliomas. Similarly, human chromosome arm 6p/canine chromosome arm 35q, containing the *HIST1* gene cluster, was frequently amplified in canine gliomas (70%) and pediatric H3 wild-type (50%) and H3-mutant gliomas (13%) but not in pediatric low-grade or adult gliomas (<5%).

We measured intratumoral heterogeneity using the Shannon Diversity Index per each patient tumor sample across canine gliomas and different molecular subtypes of human gliomas. Shannon entropy value correlated with the proportion of variants per subclone and the total number of subclones in a tumor sample, i.e., values near zero indicated lower intratumoral diversity (homogeneity or a dominant clone), while values closer to 1 or higher were associated with increased diversity and tumors consisting of more than one subclone (Wolf et al., 2019). We found that the canine gliomas showed a relatively bimodal distribution, with 25% of canine gliomas (15/60; 7 cases had no resolved clonal structure for intratumoral heterogeneity analysis) being very heterogeneous (Shannon Diversity Index  $\geq 0.45$ —third quartile) while remaining cases showed patterns suggesting clonal dominance (median Shannon Diversity Index 0.03, 95% CI: 0.02–0.07). Intratumoral heterogeneity of canine gliomas was comparable with adult IDH wild type or IDHmut-noncodels (unpaired two-tailed Wilcoxon test p value >0.18). In contrast, canine gliomas had significantly higher heterogeneity over H3 wild-type

## Figure 2. Aneuploidy Is a Major Driver of High-Grade Gliomas

(A) Focal somatic copy alterations in canine gliomas (n = 43 of 67 canine patients). Squared symbol in cell suggests either amplification (>4 copies) or deep deletion (2 copy loss) based on GISTIC2 gene-level calls (STAR Methods). Top bar plot shows patient-wise frequency of somatic variants and copy-number alterations, and right-side bar plot shows driver-wise frequency of somatic variant types, including copy-number alterations. Bottom annotations show relevant patient-specific annotations.

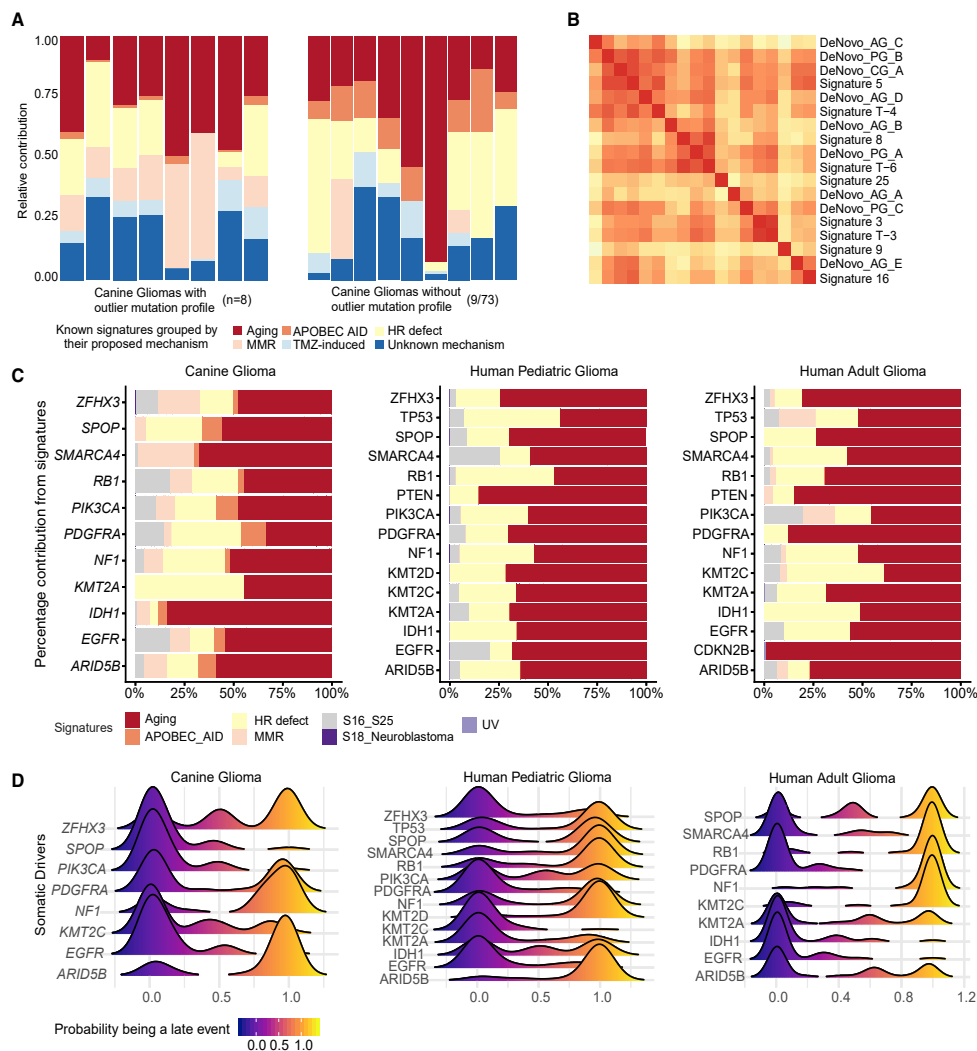
(B) Comparative aneuploidy score: box plots showing fraction of genome with aneuploidy (y axis) for canine gliomas (n = 67), H3-mutant (n = 10), and H3 wild-type (n = 13) pediatric high-grade gliomas, and human adult glioma (n = 969), separated by IDH mutation and 1p/19q codeletion status. Each box spans the first and third quartiles with the median in the center. The lower and upper whiskers extend up to 1.58 times interquartile range divided by square root of samples per box plot (displayed as dots; STAR Methods), and values outside whiskers are outliers. p values were calculated using two-sided Wilcoxon rank-sum non-parametric test (\*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, \*\*\*\*p < 0.0001).

(C) Aneuploidy metrics across shared syntenic regions of canine and human genome: Heatmap showing comparative aneuploidy across three cohorts. Each column shows the proportion of patients with the most variable arm-level aneuploidy (present or absent) for a given shared syntenic region. x-axis label shows syntenic chromosome arms for human (H) and canine (C) genome. Each row represents canine glioma and molecular subgroups of human high-grade pediatric and adult glioma as detailed in (B), plus pediatric low-grade gliomas (PG\_LGG). Colored dendrogram branches (blue, red, and black) represent three aneuploidy clusters described in the main text. Corresponding glioma driver alterations are highlighted below syntenic chromosome arms.

(D) Scatterplot showing distribution of somatic glioma driver genes with respect to their cellular prevalence (cancer cell fraction) and intratumoral heterogeneity (Shannon entropy) across canine and molecular subtypes of human pediatric and adult gliomas. Each circle represents a clonal cluster assignment per tumor sample. Size of the circle represents a major (1 clone) versus minor subclones (ranging from 2 to 4). Labeled genes represent glioma drivers shown in Figure 1A. Darker to lighter blue scale for circle and driver genes it may contain (arrows) represents the increase in intratumoral heterogeneity as measured by Shannon entropy.

See also Figure S2 and Table S3.





**Figure 3. Molecular Life History Analysis Using Mutational Signatures and Timing Analysis**

(A) Deconvolution of known human mutational signatures on canine glioma somatic variant data. Stacked bar plots show relative contribution of known human mutational signatures in individual canine patients. Signature contributions were aggregated based on their grouping into proposed mechanism. Only signatures with a relative contribution of more than a third quartile per sample are shown in the plot. Plot on the left side shows eight cases with highest mutational frequency (based on outlier mutational profile, STAR Methods) and plot on the right side shows nine representative cases with median signature contribution within interquartile range. Signatures with no proposed mechanism are grouped into the unknown category. APOBEC AID, activation-induced cytidine deaminases; HR defect, homologous repair defect; MMR, mismatch repair; TMZ-induced, alkylating agent temozolomide-associated signature.

(B) Hierarchical clustering of cosine similarities between known human mutational signatures and *de novo* signatures constructed using available whole-genome data from canine (CG), pediatric (PG), and adult (AG) data. Higher cosine similarity (red color) indicates higher resemblance of *de novo* signature to known mutational signature. Only one of three cluster groups are shown here; the complete clustering is shown in Figure S3D.

(C) Horizontal stacked bar plots represent percentage contribution of signature groups (x axis) for somatic driver mutations (y axis) found in canine and human gliomas. Each of seven signature groups represents a combination of one or more known human signatures. S16\_S25 and S18\_Neuroblastoma: signatures were previously described by Gröbner et al. (2018).

(legend continued on next page)



(p value 0.002) and H3-mutant (p value 0.007) pediatric gliomas (Figure S2D).

To better understand the potential functional versus non-functional nature of intratumor heterogeneity (Jamal-Hanjani et al., 2017), we asked whether frequent driver mutations found in canine gliomas (Figure 1A) are part of major (dominant) versus minor clone, and how these driver events compare with measured heterogeneity across molecular subtypes in human pediatric and adult gliomas (Williams et al., 2016). We observed that among less heterogeneous tumor samples (Shannon entropy near 0), shared driver events across canine and adult gliomas are part of major clones, including *PIK3CA* mutations in canine gliomas, *IDH1/2* mutations in IDH-mutant adult gliomas, and *EGFR* somatic mutations in adult IDH wild-type GBM (Figure 2D). Among tumor samples with increased heterogeneity, we found mutations in *PDGFRA* in canine gliomas (n = 7/60) and H3 wild-type pediatric gliomas (n = 2/14), whereas mutations in *TP53* (n = 9/20) and *PTEN* (3/20) were seen among IDH wild-type and IDHmut-noncode patients. We did not observe significant enrichment of driver events within minor clone(s).

Collectively, the observed high degree of aneuploidy and clonal nature of somatic drivers in canine glioma may suggest progressive genomic instability. Comparing the aneuploidy score among canine gliomas with high versus low coding mutational rate showed significant increases (Figure S2E; Wilcoxon p value 0.006) in aneuploidy among patients with a high mutational rate, suggesting that an underlying mutational process promotes genomic instability during gliomagenesis.

### DNA Damage-Related Mutational Processes Shape Somatic Driver Landscape and Maintain Genomic Instability

We leveraged known mutational signatures from adult (COSMIC v2, 1 to 30) and pediatric cancers (T1 to T12) to estimate and compare underlying mutational processes across canine and human gliomas (Alexandrov et al., 2013; Gröbner et al., 2018; Ma et al., 2018) (Table S5). The most enriched signatures across all canine gliomas (Figure S3A) were associated with aging (COSMIC signature 1, pediatric signature T1), mismatch repair deficiency (COSMIC signature 15), APOBEC-AID (COSMIC signature 2, 9), homologous repair defect signatures (COSMIC signature 8, pediatric signature T3), and signatures with unknown relevance (COSMIC signature 12, pediatric signature T10 and T11). Among the nine canine gliomas with the highest mutation rates (median coding mutation rate of 0.55 per megabase) (Figure 3A), there was significant (Wilcoxon p value 0.025) enrichment of two additional mismatch repair signatures (pediatric signature T9 or COSMIC signature 6, 15) (Figure S3B). A frameshift indel in mismatch repair gene *MSH6* was detected in one case with an outlier mutation frequency (coding mutation rate of 5.04 per MB) (Figure S3C). Among the remaining cases (median coding mutation rate of 0.25 per MB), homologous repair defect or “BRCAness” signatures (COSMIC signature 3

or pediatric signature T3, COSMIC signature 8 or pediatric signature T6) were the second most prominent signatures after clock-like signatures (COSMIC signature 1, 5). Homologous repair defect signatures have been reported to be enriched in pediatric high-grade gliomas with higher genomic instability (Gröbner et al., 2018). The known human signatures were validated by clustering *de novo* constructed signatures for all three cohorts (canine, human adult, and pediatric gliomas). Independent of cohort type, we observed significant cosine similarity (>0.8; Figures 3B and S3D) of *de novo* signatures with known homologous repair defect mutational processes (including COSMIC signature 3/pediatric signature T3, COSMIC signature 8/pediatric signature T6), further implying a role for these mutational processes in cross-species gliomagenesis.

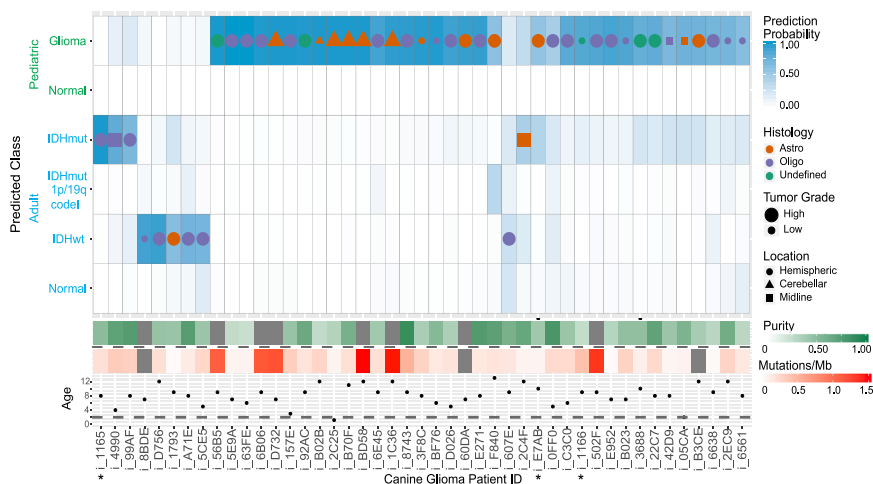
Next, we determined the relative contribution of mutational processes (with deconvoluted human signatures as a proxy) in generating mutations within significantly mutated genes, thus to identify the dominant mutational process(es) active during tumor evolution (Figure 3C). Although clock-like processes (COSMIC signature 1, 5) largely contributed to an age-related increase in mutations, including in driver genes, we found that homologous repair defect signatures (COSMIC signature 3, 8) contributed (26%, 21/81 cases) to driver mutations across all three cohorts, emphasizing that homologous repair defect can not only serve as a potential source for driver mutations but also fuel progressive genomic instability along with observed high aneuploidy (Blank et al., 2015; Targa and Rancati, 2018) in high-grade gliomas across all three cohorts.

### Comparative Molecular Timing Analysis Highlights Context-Specific Early and Late Drivers of Gliomagenesis

We inferred the sequential order of somatic alterations during gliomagenesis by estimating clonality of glioma driver events (Figure 3D) (Jolly and Van Loo, 2018; Shinde et al., 2018). In brief, significantly mutated genes were timed as occurring early (clonal) to late (subclonal) during tumor evolution based on their cancer cell fraction after accounting for tumor purity, ploidy, and copy-number status (STAR Methods). We observed clonal *PDGFRA* and *EGFR* mutations as the only shared and early event across all three cohorts. Subsequent whole chromosome 13 amplification bearing the *PDGFRA* mutant allele marked the emergence of the most recent common ancestor in six canine gliomas (Figure S3E), which grew to be a dominant clone at the time of diagnosis. *IDH1* mutation marks an initiating event in IDH-mutant human gliomas (Barthel et al., 2018). Correspondingly, *IDH1* mutations were ubiquitously timed as an initiating event (cancer cell fraction [CCF] > 0.9) in three canine and three human adult *IDH1* mutant cases, and as an early event in one case of pediatric glioma (CCF = 0.83). We observed *NF1* frameshift mutations mostly as a late event across all cohorts, whereas *PIK3CA* mutations appeared as an early event for canine and human

(D) Molecular timing of somatic drivers across canine and human gliomas: Stacked density plots, one per each of three cohorts, shows probability (x axis) of a driver event (y axis) being a late event in tumor evolution and value of <0.5 being an earlier event. Density plots for each driver event were calculated based on pairwise winning probability (where win is defined as an early event) as used in sports statistics (Bradley-Terry model). Winning probabilities were subtracted from 1 to display early events on the left side of the plot. See also Figure S3 and Table S5.

250 Cancer Cell 37, 243–257, February 10, 2020



**Figure 4. Classification of Canine Gliomas Using Human Brain Tumor Methylation Classifier**

Heatmap displaying results of L2-regularized, logistic regression classification of canine methylation profiles ( $n = 45$ ). Each column of the heatmap represents a sample, and each row in the top panel is the probability that that sample falls under a given subtype classification. The classification with the highest probability in a given sample has a symbol with symbol color, size, and shape denoting sample histology, tumor grade, and anatomical location, respectively. Panels below the probability heatmap show the tumor purity, somatic mutation rate, and age for the samples. The horizontal line on the age subpanel denotes the age of maturity for canines (2 years). \*Canine glioma patients with *IDH1* R132H somatic mutation. See also Figure S4.

pediatric gliomas. Although the relatively uniform timing patterns of these known glioma drivers suggest convergent evolution in varied contexts, i.e., presence of hotspot mutations in shared drivers (*PDGFRA*, *PIK3CA*) during clonal evolution of glioma across two species and different age groups, we also observed an oscillating pattern of timing and consequent underlying natural selection for a set of epigenetic drivers in the lysine methyltransferase (MLL) family (Rao and Dou, 2015). *MLL3* (*KMT2C*) gene mutations were clonal events in canine and pediatric gliomas but subclonal in adult gliomas, whereas *ARID5B* mutations showed the inverse pattern (Figure 3D). MLL family genes include some of the most commonly mutated genes in pediatric cancers, including gliomas (Huether et al., 2014; Sturm et al., 2014), but not in adult gliomas (Bailey et al., 2018).

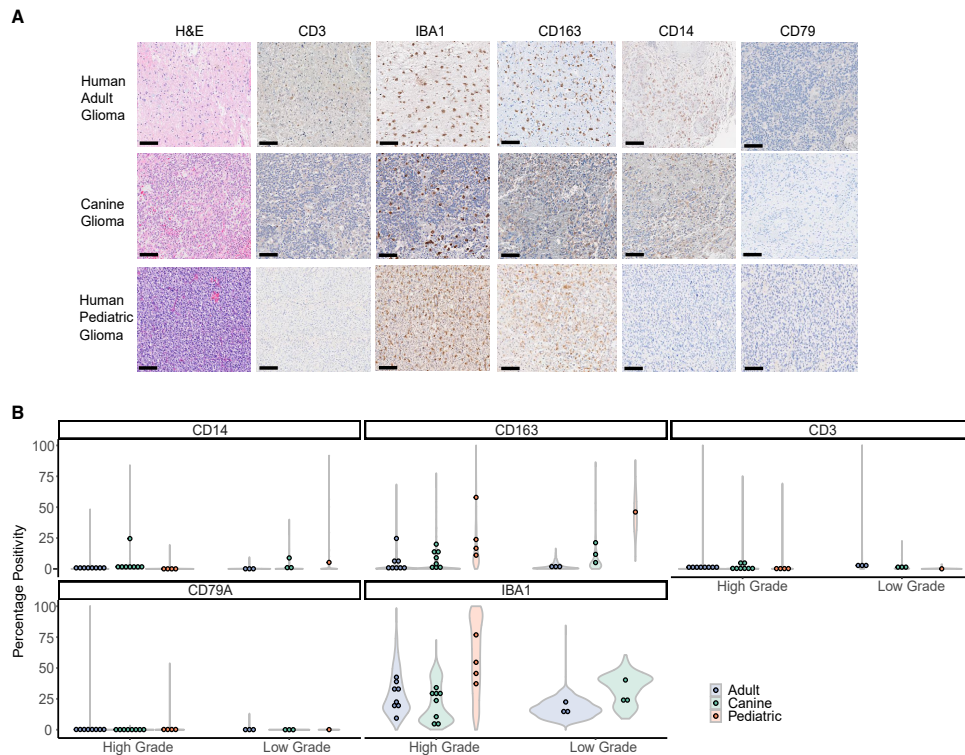
### Canine Gliomas Are Classified as Pediatric Glioma by DNA Methylation

We hypothesized that epigenetic deregulation in canine gliomas may carry a tumor-specific methylation pattern reflecting underlying tumor pathology, as has been observed across human brain tumors (Capper et al., 2018). We leveraged reduced representation bisulfite sequencing of canine gliomas to generate genome-wide DNA-methylation profiles to classify canine gliomas according to a classification model widely used for human brain tumors (Capper et al., 2018). As the human brain tumor classifier was developed using the Illumina human 450k array platform, we developed a logistic regression model to enable classification of the sequencing-based canine DNA-methylation profiles. We found that the model classified

35/45 (78%) of canine samples as pediatric glioma (Figure 4). Six of 45 (13%) samples were classified as IDH wild-type adult glioma, and 4/45 (9%) samples were classified as IDH-mutant adult glioma. Of the three samples carrying an *IDH1* R132H mutation, one was classified as IDH-mutant adult glioma, with a classification probability of 99%, while a second IDH-mutant sample had a relatively high classification probability for IDH-mutant adult glioma (40%), in parallel with a 57% pediatric glioma classification probability. The third sample had a low classification probability for IDH-mutant adult glioma (13%) and was classified as pediatric glioma with an 84% probability. Although the majority of canine samples were classified as pediatric glioma, the age of diagnosis of the patients in our canine cohort exceeded the age of sexual maturity in canines, which is reached between 10 months and 2 years of age (Thompson et al., 2017). The distribution of age of diagnosis of canine tumors classified as pediatric suggests that classification was a function of methylation profile similarity rather than chronological age. Adult human high-grade glioma tends to be restricted to the cerebral hemispheres, whereas pediatric high-grade gliomas occur throughout the central nervous system with about half of pediatric high-grade gliomas occurring in midline locations (Mackay et al., 2017). Of ten midline canine tumors (six cerebellar, four midline), eight were classified by DNA methylation as pediatric glioma and two were labeled as adult IDH mutant (Figure S4A).

DNA-methylation profiles have been used to estimate molecular age (Pai et al., 2011). We used this approach to compare the level of age acceleration in canine and human glioma. No significant difference was observed in inferred DNA-methylation age

# Mutational signatures of genotoxic cancer therapies



**Figure 5. Immunohistochemistry of Canine and Human Gliomas**  
 (A) Representative hematoxylin & eosin and immunohistochemistry staining of human adult (n = 11), canine (n = 11), and human pediatric gliomas (n = 5) using antibodies against T cells (CD3), macrophage/microglia (IBA1), M2 polarized innate immune cells (CD163), monocytes (CD14), and B cells (CD79A). Scale bars, 50  $\mu$ m.  
 (B) Violin plots represent the density of percentage positivity by field (y axis) for each of five antibodies described in (A). The points are the mean value of percentage positivity per patient within each of three cohorts, i.e., human adult (n = 11), canine (n = 11), and human pediatric gliomas (n = 5). Patients were grouped into high- versus low-grade gliomas in the absence of available molecular subtype data. See also Figure S5.

between canine tumors classified as adult glioma versus those classified as pediatric among tumors with a classification probability greater than 50% (5.945 versus 5.958, p value 0.9125), consistent with the lack of correlation observed between canine methylation classification and chronological age. The normalized mean age acceleration was significantly higher for human pediatric glioma samples (2.5) than either human adult glioma (0.8) or canine glioma samples (-0.18) (Figure S4B). Unlike human samples, the DNA-methylation-inferred age did not correlate with chronological age for canine samples (Pearson correlation coefficient 0.21), which may reflect limitations in the aging clock model derived for canids, rather than biological differences in canine tumor methylation. The DNA-methylation profile of canine glioma further corroborates the evidence that glioma in dogs is generally more similar to human pediatric glioma than human adult glioma.

## Immune Microenvironment

As spontaneous tumors arising in immune-competent hosts, canine gliomas represent an excellent resource through which to improve our understanding of how the immune system responds to and affects brain tumor development. To obtain a baseline understanding of how the canine glioma (n = 11) immune microenvironment compares with that of adult (n = 11) and pediatric human gliomas (n = 5), we used immunohistochemistry to profile the frequency of the various immune populations including monocytes (CD14), T cells (CD3), B cells (CD79A), macrophage/microglia (IBA1), and M2 polarized innate immune cells (CD163) using antibodies that had been validated for cross-species staining. Tissue segmentation was performed so that cell quantification was analyzed in a total of 33,029 fields within the gliomas. Notably, there are many key shared immunological features between the human and canine gliomas

such as the relative scarcity of both CD3<sup>+</sup> T cell and CD79A<sup>+</sup> B cells and a marked predominance of IBA1<sup>+</sup> macrophage/microglia and CD163<sup>+</sup> M2 polarized innate immune cells, especially in high-grade and pediatric gliomas as previously described (Wei et al., 2019), indicating that dogs with spontaneously arising gliomas may be valid models for the testing of immune therapeutics (Figures 5A and 5B). Our immunohistochemistry results converged with the relative immune cell fractions derived from RNA-sequencing data by using the leukocyte gene signature-based CIBERSORT deconvolution method (Newman et al., 2015), which we applied on gene expression profiles from human adult (n = 703), pediatric (n = 92), and canine glioma (n = 40) (Figure S5). The relative immune cell fractions found in each glioma type were well correlated with one another, with the low-grade pediatric glioma exhibiting the lowest correlation with high-grade canine glioma (Rho = 0.83).

## DISCUSSION

Comparative genomic oncology is a robust approach for identifying evolutionarily conserved drivers and for studying the natural history of spontaneous tumors in an immune-competent host, e.g., in domestic dogs (Decker et al., 2015a; Frampton et al., 2018; Tollis et al., 2017). Our cross-species analysis using comprehensive molecular profiling of sporadic gliomas highlights two key findings. First, convergent evolution of gliomas is observed across canine, human pediatric, and human adult gliomas, with shared molecular traits such as shared hotspot and mutually exclusive mutations in *PDGFRA* and *PIK3CA*, and in genes associated with the p53 and cell-cycle pathways, among others. This is further supported by aneuploidy being prevalent among canine and human pediatric high-grade gliomas, which are potentially under selection pressure within shared syntenic regions of the genome. Also, DNA damage-related mutational processes such as homologous recombination defects constitute a major source for progressive genomic instability, and generate somatic variations upon which natural selection acts to produce shared molecular and histopathological features of glioma. Second, the molecular landscape of canine gliomas resembles that of human pediatric gliomas based on the observed pattern of somatic alterations among non-shared drivers and DNA-methylation patterns. We did not observe canine counterparts of rare human glioma variants such as pleomorphic xanthoastrocytoma, giant cell glioblastoma, or pilocytic astrocytoma, by histopathology or by association in somatic drivers. To make a definitive claim that the canine gliomas are similar to one of the major categories of either adult or pediatric molecularly defined gliomas, additional characterization studies are needed that compare canine and human glioma in terms of cellular states (Nefel et al., 2019) as defined by single-cell transcriptomics.

Convergent evolution can reflect a footprint of adaptation to similar selective pressures (Fortunato et al., 2017). While such convergence is well appreciated in human cancers, and in particular treatment-resistant cancers (Venkatesan et al., 2017), our observation of such molecular and phenotypic convergence across two species provides a strong indicator of variations under selective pressures exerted by the tissue or ecological context (DeGregori, 2017; Schneider et al., 2017).

We note that convergent evolution should not discount a possibility of drivers unique to canine gliomas, especially within the context of germline variants (Truve et al., 2016) and non-coding regulatory regions (Lindblad-Toh et al., 2011; Villar et al., 2015). Characterization of such species-specific drivers can be of much value to identify evolutionary linchpins, which if abrogated can drive oncogenesis with similar histopathological and clinical traits. Further studies are needed to help understand how the time point at which tissue samples used in our comparative analysis were obtained, necropsy for canine samples, and diagnosis for human samples affects our results.

The molecular life history of a tumor is marked by multiple, often successive aberrations in genes (Armitage and Doll, 1954; Nowell, 1976). Accordingly, cancer is largely a disease of old age except in cases with early exposures to mutagens, e.g., germline or acquired defects in one or more hallmarks of cancer (Hanahan and Weinberg, 2011). The median age of occurrence for canine gliomas in our cohort was 9 years, i.e., dogs in their adult stage of life. However, we demonstrate that canine gliomas have a significantly lower somatic mutation rate and, consequently, a lower number of significantly mutated genes than adult human gliomas. The mutation burden of canine glioma is also less than what has been reported for other canine cancers, although a direct comparison would require additional standardization of sequencing and data-preprocessing methods (Hendricks et al., 2018; Lorch et al., 2019; Sakthikumar et al., 2018).

Canine gliomas harbor significantly higher aneuploidy than adult human high-grade gliomas, and are more similar to human pediatric gliomas (Gröbner et al., 2018; Mackay et al., 2017). We find additional support for aneuploidy as a major driver in canine and pediatric H3-mutant and H3 wild-type high-grade gliomas with the observation of aneuploidy in regions of shared synteny containing the *HIST1* and *PDGFRA* genes, known pediatric glioma drivers (Gröbner et al., 2018; Mackay et al., 2017), and in noting shared homologous repair defects as a mutational process that could drive genomic instability (Blank et al., 2015; Targa and Rancati, 2018). Recent efforts to engineer aneuploidy have provided better understanding of the functional role of aneuploidy and how it can be targeted in cancer (Bakhom and Cantley, 2018; Taylor et al., 2018). Canine high-grade gliomas carrying aneuploidy, especially among syntenic regions carrying the *HIST1* and *ACVR1* genes, can be utilized as a preclinical model for such functional screening as well as to validate recent studies showing its role in immune evasion (Bakhom et al., 2018; Davoli et al., 2017).

Tissue context and tumor microenvironment are critical factors for tumorigenesis (Haigis et al., 2019; Wang et al., 2017), and current models are unable to accurately represent the development of spontaneous tumors (Buque and Galluzzi, 2018). This renders preclinical evaluation ineffective and increases costs of clinical trials and results in minimal yields for patients. Preclinical trials of dog glioma patients enable identification of evolutionarily constrained and potentially targetable drivers, but simultaneously benefit dogs with glioma by offering treatment options that otherwise are prohibitive due to associated healthcare costs (LeBlanc et al., 2016). Future efforts leveraging results from the comparative genomics of glioma to study immune-mediated host responses can shed light on the complex interplay between

the tumor and host immune response and also aid in optimizing ongoing parallel canine clinical trials (Addissie and Klingemann, 2018) in order to improve an otherwise limited response to immunotherapies in canine and human gliomas. Our findings of canine gliomas with low tumor mutational burden but with a clonal nature of somatic drivers would be relevant in development of a preclinical model to dissect the interplay between mutation burden and immune escape during tumor evolution due to loss of clonal neoantigens (McGranahan and Swanton, 2019; Wolf et al., 2019). With respect to the immune microenvironment, differences in immune cell gene expression patterns between species could confound immune cell comparisons by under- or overestimating the presence of specific immune cell types. Despite these potential differences, comparative transcriptomic analyses of mouse and human immune cells have shown that the cells in each species exhibit a high degree of global conservation with one another, and signatures derived from murine immune cells have provided accurate immune infiltration estimates in human cancer types (Shay et al., 2013; Varn et al., 2017). Thus, the estimates in this study provide a baseline for how the relative fractions of major immune cells compare among adult, pediatric, and canine gliomas. We would have liked to have done further immune characterization including expression profiling of immune checkpoint response markers such as PD-1<sup>+</sup> tumor-infiltrating lymphocytes and PD-L1 immune and tumor expression, but these antibodies have not been validated in canines. Moving forward, signatures derived from canine immune cells will be of value in examining the presence of more specific immune cell types.

In summary, our study shows that the comparative molecular life history of gliomas details conserved drivers of glioma at both the genetic and epigenetic levels, with aneuploidy as a major hallmark of high-grade disease. Our results effectively position preclinical models of spontaneous canine glioma for use in understanding glioma drivers, and evaluating therapies targeting aneuploidy as well as immunotherapies, with relevance to all human gliomas and pediatric disease in particular.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [LEAD CONTACT AND MATERIALS AVAILABILITY](#)
- [EXPERIMENTAL MODELS AND SUBJECT DETAILS](#)
  - Canine Patients and Tissue Samples
- [METHOD DETAILS](#)
  - Published Data Sources
  - Sample Preparation
  - Sequencing Alignments, QC, and Fingerprinting
  - Somatic Variant Calling
  - Significantly Mutated Genes (SMGs) Analysis
  - Cancer Hallmark Analysis
  - Quantifying Somatic Mutation Rates
  - Somatic Copy Number Segmentation
  - Allele Specific Copy-Number Analysis
  - Aneuploidy Metrics
  - Clustering Shared Syntenic Regions

- Estimating Intra-tumoral Heterogeneity
- Mutational Signature Analysis
- Molecular Timing Analysis and Natural History of Tumors
- Class Prediction Using Methylation Data
- Immunohistochemistry
- CIBERSORT Based Expression Analysis
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)
- [DATA AND CODE AVAILABILITY](#)

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.ccell.2020.01.004>.

## ACKNOWLEDGMENTS

This work is supported by grants from the National Institutes of Health (NIH): Cancer Center Support grants P30CA16672 and P30CA034196 and Cancer Center Support Grant Supplement 3P30CA016672-41S7 (A.B.H.); R01 CA190121 (R.G.W.V.), R01 CA120813 (A.B.H.); P0 1CA207206 (J.H.R.); an unrestricted grant from Agilent Technologies (R.G.W.V.); and philanthropic support from Mr. Herb Simmons (A.B.H.). E.K. is recipient of an MD-Fellowship by the Boehringer Ingelheim Fonds and is supported by the German Academic Scholarship Foundation. 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 K99 CA226387. K.C.J. is the recipient of an American Cancer Society Fellowship (130984-PF-17-141-01-DMC). This work was partially supported (A.K.L., C.M.) by the Intramural Program of the National Cancer Institute, NIH (Z01-BC006161). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

## AUTHOR CONTRIBUTIONS

R.G.W.V., J.M.L., and A.B.H. conceived, supervised, and financially supported the study. C.E.B., P.J.D., J.W.K., J.M.L., R.A.P., J.H.R., and A.R.T. provided canine patient samples. J.W.K., A.D.M., C.R.M., B.F.P., D.R.R., C.M., K.D.W., and A.K.L. provided consensus histopathological classification on canine gliomas. Sample processing, quality control, and sequencing was performed by C.Y.N. and M.B. S.B.A. and R.G.W.V. designed analysis themes. S.B.A., K.C.J., C.E.B., and M.C. collected patient samples and curated meta-data. Data analysis was led by S.B.A. in collaboration with K.J.A., K.C.J., F.P.B., E.K., H.K., E.M.-L., and F.S.V. All authors participated in the discussion of the results. S.B.A. and R.G.W.V. wrote the manuscript. All co-authors discussed the results and commented on the manuscript and Supplementary Information.

## DECLARATION OF INTERESTS

R.G.W.V. declares equity in Boundless Bio, Inc. A.B.H. receives royalties and milestone payments for licensed intellectual property from Celldex Therapeutics, research grant support from Merck, and is a scientific board member for Caris Life Sciences. The other authors declare no competing interests.

Received: June 14, 2019

Revised: November 15, 2019

Accepted: January 10, 2020

Published: February 10, 2020

## REFERENCES

- Addissie, S., and Klingemann, H. (2018). Cellular immunotherapy of canine cancer. *Vet. Sci.* 5, <https://doi.org/10.3390/vetsci5040100>.
- Aktipis, C.A., Boddy, A.M., Gatenby, R.A., Brown, J.S., and Maley, C.C. (2013). Life history trade-offs in cancer evolution. *Nat. Rev. Cancer* 13, 883–892.



- Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L., et al. (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421.
- Alizadeh, A.A., Aranda, V., Bardelli, A., Blanpain, C., Bock, C., Borowski, C., Caldas, C., Califano, A., Doherty, M., Elsnér, M., et al. (2015). Toward understanding and exploiting tumor heterogeneity. *Nat. Med.* 21, 846–853.
- Angermueller, C., Lee, H.J., Reik, W., and Stegle, O. (2017). DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. *Genome Biol.* 18, 67.
- Armitage, P., and Doll, R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* 8, 1–12.
- Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive characterization of cancer driver genes and mutations. *Cell* 173, 371–385.e18.
- Bakhoun, S.F., and Cantley, L.C. (2018). The multifaceted role of chromosomal instability in cancer and its microenvironment. *Cell* 174, 1347–1360.
- Bakhoun, S.F., Ngo, B., Laughney, A.M., Cavallo, J.A., Murphy, C.J., Ly, P., Shah, P., Sriram, R.K., Watkins, T.B.K., Taunk, N.K., et al. (2018). Chromosomal instability drives metastasis through a cytosolic DNA response. *Nature* 553, 467–472.
- Barthel, F.P., Wesseling, P., and Verhaak, R.G.W. (2018). Reconstructing the molecular life history of gliomas. *Acta Neuropathol.* 135, 649–670.
- Blank, H.M., Sheltzer, J.M., Meehl, C.M., and Aron, A. (2015). Mitotic entry in the presence of DNA damage is a widespread property of aneuploidy in yeast. *Mol. Biol. Cell* 26, 1440–1451.
- Boussiotis, V.A., and Charest, A. (2018). Immunotherapies for malignant glioma. *Oncogene* 37, 1121–1141.
- Brennan, C.W., Verhaak, R.G., McKenna, A., Campos, B., Nourshahr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al. (2013). The somatic genomic landscape of glioblastoma. *Cell* 155, 462–477.
- Broeckx, B.J., Hitte, C., Coopman, F., Verhoeven, G.E., De Keulenaer, S., De Meester, E., Derrien, T., Alfoldi, J., Lindblad-Toh, K., Bosmans, T., et al. (2015). Improved canine exome designs, featuring ncRNAs and increased coverage of protein coding genes. *Sci. Rep.* 5, 12810.
- Brown, N.F., Carter, T.J., Ottaviani, D., and Mulholland, P. (2018). Harnessing the immune system in glioblastoma. *Br. J. Cancer* 119, 1171–1181.
- Buque, A., and Galluzzi, L. (2018). Modeling tumor immunology and immunotherapy in mice. *Trends Cancer* 4, 599–601.
- Cancer Genome Atlas Research Network, Brat, D.J., Verhaak, R.G., Aldape, K.D., Yung, W.K., Salama, S.R., Cooper, L.A., Rheinbay, E., Miller, C.R., Vitucci, M., et al. (2015). Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* 372, 2481–2498.
- Capper, D., Jones, D.T.W., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., Koelsche, C., Sahm, F., Chavez, L., Reuss, D.E., et al. (2018). DNA methylation-based classification of central nervous system tumours. *Nature* 555, 469–474.
- Ceccarelli, M., Barthel, F.P., Malta, T.M., Sabedot, T.S., Salama, S.R., Murray, B.A., Morozova, O., Newton, Y., Radenbaugh, A., Pagnotta, S.M., et al. (2016). Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* 164, 550–563.
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, 1884–1890.
- Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 213–219.
- Davoli, T., Uno, H., Wooten, E.C., and Elledge, S.J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* 355, <https://doi.org/10.1126/science.aaf8399>.
- Decker, B., Davis, B.W., Rimbault, M., Long, A.H., Karlins, E., Jagannathan, V., Reiman, R., Parker, H.G., Drögemüller, C., Corneveaux, J.J., et al. (2015a). Comparison against 186 canid whole-genome sequences reveals survival strategies of an ancient clonally transmissible canine tumor. *Genome Res.* 25, 1646–1655.
- Decker, B., Parker, H.G., Dhawan, D., Kwon, E.M., Karlins, E., Davis, B.W., Ramos-Vara, J.A., Bonney, P.L., McNeil, E.A., Knapp, D.W., and Ostrander, E.A. (2015b). Homologous mutation to human BRAF V600E is common in naturally occurring canine bladder cancer—evidence for a relevant model system and urine-based diagnostic test. *Mol. Cancer Res.* 13, 993–1002.
- Dees, N.D., Zhang, Q., Kandoth, C., Wendl, M.C., Schierding, W., Koboldt, D.C., Mooney, T.B., Callaway, M.B., Dooling, D., Mardis, E.R., et al. (2012). MuSiC: identifying mutational significance in cancer genomes. *Genome Res.* 22, 1589–1598.
- DeGregori, J. (2017). Connecting cancer to its causes requires incorporation of effects on tissue microenvironments. *Cancer Res.* 77, 6065–6068.
- DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498.
- Deshwar, A.G., Vembu, S., Yung, C.K., Jang, G.H., Stein, L., and Morris, Q. (2015). PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biol.* 16, 35.
- Dickinson, P.J., York, D., Higgins, R.J., LeCouteur, R.A., Joshi, N., and Bannasch, D. (2016). Chromosomal aberrations in canine gliomas define candidate genes and common pathways in dogs and humans. *J. Neuropathol. Exp. Neurol.* 75, 700–710.
- Fang, L.T., Afshar, P.T., Chhibber, A., Mohiyuddin, M., Fan, Y., Mu, J.C., Gibeling, G., Barr, S., Asadi, N.B., Gerstein, M.B., et al. (2015). An ensemble approach to accurately detect somatic mutations using SomaticSeq. *Genome Biol.* 16, 197.
- Fleshner, I., and Chernett, N.L. (1997). A wellness model for the geriatric population. *Home Care Provid* 2, 321–323.
- Fortunato, A., Boddy, A., Mallo, D., Aktipis, A., Maley, C.C., and Pepper, J.W. (2017). Natural selection in cancer biology: from molecular snowflakes to trait hallmarks. *Cold Spring Harb. Perspect. Med.* 7, <https://doi.org/10.1101/cshperspect.a029652>.
- Frampton, D., Schwenzer, H., Marino, G., Butcher, L.M., Pollara, G., Kriston-Vizi, J., Venturini, C., Austin, R., de Castro, K.F., et al. (2018). Molecular signatures of regression of the canine transmissible venereal tumor. *Cancer Cell* 33, 620–633.e6.
- Gentles, A.J., Newman, A.M., Liu, C.L., Bratman, S.V., Feng, W., Kim, D., Nair, V.S., Xu, Y., Khuong, A., Hoang, C.D., et al. (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat. Med.* 21, 938–945.
- Gerstung, M., Jolly, C., Leshchiner, I., Drento, S.C., Gonzalez, S., Mitchell, T.J., Rubanova, Y., Anur, P., Rosebrock, D., Yu, K., et al. (2017). The evolutionary history of 2,658 cancers. *bioRxiv*. <https://doi.org/10.1101/161562>.
- Gröbner, S.N., Worst, B.C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V.A., Johann, P.D., Balasubramanian, G.P., Segura-Wang, M., Brabetz, S., et al. (2018). The landscape of genomic alterations across childhood cancers. *Nature* 555, 321–327.
- Ha, G., Roth, A., Khattra, J., Ho, J., Yap, D., Prentice, L.M., Melnyk, N., McPherson, A., Bashashati, A., Laks, E., et al. (2014). TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome Res.* 24, 1881–1893.
- Haigis, K.M., Cichowski, K., and Elledge, S.J. (2019). Tissue-specificity in cancer: the rule, not the exception. *Science* 363, 1150–1151.
- Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674.
- Hendricks, W.P.D., Zismann, V., Sivaprakasam, K., Legendre, C., Poorman, K., Tembe, W., Perdignes, N., Kiefer, J., Liang, W., DeLuca, V., et al. (2018). Somatic inactivating PTPRJ mutations and dysregulated pathways identified in canine malignant melanoma by integrated comparative genomic analysis. *PLoS Genet.* 14, e1007589.
- Huether, R., Dong, L., Chen, X., Wu, G., Parker, M., Wei, L., Ma, J., Edmonson, M.N., Hedlund, E.K., Rusch, M.C., et al. (2014). The landscape of somatic



- mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat. Commun.* 5, 3630.
- Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J., Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A., et al. (2012). Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 150, 1107–1120.
- Iorio, F., Garcia-Alonso, L., Brammied, J.S., Martincorena, I., Wille, D.R., McDermott, U., and Saez-Rodriguez, J. (2018). Pathway-based dissection of the genomic heterogeneity of cancer hallmarks' acquisition with SLAPenrich. *Sci. Rep.* 8, 6713.
- Jamal-Hanjani, M., Wilson, G.A., McGranahan, N., Birkbak, N.J., Watkins, T.B.K., Veeriah, S., Shafi, S., Johnson, D.H., Mitter, R., Rosenthal, R., et al. (2017). Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* 376, 2109–2121.
- Jolly, C., and Van Loo, P. (2018). Timing somatic events in the evolution of cancer. *Genome Biol.* 19, 95.
- Jones, D.T., Hutter, B., Jager, N., Korshunov, A., Kool, M., Warnatz, H.J., Zichner, T., Lambert, S.R., Ryzhova, M., Quang, D.A., et al. (2013). Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. *Nat. Genet.* 45, 927–932.
- Khanna, C., Lindblad-Toh, K., Vail, D., London, C., Bergman, P., Barber, L., Breen, M., Kitchell, B., McNeil, E., Modiano, J.F., et al. (2006). The dog as a cancer model. *Nat. Biotechnol.* 24, 1065–1066.
- Koboldt, D.C., Larson, D.E., and Wilson, R.K. (2013). Using VarScan 2 for germline variant calling and somatic mutation detection. *Curr. Protoc. Bioinformatics* 44, 15.14.1–17. <https://doi.org/10.1002/0471250953.bi1504s44>.
- Koehler, J.W., Miller, A.D., Miller, C.R., Porter, B., Aldape, K., Beck, J., Brat, D., Cornax, I., Corps, K., Frank, C., et al. (2018). A revised diagnostic classification of canine glioma: towards validation of the canine glioma patient as a naturally occurring preclinical model for human glioma. *J. Neuropathol. Exp. Neurol.* 77, 1039–1054.
- Koster, J., and Rahmann, S. (2018). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* 34, 3600.
- Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571–1572.
- LeBlanc, A.K., Mazcko, C., Brown, D.E., Koehler, J.W., Miller, A.D., Miller, C.R., Bentley, R.T., Packer, R.A., Breen, M., Boudreau, C.E., et al. (2016). Creation of an NCI comparative brain tumor consortium: informing the translation of new knowledge from canine to human brain tumor patients. *Neuro Oncol.* 18, 1209–1218.
- Lee, S., Lee, S., Ouellette, S., Park, W.Y., Lee, E.A., and Park, P.J. (2017). NGSCheckMate: software for validating sample identity in next-generation sequencing studies within and across data types. *Nucleic Acids Res.* 45, e103.
- Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Maucelli, E., et al. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478, 476.
- Lorch, G., Sivaprakasam, K., Zismann, V., Perdignes, N., Contente-Cuomo, T., Nazareno, A., Facista, S., Wong, S., Drenner, K., Liang, W.S., et al. (2019). Identification of recurrent activating HER2 mutations in primary canine pulmonary adenocarcinoma. *Clin. Cancer Res.* 25, 5866–5877.
- Louis, D.N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., and Ellison, D.W. (2016). The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* 131, 803–820.
- Ma, X., Liu, Y., Liu, Y., Alexandrov, L.B., Edmonson, M.N., Gawad, C., Zhou, X., Li, Y., Rusch, M.C., Easton, J., et al. (2018). Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* 555, 371–376.
- Mackay, A., Burford, A., Carvalho, D., Izquierdo, E., Fazal-Salom, J., Taylor, K.R., Bjerke, L., Clarke, M., Vinci, M., Nandhabalan, M., et al. (2017). Integrated molecular meta-analysis of 1,000 pediatric high-grade and diffuse intrinsic pontine glioma. *Cancer Cell* 32, 520–537.e5.
- Mansour, T.A., Lucot, K., Konopelski, S.E., Dickinson, P.J., Sturges, B.K., Vernau, K.L., Choi, S., Stern, J.A., Thomasy, S.M., Doring, S., et al. (2018). Whole genome variant association across 100 dogs identifies a frame shift mutation in DISHEVELLED 2 which contributes to Robinow-like syndrome in Bulldogs and related screw tail dog breeds. *PLoS Genet.* 14, e1007850.
- Martincorena, I., Raine, K.M., Gerstung, M., Dawson, K.J., Haase, K., Van Loo, P., Davies, H., Stratton, M.R., and Campbell, P.J. (2017). Universal patterns of selection in cancer and somatic tissues. *Cell* 171, 1029–1041.e21.
- McGranahan, N., and Swanton, C. (2019). Neocantigen quality, not quantity. *Sci. Transl. Med.* 11, <https://doi.org/10.1126/scitranslmed.aax7918>.
- McGranahan, N., Favero, F., de Bruin, E.C., Birkbak, N.J., Szallasi, Z., and Swanton, C. (2015). Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci. Transl. Med.* 7, 283ra254.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The ensembl variant effect predictor. *Genome Biol.* 17, 122.
- Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 12, R41.
- Neftel, C., Laffy, J., Filbin, M.G., Hara, T., Shore, M.E., Rahme, G.J., Richman, A.R., Silverbush, D., Shaw, M.L., Hebert, C.M., et al. (2019). An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* 178, 835–849.e21.
- Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457.
- Nowell, P.C. (1976). The clonal evolution of tumor cell populations. *Science* 194, 23–28.
- Okonechnikov, K., Conesa, A., and Garcia-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32, 292–294.
- Pai, A.A., Bell, J.T., Marioni, J.C., Pritchard, J.K., and Gilad, Y. (2011). A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. *PLoS Genet.* 7, e1001316.
- Parker, H.G., VonHoldt, B.M., Quignon, P., Margulies, E.H., Shao, S., Mosher, D.S., Spady, T.C., Elkahloun, A., Cargill, M., Jones, P.G., et al. (2009). An expressed fgf4 retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science* 325, 995–998.
- Pimentel, H., Bray, N.L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* 14, 687–690.
- Pollack, I.F., Agnihotri, S., and Broniszer, A. (2019). Childhood brain tumors: current management, biological insights, and future directions. *J. Neurosurg. Pediatr.* 23, 261–273.
- Rao, R.C., and Dou, Y. (2015). Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat. Rev. Cancer* 15, 334–346.
- Sakthikumar, S., Elvers, I., Kim, J., Arendt, M.L., Thomas, R., Turner-Maier, J., Swofford, R., Johnson, J., Schumacher, S.E., Alfoldi, J., et al. (2018). SETD2 is recurrently mutated in whole-exome sequenced canine osteosarcoma. *Cancer Res.* 78, 3421–3431.
- Schneider, G., Schmidt-Suppran, M., Rad, R., and Saur, D. (2017). Tissue-specific tumorigenesis: context matters. *Nat. Rev. Cancer* 17, 239–253.
- Shay, T., Jovic, V., Zuk, O., Rothamel, K., Puyraimond-Zemmour, D., Feng, T., Wakamatsu, E., Benoist, C., Koller, D., Regev, A., and ImmGen, C. (2013). Conservation and divergence in the transcriptional programs of the human and mouse immune systems. *Proc. Natl. Acad. Sci. U S A* 110, 2946–2951.
- Shinde, J., Bayard, Q., Imbeaud, S., Hirsch, T.Z., Liu, F., Renault, V., Zucman-Rossi, J., and Letouze, E. (2018). Palimpsest: an R package for studying

- mutational and structural variant signatures along clonal evolution in cancer. *Bioinformatics* 34, 3380–3381.
- Snyder, J.M., Shofer, F.S., Van Winkle, T.J., and Massicotte, C. (2006). Canine intracranial primary neoplasia: 173 cases (1986–2003). *J. Vet. Intern. Med.* 20, 669–675.
- Song, R.B., Vite, C.H., Bradley, C.W., and Cross, J.R. (2013). Postmortem evaluation of 435 cases of intracranial neoplasia in dogs and relationship of neoplasm with breed, age, and body weight. *J. Vet. Intern. Med.* 27, 1143–1152.
- Stearns, S.C. (1992). *The Evolution of Life Histories* (Oxford University Press).
- Sturm, D., Bender, S., Jones, D.T., Lichter, P., Grill, J., Becher, O., Hawkins, C., Majewski, J., Jones, C., Costello, J.F., et al. (2014). Paediatric and adult glioblastoma: multiform (epi)genomic culprits emerge. *Nat. Rev. Cancer* 14, 92–107.
- Targa, A., and Rancati, G. (2018). Cancer: a CINful evolution. *Curr. Opin. Cell Biol.* 52, 136–144.
- Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 47, D941–D947.
- Taylor, A.M., Shih, J., Ha, G., Gao, G.F., Zhang, X., Berger, A.C., Schumacher, S.E., Wang, C., Hu, H., Liu, J., et al. (2018). Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* 33, 676–689.e3.
- Thompson, M.J., vonHoldt, B., Horvath, S., and Pellegrini, M. (2017). An epigenetic aging clock for dogs and wolves. *Aging (Albany NY)* 9, 1055–1068.
- Tollis, M., Schiffman, J.D., and Boddy, A.M. (2017). Evolution of cancer suppression as revealed by mammalian comparative genomics. *Curr. Opin. Genet. Dev.* 42, 40–47.
- Truve, K., Dickinson, P., Xiong, A., York, D., Jayashankar, K., Pielberg, G., Koltochian, M., Muren, E., Fuxelius, H.H., Weishaupt, H., et al. (2016). Utilizing the dog genome in the search for novel candidate genes involved in glioma development—genome wide association mapping followed by targeted massive parallel sequencing identifies a strongly associated locus. *PLoS Genet.* 12, e1006000.
- Varn, F.S., Wang, Y., Mullins, D.W., Fiering, S., and Cheng, C. (2017). Systematic pan-cancer analysis reveals immune cell interactions in the tumor microenvironment. *Cancer Res.* 77, 1271–1282.
- Venkatesan, S., and Swanton, C. (2016). Tumor evolutionary principles: how intratumor heterogeneity influences cancer treatment and outcome. *Am. Soc. Clin. Oncol. Educ. Book* 35, e141–e149.
- Venkatesan, S., Birkbak, N.J., and Swanton, C. (2017). Constraints in cancer evolution. *Biochem. Soc. Trans.* 45, 1–13.
- Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., et al. (2015). Enhancer evolution across 20 mammalian species. *Cell* 160, 554–566.
- Wang, Q., Hu, B., Hu, X., Kim, H., Squatrito, M., Scarpace, L., deCarvalho, A.C., Lyu, S., Li, P., Li, Y., et al. (2017). Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. *Cancer Cell* 32, 42–56.e6.
- Wei, J., Chen, P., Gupta, P., Ott, M., Zamlar, D., Kassab, C., Bhat, K.P., Curran, M.A., de Groot, J.F., and Heimberger, A.B. (2019). Immune biology of glioma associated macrophages and microglia: functional and therapeutic implications. *Neuro Oncol.* <https://doi.org/10.1093/neuroonc/noz212>.
- Williams, M.J., Werner, B., Graham, T.A., and Sottoriva, A. (2016). Functional versus non-functional intratumor heterogeneity in cancer. *Mol. Cell. Oncol.* 3, e1162897.
- Wilm, A., Aw, P.P., Bertrand, D., Yeo, G.H., Ong, S.H., Wong, C.H., Khor, C.C., Petric, R., Hibberd, M.L., and Nagarajan, N. (2012). LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res.* 40, 11189–11201.
- Wolf, Y., Bartok, O., Patkar, S., Eli, G.B., Cohen, S., Litchfield, K., Levy, R., Jimenez-Sanchez, A., Trabish, S., Lee, J.S., et al. (2019). UVB-induced tumor heterogeneity diminishes immune response in melanoma. *Cell* 179, 219–235.e21.
- Wong, K., van der Weyden, L., Schott, C.R., Foote, A., Constantino-Casas, F., Smith, S., Dobson, J.M., Murchison, E.P., Wu, H., Yeh, I., et al. (2019). Cross-species genomic landscape comparison of human mucosal melanoma with canine oral and equine melanoma. *Nat. Commun.* 10, 353.
- Zhang, J., Wu, G., Miller, C.P., Tatevossian, R.G., Dalton, J.D., Tang, B., Orisme, W., PUNCHIHewa, C., Parker, M., Qaddoumi, I., et al. (2013). Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. *Nat. Genet.* 45, 602–612.

# Mutational signatures of genotoxic cancer therapies



## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
CD3 antibody	Agilent	Cat# M725429-2; RRID: AB_2631163
CD79A antibody	Agilent	Cat# M705029-2; RRID: AB_2244527
CD163 antibody	MyBioSource	Cat# MBS9409179; RRID: N/A
IBA1 antibody	Wako	Cat# 019-19741; RRID: AB_839504
CD4 antibody	GeneTex	Cat# GTX84720; RRID: AB_10727465
CD14 antibody	Novus	Cat# NB100-77758; RRID: AB_1083332
<b>Biological Samples</b>		
Canine Glioma Patient Samples	This paper	Detailed under <a href="#">Table S1</a>
<b>Critical Commercial Assays</b>		
AllPrep DNA/RNA Mini Kit	Qiagen	N/A
AllPrep DNA/RNA FFPE Kit	Qiagen	N/A
KAPA Hyper Prep Kit (Illumina)	KAPA Biosystems/Roche	N/A
SeqCap EZ Canine Exome Custom Design	Roche Nimblegen	canine 140702_canFam3_exomeplus_BB_EZ_HX1 probe set
Nimblegen SepCap EZ Kit	Roche Nimblegen	N/A
KAPA Stranded mRNA-Seq kit	KAPA Biosystems/Roche	N/A
Premium RRBS Kit	Diagenode	N/A
<b>Deposited Data</b>		
DNA sequencing data - WGS and Exome	This paper	NCBI SRA Accession ID: PRJNA579792
RNA sequencing data	This paper	NCBI SRA Accession ID: PRJNA579792
RRBS sequencing data	This paper	NCBI SRA Accession ID: PRJNA579792
<b>Software and Algorithms</b>		
bwa	v0.7.15-r1140	<a href="http://bio-bwa.sourceforge.net/">http://bio-bwa.sourceforge.net/</a>
Genome Analysis ToolKit (GATK)	v4.0.8.1	<a href="https://software.broadinstitute.org/gatk/">https://software.broadinstitute.org/gatk/</a>
Qualimap	v2.2.1	<a href="http://qualimap.bioinfo.cipf.es/">http://qualimap.bioinfo.cipf.es/</a>
fastp	v0.19.5	<a href="https://github.com/OpenGene/fastp">https://github.com/OpenGene/fastp</a>
kallisto	v0.45.0	<a href="https://pachterlab.github.io/kallisto">https://pachterlab.github.io/kallisto</a>
sleuth	v0.30.0	<a href="https://pachterlab.github.io/sleuth">https://pachterlab.github.io/sleuth</a>
FastQC	v0.11.7	<a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc">https://www.bioinformatics.babraham.ac.uk/projects/fastqc</a>
TrimGalore	v0.5.0	<a href="https://github.com/FelixKrueger/TrimGalore">https://github.com/FelixKrueger/TrimGalore</a>
Bismark Bisulfite Mapper	v0.19.1	<a href="https://github.com/FelixKrueger/Bismark">https://github.com/FelixKrueger/Bismark</a>
Bowtie2	v2.2.3	N/A
NGSCheckMate	v1.0.0	<a href="https://github.com/parklab/NGSCheckMate">https://github.com/parklab/NGSCheckMate</a>
Mutect2 - GATK4	v4.0.8.1	<a href="https://software.broadinstitute.org/gatk/">https://software.broadinstitute.org/gatk/</a>
VarScan2	v2.4.2	<a href="https://github.com/dkoboldt/varscan">https://github.com/dkoboldt/varscan</a>
LoFreq	v2.1.3.1	<a href="https://github.com/CSB5/lofreq">https://github.com/CSB5/lofreq</a>
SomaticSeq	v3.1.0	<a href="https://github.com/bioinform/somaticseq">https://github.com/bioinform/somaticseq</a>
Variant Effect Predictor (VEP)	v91	<a href="https://github.com/Ensembl/ensembl-vep">https://github.com/Ensembl/ensembl-vep</a>
dNdScv	0.0.1.0	<a href="https://github.com/im3sanger/dndscv">https://github.com/im3sanger/dndscv</a>
MuSiC2	v0.2	<a href="https://github.com/ding-lab/MuSiC2">https://github.com/ding-lab/MuSiC2</a>
GISTIC2	v2.0.22	<a href="ftp://ftp.broadinstitute.org/pub/GISTIC2.0/GISTICDocumentation_standalone.htm">ftp://ftp.broadinstitute.org/pub/GISTIC2.0/GISTICDocumentation_standalone.htm</a>
HMMCopy	v1.22.0	<a href="http://bioconductor.org/packages/release/bioc/html/HMMcopy.html">http://bioconductor.org/packages/release/bioc/html/HMMcopy.html</a>

(Continued on next page)

e1 Cancer Cell 37, 243–257.e1–e7, February 10, 2020

<b>Continued</b>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
TitanCNA	v1.19.1	<a href="https://github.com/gavinha/TitanCNA">https://github.com/gavinha/TitanCNA</a>
Snakemake	v5.2.1	<a href="https://snakemake.readthedocs.io/en/stable/">https://snakemake.readthedocs.io/en/stable/</a>
flowr	v0.9.10	<a href="https://github.com/sahilseth/flowr">https://github.com/sahilseth/flowr</a>
NMF R package	v0.21.0	<a href="https://cran.r-project.org/web/packages/NMF">https://cran.r-project.org/web/packages/NMF</a>
Entropy R package	v1.2.1	<a href="https://cran.r-project.org/web/packages/entropy">https://cran.r-project.org/web/packages/entropy</a>
outliers R package	v0.14	<a href="https://cran.r-project.org/web/packages/outliers/">https://cran.r-project.org/web/packages/outliers/</a>
MutationalPattern R package	v1.6.2	<a href="https://bioconductor.org/packages/release/bioc/html/MutationalPatterns.html">https://bioconductor.org/packages/release/bioc/html/MutationalPatterns.html</a>
Palimpsest R package	v1.0.0	<a href="https://github.com/FunGeST/Palimpsest">https://github.com/FunGeST/Palimpsest</a>
BradleyTerryScalable R package	0.1.0.9000	<a href="https://cran.r-project.org/web/packages/BradleyTerryScalable/vignettes/BradleyTerryScalable.html">https://cran.r-project.org/web/packages/BradleyTerryScalable/vignettes/BradleyTerryScalable.html</a>
DNANexus app for St Jude Cloud data analysis	v1.1.6 (This paper)	<a href="https://dxapp.verhaaklab.com/dnanexus_ngsapp">https://dxapp.verhaaklab.com/dnanexus_ngsapp</a>
CIBERSORT	webservice	<a href="https://cibersort.stanford.edu">https://cibersort.stanford.edu</a>

## LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for data resources should be directed to and will be fulfilled by the Lead Contact, Roel Verhaak ([roel.verhaak@jax.org](mailto:roel.verhaak@jax.org)). This study did not generate new unique reagents.

## EXPERIMENTAL MODELS AND SUBJECT DETAILS

### Canine Patients and Tissue Samples

Tissue samples from canine patients with gliomas were acquired with material transfer agreements from Auburn University College of Veterinary Medicine, Colorado State University, Texas A&M College of Veterinary Medicine & Biomedical Sciences, UC Davis School of Veterinary Medicine and Virginia-MD College of Veterinary Medicine. Tissue samples from resected tumor (n=83) and matched normal tissue (n=67 or paired cases) were collected at the surgical treatment or immediately following euthanasia. There were also four additional dog patients where we had adequate DNA and RNA for methylation (n=48) and RNA-seq (n=40) profiling but unable to do WGS/Exome sequencing because of failed library preparation (Table S1). Matched normal tissue were from post-necropsy sample of contra-lateral healthy brain tissue (n=38), white blood cells (n=13), and remaining 17 samples from other tissues. Samples were archived in snap-frozen (n=37/67 paired cases; n=8/16 tumor-only cases) and Formalin-Fixed Paraffin-Embedded (FFPE, n=30/67 paired cases; n=8/16 tumor-only cases) state. Samples were then shipped to sequencing core facilities for sample preparation, quality control and sequencing (see STAR Methods Details below).

## METHOD DETAILS

### Published Data Sources

For comparison to human glioma, we downloaded both - raw sequencing data and processed tables for human pediatric and adult gliomas with appropriate controlled-data access agreements where needed. We used published mutation rates (Figure 1D) and mutational signatures (Figure 3) from pan-cancer datasets from adults (n=3,281) and pediatric (n=961) cohorts (Alexandrov et al., 2013; Bailey et al., 2018; Gröbner et al., 2018). For aneuploidy and molecular life history analysis (details below), we downloaded raw sequencing data and analyzed whole genomes from 53 pediatric gliomas (Ma et al., 2018; St. Jude Cloud Pediatric Cancer Genome Project, <https://pecan.stjude.cloud>), SNP6 data from adult gliomas - IDHwt (n=517), IDHmut-codel (n=171), and IDHmut-noncodel (n=281) cases (Ceccarelli et al., 2016), as well as whole genomes from 23 adult GBMs (Brennan et al., 2013). For coding mutation rate calculation, we used a subset of TCGA glioma set where exome/whole genome based variant calls were available: IDHwt (n=371), IDHmut-non-codel (n=268), and IDHmut-codel (n=169).

### Sample Preparation

**DNA/RNA extraction** - Genomic DNA and total RNA of fresh frozen tissue and FFPE tissue from paraffin scrolls were extracted simultaneously using AllPrep DNA/RNA Mini Kit (Qiagen) and AllPrep DNA/RNA FFPE Kit (Qiagen) according to the manufacturer's instructions, respectively. Additional DNase treatment was performed on-column for RNA purification. **WGS sample preparation** - 200-400ng of DNA was sheared to 400bp using a LE220 focused-ultrasonicator (Covaris) and size selected using Ampure XP beads (Beckman Coulter). The fragments were treated with end-repair, A-tailing, and ligation of Illumina compatible adapters (Integrated

DNA Technologies) using the KAPA Hyper Prep Kit (Illumina) (KAPA Biosystems/ Roche). For FFPE samples, 5 to 10 cycles of PCR amplification were performed. Quantification of libraries were performed using real-time qPCR (Thermo Fisher). Libraries were sequenced paired end reads of 151bp on Illumina HiSeq X-Ten (Novogene). **WES sample preparation** - Sample were prepared as described above in the WGS sample preparation, targeting 200bp with PCR amplification. Target capture was performed using Seq-Cap EZ Canine Exome Custom Design (canine 140702\_canFam3\_exomeplus\_BB\_EZ\_HX1 probe set) (Broeckx et al., 2015) (Roche Nimblegen). Briefly, WGS libraries were hybridized with capture probes using Nimblegen SepCap EZ Kit (Roche Nimblegen) according to manufacturer's instruction. Captured fragments were PCR amplified and purified using Ampure XP beads. Quantification of libraries were performed using real-time qPCR (Thermo Fisher). Libraries were sequenced paired end of 76bp on HiSeq4000 (Illumina). **RNA-seq sample preparation** - RNA-seq libraries were prepared with KAPA Stranded mRNA-Seq kit (Kapa Biosystem/ Roche) according to manufacturer's instruction. First, poly A RNA was isolated from 300ng total RNA using oligo-dT magnetic beads. Purified RNA was then fragmented at 85°C for 6 mins, targeting fragments range 250-300bp. Fragmented RNA is reverse-transcribed with an incubation of 25°C for 10mins, 42°C for 15 mins and an inactivation step at 70°C for 15mins. This was followed by second strand synthesis at 16°C, 60 mins. Double stranded cDNA (dscDNA) fragments were purified using Ampure XP beads (Beckman). The dscDNA were then A-tailed, and ligated with illumina compatible adaptors (IDT). Adaptor-ligated DNA was purified using Ampure XP beads. This is followed by 10 cycles of PCR amplification. The final library was cleaned up using Ampure XP beads. Quantification of libraries were performed using real-time qPCR (Thermo Fisher). Sequencing was performed on HiSeq4000 (Illumina) generating paired end reads of 75bp. **Reduced Representation Bisulfite Sequencing (RRBS) sample preparation** - Library preparation for RRBS was performed using Premium RRBS Kit (Diagenode) according to manufacturer's instructions. Briefly, 100ng of DNA was used for each sample, which was enzymatically digested, end-repaired and ligated with an adaptor. Subsequently, 8 samples with different adaptors were pooled together and subjected to bisulfite treatment. After purification steps following bisulfite conversion, the pooled DNA was amplified with 9-14 cycles of PCR and then cleaned up with Ampure XP beads. Quantification of libraries were performed using real-time qPCR (Thermo Fisher). Libraries were sequenced single end 101bp on HiSeq2500 (Illumina).

## Sequencing Alignments, QC, and Fingerprinting

**DNA alignments** - DNA alignments for whole genome (WGS) and exome sequencing was done using bwa-mem (version 0.7.15-r1140) (Fleshner and Chernett, 1997) with -M -t 12 argument and against CanFam3.1 reference genome from UCSC, <https://genome.ucsc.edu/cgi-bin/hgGateway?db=canFam3> (md5: 112bc809596d22c896d7e9bcb6e68ede6). For each sample, fastq files were aligned per read group and then merged using Picard tools (v2.18.0, <http://broadinstitute.github.io/picard>) *SortSam* command to make an interim bam file. Final, analysis-ready bam file per sample – tumor and normal bam, if available – was created by series of steps following best practices guidelines from GATK4 (version 4.0.8.1) (DePristo et al., 2011), namely *MarkDuplicates*, *Indel Realignment*, and *Base Quality Score Recalibration* (BQSR). Alignment QC metrics were calculated using GATK4 *DepthOfCoverage* (for WGS) and *CollectHSMetrics* (for exome) as well as Qualimap (version 2.2.1) (Okonechnikov et al., 2016) *bamqc* for merged bam files. Coverage statistics were also based on regions of interest (ROIs) which consisted of exonic region-level annotations for biotypes: protein-coding gene, microRNA, lincRNA, and pseudogene from Ensembl gene annotations for canine genome (v91 and higher). We flagged samples as failed QC if merged bam file has a genome-wide coverage of < 30% or > 75% of ROIs have 30% or lesser coverage. Accordingly, three samples (of three cases) failed QC step and they were removed from all analyses. Note that 83 cases in patient tissues and samples section represent all cases which passed QC at WGS, exome, RNA-seq, and methylation level data pre-processing. **RNA alignments** - Raw fastq files from paired-end RNA-seq assay for 40 tumor samples and 3 matched normal tissue samples were first preprocessed through *fastp* (version 0.19.5) (Chen et al., 2018) to perform read-based quality pruning, including adapter trimming. Resulting fastq files were then used as input for *kallisto quant* (version 0.45.0) – a pseudoalignment based method to quantify RNA abundance at transcript-level in transcripts per million (TPM) counts format. We then used *sleuth* R package (version 0.30.0) (Pimentel et al., 2017) to output model-based, gene-level normalized TPM matrix which was also corrected for potential batch effects due to RNA-seq data derived from two sequencing core facilities and tissue archival (snap-frozen vs FFPE). Detailed workflow, including command-line parameters for model fitting are in the software code repository (See [Data and Code Availability](#)). **RRBS alignments** - Raw fastq files from RRBS assay for 45 tumor samples were processed through FastQC (version 0.11.7, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and Trim Galore (version 0.5.0, <https://github.com/FelixKrueger/TrimGalore>) for quality control, filtering low quality base calls, and adapter trimming. Trimmed reads were then mapped to a bisulfite converted reference genome (canFam3.1, obtained from Ensembl release 85) using the Bismark Bisulfite Mapper (v0.19.1) with the Bowtie2 short read aligner (v2.2.3) (Krueger and Andrews, 2011), allowing for one non-bisulfite mismatch per read. Cytosine methylation calls were made for the mapped reads using the Bismark methylation extractor (version 0.19). The resulting methylation values were obtained as  $\beta$ -values, calculated as the ratio of methylated to total reads at a given CpG site. **DNA fingerprinting** - DNA fingerprinting for each of WGS and exome tumor-normal and tumor-only bam files was done using *NGSCheckMate* tool (version 1.0.0) (Lee et al., 2017). Germline snps in protein-coding regions was used as a variant reference panel to allow simultaneous fingerprinting of WGS and exome libraries. *NGSCheckMate* does sample pairing QC based on shared germline variants found in samples (tumor and normal tissue from the same patient) and also model difference between samples (or libraries) based on sequencing depth-dependent variation in allelic fraction of reference variants. Fingerprint results for WGS and exome samples from 81 canine glioma did not show mixture of tumor-normal or cross-patient sample contamination (See [Figure S1F](#)).

## Somatic Variant Calling

Somatic variant calls were called on the merged whole genome and exome bam files using three callers: GATK4 (version 4.0.8.1) (McKenna et al., 2010) Mutect2 (Cibulskis et al., 2013), VarScan2 (version 2.4.2), and LoFreq (version 2.1.3.1) (Wilm et al., 2012). Matching and fingerprint validated WGS and exome files per sample were merged using Picard tools (v2.18.0, <http://broadinstitute.github.io/picard>), *MergeSamFiles* command. Three somatic callers were then run in either paired tumor – matched normal (n=67) or tumor-only (n=14) mode. Mutect2 was first run in panel-of-normals (PON) mode using matched normal samples. Resulting PON file was used for calling somatic variant calls using Mutect2 in both, paired and tumor-only mode along with options: `-germline-resources 58indiv.unifiedgenotyper recalibrated_95.5_filtered.pass_snp.fill_tags.vcf.gz -af-of-alleles-not-in-resource 0.008621`. Tumor-only Mutect2 mode was run using default arguments and paired Mutect2 calls had following arguments: `-initial-tumor-lod 2.0 -normal-lod 2.2 -tumor-lod-to-emit 3.0 -pcr-indel-model CONSERVATIVE`. Throughout the process of using GATK4 based tools, including Mutect2, we followed best practices guidelines (DePristo et al., 2011) where practical for canine genome, e.g., in contrast to human genome, population level resources are limited for canine genome. VarScan2 paired mode was run with a command: *somatic* and arguments: `-min-coverage 8 -min-coverage-normal 6 -min-coverage-tumor 8 -min-reads2 2 -min-avg-qual 15 -min-var-freq 0.08 -min-freq-for-hom 0.75 -tumor-purity 1.0 -strand-filter 1 -somatic-p-value 0.05 -output-vcf 1`. VarScan2 tumor-only mode was run using command: *mpileup2cns* and arguments: `-min-coverage 8 -min-reads2 2 -min-avg-qual 15 -min-var-freq 0.08 -min-freq-for-hom 0.75 -strand-filter 1 -p-value 0.05 -variants-output-vcf 1`. LoFreq paired mode was run using command: *somatic* and arguments: `-threads 4 -call-indels -min-cov 7 -verbose` and tumor-only mode was run using command: *call* and arguments: `-call-indels -sig 0.05 -min-cov 7 -verbose -s`. Resulting raw somatic calls - single nucleotide variants (SNV) and small insertions and/or deletions (Indels) - from three callers were then subject to filtering based on caller-specific filters and hard filters. Briefly, Mutect2 calls were subject to extensive filtering based on germline risk, artifacts arising due to sequencing platforms, tissue archival (FFPE), repeat regions, etc. See [Data and Code Availability](#) and <https://software.broadinstitute.org/gatk/documentation/article?id=11136> for detailed parameters. VarScan2 somatic filters were applied as per developer's guidelines (Koboldt et al., 2013). Hard filters were based upon filtering out variants present in dbSNP and PONs created via GATK4 Mutect2. Filtered somatic calls from three callers (in VCF version 4.2 format) were then subject to consensus somatic calls using SomaticSeq (version 3.1.0) (Fang et al., 2015) in majority voting mode with priority given to Mutect2 filtered (PASS) calls followed by consensus voting based on calls present in VarScan2 and LoFreq filtered calls. Resulting consensus VCF file for 81 cases were finally converted to Variant Effect Predictor (VEP version 91) (McLaren et al., 2016) annotated vcfs and Mutation Annotation Format (MAF, [https://docs.gdc.cancer.gov/Data/File\\_Formats/MAF\\_Format](https://docs.gdc.cancer.gov/Data/File_Formats/MAF_Format)) using *vcf2maf* utility (<https://github.com/mskcc/vcf2maf>). Annotated VCFs and MAFs were used for all of downstream analyses.

## Significantly Mutated Genes (SMGs) Analysis

SMG analysis in canine gliomas (Figures 1A, 1C, and 2A) with paired tumor-normal samples (n=57) was performed using dNdS (Martincorena et al., 2017) and MuSiC2 (version 0.2) (Dees et al., 2012). We excluded tumor-only cases for being conservative in SMG analysis and minimize false-positives. Also, MuSiC2 required matched normal tissue required matched normal tissue for SMG analysis. Detailed parameters for SMG analysis are in the software code repository (See [Data and Code Availability](#)). Detailed output from both methods are in [Table S2](#).

## Cancer Hallmark Analysis

Cancer hallmarks were defined according to published ten hallmarks (Hanahan and Weinberg, 2011) and one additional hallmark, i.e. epigenetic (Imielinski et al., 2012). A pool of 268 known glioma (Ceccarelli et al., 2016; Mackay et al., 2017) and pan-cancer driver genes (Gröbner et al., 2018; Bailey et al., 2018) were mapped to hallmarks following a previously published computer-assisted manual curation method (Table S4) (lorio et al., 2018). Based on WHO molecular classification of brain tumors, somatic SNV and copy-number data from patients with human adult gliomas (AG) (Ceccarelli et al., 2016) were stratified into IDHwt (n=373), IDHmut-codel (n=169) and IDHmut-noncodel (n=268) subgroups while corresponding data from patients with human pediatric glioma (PG) (Mackay et al., 2017) were subgrouped based on mutations in Histone H3 gene, namely H3mut (n=200) and high-grade H3wt (n=126). For canine patients with glioma (CG), we used somatic mutations and copy-number calls from 67 cases with paired tumor-normal samples. For each of the six cohorts coding mutations were mapped to eleven hallmarks and coverage adjusted relative proportions of patients harboring an alteration in a given hallmark were calculated. For comparisons between cohorts a two-sided Fisher's exact test was applied (Table S4).

## Quantifying Somatic Mutation Rates

Somatic mutations (SNVs and Indels) rate was estimated within coding genes and adjusted based on relative per-base coverage with minimum coverage of 30X in coding regions (Figure 1D). Coding mutation rates for human pediatric (n=961) and adult cancers (n=3,800, includes 811 adult gliomas) were taken from published studies (Ma et al., 2018; Ceccarelli et al., 2016; Gröbner et al., 2018).

## Somatic Copy Number Segmentation

Somatic copy-numbers were called for paired tumor-normal cases (n=67) using HMMCopy tool (version 1.22.0) using author's recommendations. In brief, GC counts and mappability files for CanFam3.1 genome were generated with 1000 bp window size. Read counts for each of tumor and normal bam files were generated using 1000 bp window size. Resulting count, mappability



# Mutational signatures of genotoxic cancer therapies

CellPress

and count files were feed into HMMCopy algorithm (<http://bioconductor.org/packages/release/bioc/html/HMMcopy.html>) and segmentations were called using Viterbi algorithm. Segmented copy number calls were used to generate Integrated Genome Viewer (IGV) copy-number plots and GISTIC2 (version 2.0.22) based somatic copy number significance (Mermel et al., 2011), including calling gene-level deep deletions, loss-of-heterozygosity (LOH), and amplifications (Figure 2A) as well as inferring aneuploidy metrics (Figures 2B and 2C). Segmented copy number for pediatric gliomas (n=53) were called by using cloud-based TitanCNA workflow ([https://dxapp.verhaaklab.com/dnanexus\\_ngsapp](https://dxapp.verhaaklab.com/dnanexus_ngsapp)). Segmented copy number for adult gliomas were derived from SNP6 based platform from the TCGA Broad Firehose platform (version stddata\_2016\_01\_28) with following download url: [http://gdac.broadinstitute.org/runs/stddata\\_2016\\_01\\_28/data/GBM/20160128/gdac.broadinstitute.org\\_GBM.Merge\\_snp\\_genome\\_wide\\_snp\\_6\\_broad\\_mit\\_edu\\_Level\\_3\\_segmented\\_scna\\_minus\\_germline\\_cnv\\_hg19\\_seg.Level\\_3.2016012800.0.0.tar.gz](http://gdac.broadinstitute.org/runs/stddata_2016_01_28/data/GBM/20160128/gdac.broadinstitute.org_GBM.Merge_snp_genome_wide_snp_6_broad_mit_edu_Level_3_segmented_scna_minus_germline_cnv_hg19_seg.Level_3.2016012800.0.0.tar.gz) and [http://gdac.broadinstitute.org/runs/stddata\\_2016\\_01\\_28/data/LGG/20160128/gdac.broadinstitute.org\\_LGG.Merge\\_snp\\_genome\\_wide\\_snp\\_6\\_broad\\_mit\\_edu\\_Level\\_3\\_segmented\\_scna\\_minus\\_germline\\_cnv\\_hg19\\_seg.Level\\_3.2016012800.0.0.tar.gz](http://gdac.broadinstitute.org/runs/stddata_2016_01_28/data/LGG/20160128/gdac.broadinstitute.org_LGG.Merge_snp_genome_wide_snp_6_broad_mit_edu_Level_3_segmented_scna_minus_germline_cnv_hg19_seg.Level_3.2016012800.0.0.tar.gz) Only primary tumor cases from TCGA GBM (n=577) and TCGA LGG (n=513) cohort were used for downstream analyses, i.e., pathway analysis (Figure 1C) and aneuploidy metrics (Figures 2B–2D).

## Allele Specific Copy-Number Analysis

We derived allele-specific copy numbers and copy-number based clonality inference (including purity and ploidy estimates) using TitanCNA algorithm (version 1.19.1) (Ha et al., 2014). *Snakemake* (version 5.2.1) based workflow (Koster and Rahmann, 2018) was implemented using default arguments and genome-specific germline dbSNP resource for WGS paired tumor-normal samples from 67 canine patients. For pediatric gliomas (n=53) and adult gbms with WGS data (n=23), allele-specific copy-number calls were used from TitanCNA workflow. Allele-specific copy-numbers were used for mutational signature and molecular timing analysis (Figure 3).

## Aneuploidy Metrics

The simplest 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 (Taylor et al., 2018). 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 38 (total number of autosomal chromosome arms – given all of canine chromosomes are either acrocentric or telocentric).

## Clustering Shared Syntenic Regions

Shared syntenic regions between CanFam3.1 and hg19 reference genome were downloaded from Ensembl BioMart (version 94) database using orthologous mapped Ensembl gene ids. Arm-level synteny was based on arm-level aneuploidy scores of shared syntenic regions in the respective, canine and human genomes. Hierarchical clustering of proportion of patients per molecular subtype having syntenic arm-level aneuploidy was then carried out for each of canine, human pediatric and adult cohort (Figure 2C).

## Estimating Intra-tumoral Heterogeneity

We estimated patient-level ITH based on whole-genome derived subclonal structure and number of somatic variants in each of these subclones. Subclonal structure and cellular prevalence or cancer cell fraction of each tumor subclone (and underlying somatic variants) was derived using TITAN allele-specific copy number calls. Since accuracy of inferred subclonal structure depends largely on sequencing read depth and number of somatic variants per inferred subclone, we limited estimation of subclonal structure for maximum five subclones per patient given a minimum sequencing read depth of 30X for whole genome data we had across all three cohorts. Shannon entropy was then calculated using entropy function in the R package: entropy by taking number of somatic variants per subclone per patient as a vector. A resulting Shannon entropy value was used to plot figures along with cancer cell fraction and number of clones derived per patient. We acknowledge that our estimation of ITH and resolving subclonal structure can be improved with higher depth of sequencing (100X or more) to detect subclonal structures (number of clones) (Deshwar et al., 2015).

## Mutational Signature Analysis

Mutational signature analysis was performed in two-parts. First, de-novo signatures (Figure 3B) were constructed for canine (n=81), human pediatric (n=53) and adult cohort (n=23) using somatic snvs from whole-genome data. Signatures were constructed using non-negative matrix factorization (*nmf* R package, version 0.21.0) with brunet approach and 100 iterations with expected range of signatures between 2 to 10. Optimal signatures were then selected using *nmfEstimateRank* function to match number of de-novo signatures (clusters) – 1 where inflection point for cophenetic correlation coefficient was observed. Accordingly, three de-novo signatures were found in canine and human pediatric gliomas while five in adult glioblastoma cohort. In the second part, known human mutational signatures from COSMIC (v2, n=30) and published pediatric cancer signature from two studies, T1 to T11 (Ma et al., 2018) and P1 (Gröbner et al., 2018) were pooled together and used to deconvolute (*MutationalPattern* R package, version 1.6.2) mutational

e5 Cancer Cell 37, 243–257.e1–e7, February 10, 2020



trinucleotide context (n=96) from somatic snvs in each of three cohorts. Somatic ultra-hypermutation cases from pediatric (n=3) and adult cohort (n=1) were excluded from signature analysis. Cosine similarities of known signatures with de-novo signatures was then calculated and clustered using hierarchical clustering (Figure 3B). Absolute and relative contribution of known signatures per sample was then quantified using *fit\_to\_signatures* function which finds the linear combination of signatures that closely resembles 96 context based mutational matrix by solving the nonnegative least-squares constraints problem. We then selected top contributing signatures per cohort based on signatures which contributed per sample to higher than 3<sup>rd</sup> quartile of median value of each signature's contribution (rowMedian) per cohort (Figure S3A). Top contributing signatures were further calculated using *outlier profiling* on canine patients showing highest mutational load (>3<sup>rd</sup> quartile of median coding mutation rate per megabase) and plotted in Figure 3A. Outlier sample detection was done using *car::outlierTest* function in R to label true outliers from entire cohort (2 cases) while correcting for confounding effects due to type of tissue archival (snap-frozen vs ffpe) and analysis type (tumor-matched normal versus tumor-only somatic variant calling). This was followed by second run of outlier by first excluding true outliers (2 cases), and then labelling outliers (six cases) based on chi-squared statistics using *outliers::scores* function in R. Signatures contributing to driver mutations (Figure 3C) were calculated based on first getting relative proportion of trinucleotide context per snv and then finding known signatures with maximum value for the same trinucleotide context. Known signatures were combined to a single group where they are shown in literature as potential underlying process, e.g., aging group is associated with COSMIC signature 1 and 5, and show significant cosine similarity (> 0.9) with pediatric signatures T1 and T4, respectively. Table S5 provides mapping between signature and known/proposed mechanisms, if any.

### Molecular Timing Analysis and Natural History of Tumors

Probabilistic estimation of relative timing of driver mutations (among 79 observed somatic snvs in cancer driver genes) was based on existing methods (Gerstung et al., 2017; Jolly and Van Loo, 2018) with several steps carried out using Palimpsest R package (version 1.0.0; <https://github.com/FunGeST/Palimpsest>) (Shinde et al., 2018) and custom R scripts based on published approach (McGranahan et al., 2015): First step involved categorizing somatic drivers into clonal vs subclonal events using estimated cancer cell fraction (CCF) which is estimated fraction of cancer cells with a somatic snv. CCF per somatic snv was a product of variant allelic fraction (VAF) of a somatic snv, adjusted by local copy number of gene locus and whole tumor sample (ploidy) as well as purity estimate (tumor cell content) inferred from TitanCNA algorithm (Detailed under copy number estimation section above). A clonal (early) vs subclonal (late) mutation was then classified based on upper boundary of CCF was above 0.95 (clonal) or not (subclonal). Second, we timed copy number gain and copy-neutral LOH regions based on VAF of somatic snvs in these copy regions, i.e., early mutations prior to copy gain will have higher VAF relative to VAF of late mutations after copy gain. Third, we ordered mutations in four sequential categories: early clonal, early subclonal, late clonal, and late subclonal. We note here that early subclonal and late clonal categories are result of underlying parallel and/or convergent evolution of multiple clones (Venkatesan and Swanton, 2016) and/or a technical limitation (given ~60X depth of merged bam files and lack of spatial sequencing data) in resolving polyclonal structure of a tumor sample (Deshwar et al., 2015). We then tally frequency of each of these four categories per somatic driver mutation and get the average frequency of each category per driver mutation at cohort (canine, pediatric, adult) level. These average frequencies are converted to winning tables, similar to sports statistics where each driver mutation competes with remaining driver mutations with winning being an early somatic event based on order of events using clonality (Jolly and Van Loo, 2018) (step 3). Finally, a winning table is then passed to Bradley-Terry model (*BradleyTerryScalable* R package, version 0.1.0.9000) to estimate winning probability (driver event being an early event) based on a Bayesian maximum a posteriori probability (MAP) estimate. Resulting winning probability per driver mutation is subtracted from 1 to plot multiple density plots (*ggridges* R package, version: 0.5.1.9000) with X-axis now showing a probability of event being a late event (Figure 3D). We note that density plots are based on kernel density estimates and thus, may extend their tails (probability distribution) beyond 1 or less than zero (<https://serialmentor.com/dataviz/histograms-density-plots.html>).

### Class Prediction Using Methylation Data

To compare the methylation patterns of human and canine glioma, the LIBLINEAR library was used to fit an L2-regularized logistic regression classifier. Model training and validation was performed on the human glioma samples and normal controls in the GSE109381 dataset (Capper et al., 2018), with the methylation status of CpGs located in regions of the human genome orthologous to canine CpG islands used to predict DNA methylation-based subtypes of glioma. The methylation categories designated as regression outcome variables were derived from the World Health Organization classification of gliomas: IDH-wild-type adult glioma, IDH-mutant, 1p/19q-intact adult glioma, IDH-mutant, 1p/19q-codeleted adult glioma, adult normal control, pediatric glioma, and pediatric normal control. After model fitting, the logistic regression classifier was applied to the canine samples, using the  $\beta$ -values of CpGs orthologous to the selected 11,495 Illumina 450K probes as input data. For classifier CpG sites in the canine samples with no methylation observations,  $\beta$ -values were predicted using the DNA module of the DeepCpG algorithm, a deep learning algorithm that predicts methylation state based on local DNA sequence context (Angermueller et al., 2017). The logistic regression classifier outputs the probability that a sample matches a given methylation category. Category probabilities were calculated for the canine samples, and these probabilities were compared with sample age, anatomical location, tumor grade, tumor purity, and mutation rate (Figure 4).

## Immunohistochemistry

### Staining

Hematoxylin & Eosin staining was used to classify glioma grade and lineage. The immunohistochemistry panel included those antibodies that have been documented to work in canine tissues and include myeloid microglia/macrophages (IBA1), monocytes (CD14) and their M2 skew subtype (CD163), and lymphoid T cells (CD3) and B cells (CD79a). Slides with 5µm sections, were deparaffinized and rehydrated in a dry incubator (60°C for 1 hour), xylene, and histological grade ethanol. Antigen retrieval was performed using citrate buffer and a pressure cooker at 120°C for 12 minutes. Quenching for endogenous peroxidase was performed with 3% H<sub>2</sub>O<sub>2</sub> for 15 minutes at room temperature. Non-specific binding was minimized using ready-to-use protein blocker (Dako) applied for 15 minutes before the application of the primary antibody overnight at 4°C. All the washing was done using 1x T-PBS buffer mixed with 0.1% Tween 20. The biotinylated secondary antibody was applied for 30 minutes at room temperature followed by three washes with buffer for 5 minutes each. Color development was performed using the DAKO DAB kit and color change was monitored until an appropriate detectable level was achieved (10-60 sec depending on the antibody). Slides were counterstained with hematoxylin (25 seconds) and bluing buffer, then rehydrated and cover-slipped with long lasting mounting solution. The immunohistochemistry quantification were done blindly relative to the tumor pathology. Scanning and tissue segmentation - Scanning and analysis were performed using the PerkinElmer Vectra Automated Quantitative Pathology Imaging System and the inForm Cell Analysis software (ver 2.4). Slides were scanned twice on low- and high-power fields as follows: the first scan was of the whole slide on low power field (10x) for manual tissue segmentation to identify three tumor regions/categories as necrotic center, tumor and invasive edge under the neuropathologist's supervision/direction. For each region, every fourth field was imaged (25%) on high-power field (20x) and resulted in 21 to 274 fields per slide, which varies according to the size of the tissue and presence or absence of necrosis. For the training set, heterogeneous fields were randomly selected to include tissue, non-tissue and damaged areas. Hematoxylin and DAB was used to identify the nuclei. Positive and negative cells were distinguished visually and three optical densities (OD) thresholds were set accordingly. The thresholds allowed 4-bin (0 = negative, +1 = weak positive, +2 = intermediate, +3 = strong positive) sorting of cells depending on the positivity and its intensity. The intermediate positivity threshold was calculated as the midpoint after setting the lower and higher threshold. The algorithm of the training set was applied for all the high-power fields captured. The results were inspected and the nonspecific and defective fields were removed before compiling the dataset. The same process was applied for all seven markers (Figure 5).

### CIBERSORT Based Expression Analysis

Processed RNA-seq expression matrices from canine (n=40; 25 HGG, 14 LGG, 1 unknown grade), adult (n=703; 529 LGG, 174 GBM), and pediatric glioma (n=92; 42 LGG, 50 HGG) were each run as separate jobs into the CIBERSORT webserver (<https://cibersort.stanford.edu>) and processed in relative mode using the following parameters: Signature Genes: LM22 CIBERSORT default, Permutations run: 100, Quantile normalization disabled (Newman et al., 2015). The resulting cellular fraction tables were then collapsed from 22 cell types into 11 based on lineage, using groupings from a prior publication (Gentles et al., 2015).

### QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed using R version 3.6.1. Statistical details for analyses are described in the respective sub-section under the [Method Details](#) section above and summarized in figure legends. p value of <0.05 were considered statistically significant. No statistical methods were used to predetermine study sample size.

### DATA AND CODE AVAILABILITY

Sequencing data generated during this study is available in the Binary Alignment Map (BAM) format at the NCBI SRA database with the BioProject accession ID PRJNA579792 [URL: <https://dataview.ncbi.nlm.nih.gov/object/PRJNA579792>]. Software code used to generate figures is available at <https://github.com/TheJacksonLaboratory/canineglioma> and documented at the URL, <https://canineglioma.verhaaklab.com>.

# Chapter 7

## Discussion

This chapter serves to summarize, discuss, and contextualize the previous chapters. My analyses specifically followed the hypothesis that longitudinal and molecular characterization of cancer may help uncover the clonal trajectories of cancer, understand the impact of treatment on the cancer genome and reveal potential therapeutic avenues within a precision oncology approach.

In **Chapter 2**, I provided an overview of the field of molecular characterization of cancer, the high unmet clinical need in glioma patient care and the potential opportunities for longitudinal analysis of cancer to contribute to an improved understanding of therapeutic effects and resistance mechanisms. Moreover, I highlight the gaps in knowledge on the genomic effects of radiotherapy and alkylating agent – based chemotherapy, which represent the clinical gold standard in glioma treatment.

In **Chapter 3**, I present results derived from my participation in the analysis of matched initial (pre-treatment) and recurrent (post-treatment) glioma samples sequenced at the WGS/WES and RNA-seq levels as part of the GLASS consortium. Overall, this study revealed that most genomic alterations were shared between initial and recurrent gliomas, with little evidence of directed and predictable evolutionary routes. These results are consistent with a previous study of longitudinally analyzed IDHwt gliomas, suggesting that the most aggressive glioma subtype, GBM, formed its genomic repertoire prior to diagnosis and that treatment had no significant impact on the clonal evolution of the analyzed tumors samples<sup>66</sup>. However, IDHmut glioma samples classified as low-grade gliomas according to the 2016 WHO classification<sup>26</sup>, showed patterns of genomic alterations associated with an increase in grade. I quantified the proportion of the alkylating agent-associated hypermutator phenotype, indicating the highest susceptibility for the IDHmut-noncode1 group. Mutational signature analysis further revealed a dominance of the single base substitution (SBS) 11 signature, which was confirmed in a subsequent independent analysis<sup>67</sup>. Moreover, IDHmut gliomas were characterized by a significant increase in genome-wide and arm-level aneuploidy, which was typically preceded by the acquisition of *CDKN2A*

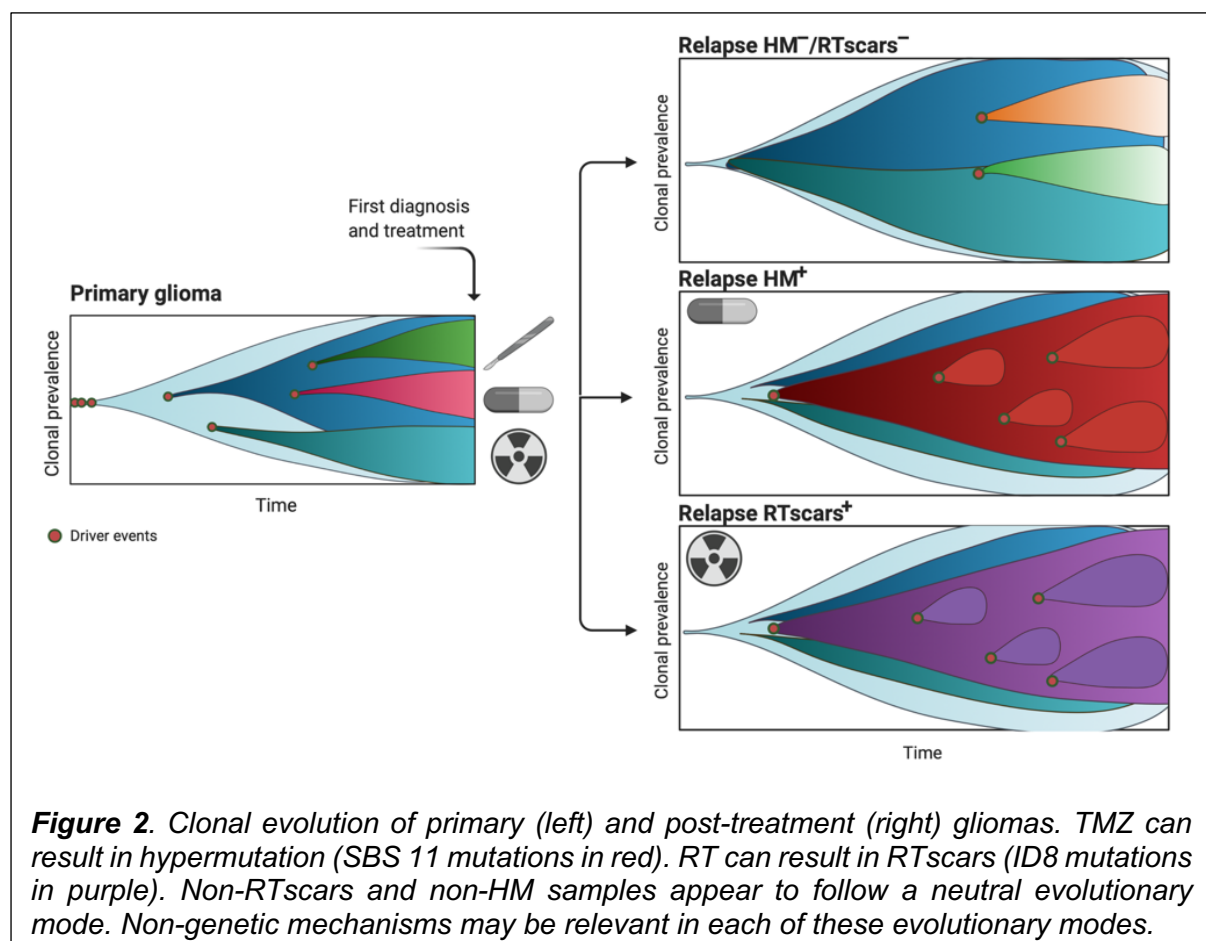
## Mutational signatures of genotoxic cancer therapies

homozygous deletions in recurrent samples. These features appeared to be clinically relevant as they were associated with poorer overall survival.

In **Chapter 4**, I addressed the question of whether RT has a specific imprint on cancer genomes and analyzed a pan-cancer metastatic dataset in addition to the GLASS dataset. While RT leads to DNA damage, e.g., in the form of DSBs, repair pathways can be activated. However, previous literature has attributed distinct repair pathways to RT-induced DNA damage, and studies performed with low samples sizes yielded conflicting results. In my analyses, I identified a significant increase in small deletions that could be specifically attributed to RT. These small deletions had distinct genomic features, including a length of 5-15 bp, a random genomic distribution across the genome and a corresponding independence of intrinsically mutagenic genomic sites, as well as a lack of microhomology at breakpoint sequences. These results provided important information about the repair pathways associated with RT-induced DNA damage by nominating canonical non-homologous end-joining (cNHEJ) as the preferred pathway for repairing RT-induced DSBs. Importantly, I was able to link this genomic outcome to a specific indel mutational signature, i.e. ID8, and provide a quantitative measure with potential translational value. Indeed, a high burden of RT-associated deletion signature was associated with worse survival outcomes. Furthermore, we observed that RT was associated with an increase in larger structural variants, particularly for inversions and large deletions. It is reasonable to speculate that these, similar to small deletions, may result from erroneous repair of RT-induced DSBs. These changes were accompanied by an increase in aneuploidy, which was particularly driven by aneuploidy arm level loss events. These results suggest that RT may induce chromosome segregation errors during mitosis. Importantly, focal deletions on chromosome 9p21 leading to homozygous deletions in *CDKN2A* are rare in primary IDHmut gliomas but are of great clinical importance as negative prognostic markers, as recently recognized in the 2021 update of the WHO classification for central nervous system tumors (CNS5)<sup>68</sup>. My observations that acquisition of *CDKN2A* homozygous deletions were enriched exclusively in recurrent IDHmut glioma patients who had received RT suggests that RT may be mechanistically involved in the emergence of this poor prognostic biomarker and underscores the value of performing NGS analysis of post-treatment, recurrent tumors.

## Mutational signatures of genotoxic cancer therapies

In **Chapter 5**, I participated in the follow-up analyses of the GLASS consortium, in which we analyzed DNA and RNA sequencing data from 304 patients with IDHmut and IDHwt glioma. We integrated genomic information with gene expression data that allowed us to infer cancer cellular states as well as the decomposition of the bulk tumor into malignant cells and cells of the microenvironment. In particular, I was able to detect an increase in the proliferative stem-like cell state in IDHwt and IDHmut gliomas. This non-genetic longitudinal change was particularly driven by known genomic markers affecting survival, i.e. TMZ-associated hypermutation and RT-associated *CDKN2A* homozygous deletions. The genomic resistance markers appear to correspond with an increasingly malignant and proliferative phenotype, which is underscored by a survival analysis indicating significantly worse survival for patients with a longitudinal increase in the proliferative stem like cell state. **Chapter 5**, in combination with **Chapters 3-4**, provides an integrative longitudinal model of genomic and non-genetic evolution of gliomas in response to treatment, depicting three modes of evolution (Fig 2).



## **Mutational signatures of genotoxic cancer therapies**

In **Chapter 6**, I performed a molecular analysis of 83 spontaneous canine gliomas and compared the molecular profiles with those of human adult and pediatric gliomas. Specifically, I found the genomic alteration pathways to be largely conserved between species, with canine gliomas displaying a high similarity to human histone H3-wildtype pediatric gliomas. The translational pipeline in gliomas has been largely unsuccessful in the past two decades, as exemplified by negative trial results for immunotherapies<sup>69,70</sup>. With an immunocompetent host environment, preclinical trials in canine glioma patients could provide novel translational opportunities, while simultaneously providing treatment options for dogs with gliomas.

## **Mutational signatures of cancer therapies in the context of precision oncology**

Mutational signatures are a powerful tool for analyzing the underlying mechanistic patterns of mutations at the single base substitution (SBS), double base substitution (DBS) and insertion-deletion (ID) levels. In my work, I identified a strong enrichment of the SBS11 signature in TMZ-associated hypermutant samples and ID8 signature in samples with an RT-associated deletion signature. While I did not observe an impact of HM on overall survival, survival after recurrence was significantly shorter in HM patients compared to patients without HM, which was further confirmed in an independent cohort<sup>67</sup>. The RT-associated deletion signature resulted in worse overall survival in patients that were classified in the highest tertile based on newly acquired deletion burden after RT. More sophisticated clinical analysis of the RT-associated deletion signature revealed no survival differences for the surgical interval time point (survival between primary tumor and recurrence), but significantly worse post-recurrence survival. Taken together, SBS11 (marker for hypermutation) and ID8 (marker for RT-associated deletion signature) marked tumors that were resistant towards TMZ and RT, respectively, suggesting that continuation of the same treatment modality would not provide further clinical benefit.

Based on these considerations, the identified markers could be translated as potential prognostic biomarkers into the clinical setting and help in therapeutic decision making in the relapse setting where there is no evidence for a standardized regimen in gliomas<sup>71</sup>. In addition to this prognostic value, the SBS11 and ID8 signatures may have predictive value that needs to be tested in prospective clinical trials<sup>72</sup>. I hypothesize that glioma patients harboring a high burden of the hypermutation-

## Mutational signatures of genotoxic cancer therapies

associated SBS11 signature are unlikely to benefit from further TMZ treatment. Similarly, cancer patients with a high burden of the RT-associated ID8 signature are unlikely to benefit from further RT. Instead, alternative treatment strategies could be considered in these instances. For example, an MMR-independent compound that has been shown to be effective in in TMZ-resistant gliomas should be considered for cases with hypermutation<sup>73</sup>. Since the underlying mutational mechanism for the ID8 signature is the erroneous repair of RT-induced DSBs via cNHEJ, compounds targeting the DNA damage response pathway should be considered as a combination therapy with RT<sup>74</sup>. In this context, ID8 could serve as a biomarker for differential susceptibility.

Having provided valuable insights into the biological processes underlying somatic mutagenesis, mutational signatures are gaining translational and clinical importance<sup>75</sup>. Cancers with homologous recombination deficiency (HRD) have been shown to be predictive of increased response to platinum-based treatment<sup>76</sup> and targeted therapy with poly(ADP)-ribose polymerase (PARP) – inhibitors in a synthetic lethality approach<sup>77</sup>. Novel algorithms based on genomic signatures have high predictive value in terms of sensitivity and specificity in predicting HRD. The tool ‘HRDetect’ predicted *BRCA1/BRCA2* deficiency with a sensitivity of 98.7% (AUC: 0.98)<sup>78</sup> and was successfully validated in a real-world cohort<sup>79</sup>, supporting its translational and clinical value. The Classifier of HOmologous Recombination Deficiency (CHORD) algorithm, developed by the HMF group and was also used for my analyses in **Chapter 4**, identified bi-allelic inactivation of *BRCA1*, *BRCA2*, *RAD51C* or *PALB2* as common mechanisms for HRD and is another example for a readily available tool for diagnostic and therapy decision purposes in the clinic<sup>80</sup>.

Mismatch repair (MMR) deficiency is associated with an increased tumor mutational burden (TMB) due to increased numbers of mutations at short tandem repeats<sup>81</sup>. Importantly, patients with MMR deficient cancers respond remarkably well to immunotherapy with immune checkpoint inhibitors (ICIs)<sup>82-84</sup>, which led to the first tumor-agnostic approval of the ICI pembrolizumab for any solid cancer type with MMR deficiency by the US Food and Drug Administration (FDA) in 2017<sup>85</sup>. One of the established clinical biomarkers for predicting the efficacy of ICI is TMB<sup>86</sup>, which reflects the number of mutations per megabase (mut/Mb). It is increasingly recognized that patients with a TMB of more than 10 mut/Mb are the most promising target group for



## **Mutational signatures of genotoxic cancer therapies**

a successful response to ICI, but this biomarker cutoff leading to a binary classification has been the subject of controversy<sup>87</sup>. Improved detection of MMR deficient cancer samples has recently been demonstrated by using MMR deficiency – associated mutational signatures in the ‘MMRDetect’ algorithm, which limits false-positive and false-negative mutational calls<sup>48</sup>. Furthermore, a comprehensive meta-analysis incorporating genomic and transcriptomic sequencing data revealed that, in addition to the quantity of mutations measured as TMB, the quality of mutations (clonal vs. subclonal, SBS vs. ID etc.) played a major role in the response to ICIs<sup>88</sup>. A multivariable predictor incorporating multiple levels of biomarkers outperformed the use of TMB alone for predicting response to ICIs<sup>88</sup>.

As WGS becomes increasingly implemented in clinical practice due to decreasing sequencing costs and improved turnaround times of sequencing reports<sup>89-93</sup>, it is presumably only a matter of time before mutational signatures become an important component of clinical decision making.

## **Longitudinal phenotypic plasticity in response to therapy**

In **Chapters 3-4**, I comprehensively analyzed the genomic clonal evolution of cancer in response to treatment. In addition to genomic evolution, gliomas are characterized by marked transcriptional intratumoral heterogeneity and phenotypic plasticity<sup>94-100</sup>, which may increase their fitness advantage. These features are expected to further help bypass the evolutionary bottleneck caused by treatment. Cancer cells can activate a variety of non-genetic adaptive mechanisms that enable them to survive cancer treatment<sup>101</sup>. One converging mode of adaptation to therapeutic stress in various cancers appears to be the acquisition of cancer stemness signatures that impact tumor progression and clinical outcomes in patients. For example, B-cell lymphomas appear to enter a treatment-induced state of cancer cell senescence after chemotherapy<sup>102</sup>. It has been reported that a cell state switch involving upregulation of stem cell properties through activation of Wnt signaling enables cancer cells to have proliferative potential and escape cell-cycle blockade<sup>102</sup>. Consequently, intrinsic plasticity of cancer cells toward a stemness signature was enriched in relapsed B-cell lymphomas, indicating an adaptive growth advantage<sup>102</sup>. Strikingly similar mechanisms of treatment-induced stemness signature activation associated with drug-resistance and tumor recurrence have been observed in acute

## Mutational signatures of genotoxic cancer therapies

myeloid leukemia (AML)<sup>103-105</sup>, colorectal cancer<sup>106</sup>, breast cancer<sup>107</sup>, melanoma<sup>108</sup>, and basal cell carcinoma<sup>109</sup>, among others.

This consistent similarity in both hematologic and solid tumors suggested that similar phenotypic changes related to the stem cell signature might also play an important role in resistance to RT and TMZ in gliomas. Indeed, my analyses presented in **Chapter 6** highlighted non-genetic evolutionary mechanisms, including an increase in a proliferative stemness signature that associated with an aggressive phenotype. Identification and targeting specific stemness signature pathways and genes represent a promising approach to overcome treatment failure. Recent studies have highlighted the role of epigenetic regulators in shaping distinct transcriptional cell states in gliomas<sup>99,100</sup> and I suspect that a concept targeting such regulators may be an effective strategy to increase the efficacy of TMZ and RT, which needs to be tested in prospective studies.

## Integration of genetic and non-genetic treatment effects

In this thesis, I sought to characterize the patterns of mutations associated with RT and TMZ. Additionally, I found evidence of a quantitative increase in copy number (CN) alterations and structural variants (SV) that were related to RT. However, the qualitative nature of these larger genomic alterations remains unclear. In analogy to mutational signatures, comprehensive catalogs of CN signatures<sup>110,111</sup> and SV signatures (SVs)<sup>112</sup> have recently been established. Applying these tools, which were used only for pre-treatment primary cancer tissues, to the post-treatment GLASS and HMF datasets opens the opportunity of gaining insight into the mechanistic basis of previous quantitative observations, as presented in **Chapter 4**. Moreover, correlation of CN- and SV- signatures with treatment-associated mutational signatures may provide a detailed view of the whole genome effects of therapies.

Cancer treatment is expected to modulate the interaction of cancer cells with their tumor microenvironment<sup>113</sup>. Specifically, longitudinal changes at the genomic level (including mutational, CN- and SV-signatures), could change the interplay with the immune system, potentially opening untapped opportunities for immunotherapies<sup>114,115</sup>. Chemotherapy and RT could lead to the generation of immunogenic neoantigens that could abrogate a local tumor immune response and

## **Mutational signatures of genotoxic cancer therapies**

enhance the immunogenicity of immune-cold tumors<sup>116-118</sup>. Further follow-up analyses are needed to rigorously test these hypotheses.

Ultimately, the goal of translational medicine is to provide new insights for the patient bedside and to improve patient outcomes. Genomics-based biomarkers, preferably at multiple levels using multi-omic approaches, have the potential to support clinical decision-making processes related to cancer diagnosis and prognosis, as well as prediction of treatment response and resistance. With the use of machine learning models and novel computational frameworks, the analysis of multimodal datasets is becoming increasingly feasible. In addition to the clinical and genomic features described in detail in this thesis, the use of data collected in routine clinical practice, such as radiological imaging, including computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET)<sup>119</sup>, as well as hematoxylin & eosin (H&E)-stained pathological slide images<sup>120-122</sup>, provide rich analytical outputs with clinical utility. The integration of multiple variables<sup>88</sup> and modalities<sup>123</sup> of clinical, molecular, radiological and histological imaging data into predictive models holds tremendous potential for biomarker development that will drive 'precision oncology' in the coming years and decades<sup>124</sup>.

# Chapter 8

## Appendices

### List of abbreviations

53BP1	Tumor suppressor P53-binding protein 1
a-EJ	Alternative end joining
AML	Acute myeloid leukemia
APOBEC	apolipoprotein B mRNA editing enzyme, catalytic polypeptide
bp	Base pairs
BRCA1/2	Breast cancer type 1/2 susceptibility protein
C	Cytosine
CN	Copy number
CNV	Copy number variation
c-NHEJ	Canonical/classical non-homologous end joining
CDKN2A	Cyclin-dependent kinase inhibitor 2A
CHORD	Classifier of HOMologous Recombination Deficiency
CNS	Central nervous system
CT	Connecticut
CT	Computed tomography
CtIP	Carboxy-terminal binding protein interacting protein
d	Day
DNA	Deoxyribonucleic acid
DNA-PK	DNA-dependent protein kinase
DSB	Double strand break
FDA	Food and Drug Administration
GBM	Glioblastoma
G	Guanine
GLASS	Glioma Longitudinal Analysis Consortium
Gy	Gray
H&E	Hematoxylin & eosin
HM	Hypermutation
HMF	Hartwig Medical Foundation
HR	Homologous recombination
HRD	Homologous recombination deficiency
ICI	Immune checkpoint inhibitor
ID	Insertion-deletion
IDH	Isocitrate dehydrogenase
IDHwt	<i>IDH</i> wild type
IDHmut	<i>IDH</i> mutant
IDHmut-codel	<i>IDH</i> mutant with chromosome arm 19/19q-codeletion
IDHmut-noncodel	<i>IDH</i> mutant without chromosome arm 19/19q-codeletion
i.e.	id est
indel	Insertion and deletion
Mb	Megabase

## Mutational signatures of genotoxic cancer therapies

mg	Milligram
MGMT	O(6)-methylguanine-DNA methyltransferase
MLH1	MutL protein homolog 1
MMR	Mismatch repair
MRI	Magnetic resonance imaging
MRN	MRE11-RAD50-NBS1 complex
MSH2	MutS-homolog 2
MSH6	MutS-homolog 6
mut	Mutation
m <sup>2</sup>	Square meter
NMF	Non-negative matrix factorization
NGS	Next generation sequencing
O6-MeG	O6-methylguanine
PARP	poly(ADP)-ribose polymerase
PET	Positron emission tomography
PMS2	PMS1 homolog 2
RCT	Randomized controlled trial
RNA	Ribonucleic acid
RT	Radiotherapy
SBS	Single base substitution
seq	sequencing
SSA	Single strand annealing
SSB	Single strand break
SV	Structural variant
T	Thymine
TMB	Tumor mutational burden
TMZ	Temozolomide
USA	United States of America
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World Health Organization

## List of Figures

Figure 1	DSB repair pathways
Figure 2	Clonal evolution of primary and post-treatment gliomas

## List of Tables

N/A

## Mutational signatures of genotoxic cancer therapies

### References

1. Chang, H.H.Y., Pannunzio, N.R., Adachi, N. & Lieber, M.R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* **18**, 495-506 (2017).
2. Sung, H., *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209-249 (2021).
3. Siegel, R.L., Miller, K.D., Fuchs, H.E. & Jemal, A. Cancer statistics, 2022. *CA Cancer J Clin* **72**, 7-33 (2022).
4. Mateo, J., *et al.* Delivering precision oncology to patients with cancer. *Nat Med* **28**, 658-665 (2022).
5. Pich, O., *et al.* The mutational footprints of cancer therapies. *Nat Genet* **51**, 1732-1740 (2019).
6. Miller, K.D., *et al.* Brain and other central nervous system tumor statistics, 2021. *CA Cancer J Clin* **71**, 381-406 (2021).
7. Ostrom, Q.T., Cioffi, G., Waite, K., Kruchko, C. & Barnholtz-Sloan, J.S. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2014-2018. *Neuro Oncol* **23**, iii1-iii105 (2021).
8. Stupp, R., *et al.* Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* **352**, 987-996 (2005).
9. Keime-Guibert, F., *et al.* Radiotherapy for glioblastoma in the elderly. *N Engl J Med* **356**, 1527-1535 (2007).
10. Hegi, M.E., *et al.* MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med* **352**, 997-1003 (2005).
11. Gilbert, M.R., *et al.* A randomized trial of bevacizumab for newly diagnosed glioblastoma. *N Engl J Med* **370**, 699-708 (2014).
12. Chinot, O.L., *et al.* Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma. *N Engl J Med* **370**, 709-722 (2014).
13. Wick, W., *et al.* Lomustine and Bevacizumab in Progressive Glioblastoma. *N Engl J Med* **377**, 1954-1963 (2017).
14. Perry, J.R., *et al.* Short-Course Radiation plus Temozolomide in Elderly Patients with Glioblastoma. *N Engl J Med* **376**, 1027-1037 (2017).
15. Herrlinger, U., *et al.* Lomustine-temozolomide combination therapy versus standard temozolomide therapy in patients with newly diagnosed glioblastoma with methylated MGMT promoter (CeTeG/NOA-09): a randomised, open-label, phase 3 trial. *Lancet* **393**, 678-688 (2019).
16. Stummer, W., *et al.* Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. *Lancet Oncol* **7**, 392-401 (2006).
17. Sanai, N. & Berger, M.S. Surgical oncology for gliomas: the state of the art. *Nat Rev Clin Oncol* **15**, 112-125 (2018).
18. Molinaro, A.M., *et al.* Association of Maximal Extent of Resection of Contrast-Enhanced and Non-Contrast-Enhanced Tumor With Survival Within Molecular Subgroups of Patients With Newly Diagnosed Glioblastoma. *JAMA Oncol* **6**, 495-503 (2020).
19. Cancer Genome Atlas Research, N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061-1068 (2008).

## Mutational signatures of genotoxic cancer therapies

20. Parsons, D.W., *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807-1812 (2008).
21. 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* **17**, 98-110 (2010).
22. Cancer Genome Atlas Research, N., *et al.* Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med* **372**, 2481-2498 (2015).
23. Ceccarelli, M., *et al.* Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* **164**, 550-563 (2016).
24. Wang, Q., *et al.* Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell* **32**, 42-56 e46 (2017).
25. Capper, D., *et al.* DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469-474 (2018).
26. Louis, D.N., *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* **131**, 803-820 (2016).
27. Consortium, G. Glioma through the looking GLASS: molecular evolution of diffuse gliomas and the Glioma Longitudinal Analysis Consortium. *Neuro Oncol* **20**, 873-884 (2018).
28. Fu, D., Calvo, J.A. & Samson, L.D. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat Rev Cancer* **12**, 104-120 (2012).
29. Lomax, M.E., Folkes, L.K. & O'Neill, P. Biological consequences of radiation-induced DNA damage: relevance to radiotherapy. *Clin Oncol (R Coll Radiol)* **25**, 578-585 (2013).
30. Lord, C.J. & Ashworth, A. The DNA damage response and cancer therapy. *Nature* **481**, 287-294 (2012).
31. Behjati, S., *et al.* Mutational signatures of ionizing radiation in second malignancies. *Nat Commun* **7**, 12605 (2016).
32. Choi, D.K., Helenowski, I. & Hijiya, N. Secondary malignancies in pediatric cancer survivors: perspectives and review of the literature. *Int J Cancer* **135**, 1764-1773 (2014).
33. Sharma, R.A., *et al.* Clinical development of new drug-radiotherapy combinations. *Nat Rev Clin Oncol* **13**, 627-642 (2016).
34. Petroni, G., Cantley, L.C., Santambrogio, L., Formenti, S.C. & Galluzzi, L. Radiotherapy as a tool to elicit clinically actionable signalling pathways in cancer. *Nat Rev Clin Oncol* **19**, 114-131 (2022).
35. Cytlak, U.M., *et al.* Immunomodulation by radiotherapy in tumour control and normal tissue toxicity. *Nat Rev Immunol* **22**, 124-138 (2022).
36. Scully, R., Panday, A., Elango, R. & Willis, N.A. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol* **20**, 698-714 (2019).
37. Daley, J.M., Niu, H., Miller, A.S. & Sung, P. Biochemical mechanism of DSB end resection and its regulation. *DNA Repair (Amst)* **32**, 66-74 (2015).
38. De Ioannes, P., Malu, S., Cortes, P. & Aggarwal, A.K. Structural basis of DNA ligase IV-Artemis interaction in nonhomologous end-joining. *Cell Rep* **2**, 1505-1512 (2012).



## Mutational signatures of genotoxic cancer therapies

39. Ramsden, D.A., Carvajal-Garcia, J. & Gupta, G.P. Mechanism, cellular functions and cancer roles of polymerase-theta-mediated DNA end joining. *Nat Rev Mol Cell Biol* **23**, 125-140 (2022).
40. Zhang, J., Stevens, M.F. & Bradshaw, T.D. Temozolomide: mechanisms of action, repair and resistance. *Curr Mol Pharmacol* **5**, 102-114 (2012).
41. Wick, W., *et al.* MGMT testing--the challenges for biomarker-based glioma treatment. *Nat Rev Neurol* **10**, 372-385 (2014).
42. Hunter, C., *et al.* A hypermutation phenotype and somatic MSH6 mutations in recurrent human malignant gliomas after alkylator chemotherapy. *Cancer Res* **66**, 3987-3991 (2006).
43. Johnson, B.E., *et al.* Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* **343**, 189-193 (2014).
44. Alexandrov, L.B., *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-421 (2013).
45. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
46. Nik-Zainal, S., *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979-993 (2012).
47. Devarajan, K. Nonnegative matrix factorization: an analytical and interpretive tool in computational biology. *PLoS Comput Biol* **4**, e1000029 (2008).
48. Zou, X., *et al.* A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nat Cancer* **2**, 643-657 (2021).
49. Alexandrov, L.B., *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94-101 (2020).
50. Szikriszt, B., *et al.* A comprehensive survey of the mutagenic impact of common cancer cytotoxics. *Genome Biol* **17**, 99 (2016).
51. Kucab, J.E., *et al.* A Compendium of Mutational Signatures of Environmental Agents. *Cell* **177**, 821-836 e816 (2019).
52. Boot, A., *et al.* In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. *Genome Res* **28**, 654-665 (2018).
53. Poon, S.L., McPherson, J.R., Tan, P., Teh, B.T. & Rozen, S.G. Mutation signatures of carcinogen exposure: genome-wide detection and new opportunities for cancer prevention. *Genome Med* **6**, 24 (2014).
54. Liu, D., *et al.* Mutational patterns in chemotherapy resistant muscle-invasive bladder cancer. *Nat Commun* **8**, 2193 (2017).
55. Li, B., *et al.* Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia. *Blood* **135**, 41-55 (2020).
56. Yang, F., *et al.* Chemotherapy and mismatch repair deficiency cooperate to fuel TP53 mutagenesis and ALL relapse. *Nat Cancer* **2**, 819-834 (2021).
57. Lopez, G.Y., *et al.* The genetic landscape of gliomas arising after therapeutic radiation. *Acta Neuropathol* **137**, 139-150 (2019).
58. Phi, J.H., *et al.* Genomic analysis reveals secondary glioblastoma after radiotherapy in a subset of recurrent medulloblastomas. *Acta Neuropathol* **135**, 939-953 (2018).
59. Davidson, P.R., Sherborne, A.L., Taylor, B., Nakamura, A.O. & Nakamura, J.L. A pooled mutational analysis identifies ionizing radiation-associated mutational signatures conserved between mouse and human malignancies. *Sci Rep* **7**, 7645 (2017).

## Mutational signatures of genotoxic cancer therapies

60. Rose Li, Y., *et al.* Mutational signatures in tumours induced by high and low energy radiation in Trp53 deficient mice. *Nat Commun* **11**, 394 (2020).
61. Sherborne, A.L., *et al.* Mutational Analysis of Ionizing Radiation Induced Neoplasms. *Cell Rep* **12**, 1915-1926 (2015).
62. Barthel, F.P., *et al.* Longitudinal molecular trajectories of diffuse glioma in adults. *Nature* **576**, 112-120 (2019).
63. Kocakavuk, E., *et al.* Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer. *Nat Genet* **53**, 1088-1096 (2021).
64. Varn, F.S., *et al.* Glioma progression is shaped by genetic evolution and microenvironment interactions. *Cell* **185**, 2184-2199 e2116 (2022).
65. Amin, S.B., *et al.* Comparative Molecular Life History of Spontaneous Canine and Human Gliomas. *Cancer Cell* **37**, 243-257 e247 (2020).
66. 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* **35**, 692-704 e612 (2019).
67. Touat, M., *et al.* Mechanisms and therapeutic implications of hypermutation in gliomas. *Nature* **580**, 517-523 (2020).
68. Louis, D.N., *et al.* The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol* **23**, 1231-1251 (2021).
69. Sampson, J.H., Gunn, M.D., Fecci, P.E. & Ashley, D.M. Brain immunology and immunotherapy in brain tumours. *Nat Rev Cancer* **20**, 12-25 (2020).
70. Reardon, D.A., *et al.* Effect of Nivolumab vs Bevacizumab in Patients With Recurrent Glioblastoma: The CheckMate 143 Phase 3 Randomized Clinical Trial. *JAMA Oncol* **6**, 1003-1010 (2020).
71. Weller, M., *et al.* EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. *Nat Rev Clin Oncol* **18**, 170-186 (2021).
72. Price, J.M., Prabhakaran, A. & West, C.M.L. Predicting tumour radiosensitivity to deliver precision radiotherapy. *Nat Rev Clin Oncol* (2022).
73. Lin, K., *et al.* Mechanism-based design of agents that selectively target drug-resistant glioma. *Science* **377**, 502-511 (2022).
74. Groelly, F.J., Fawkes, M., Dagg, R.A., Blackford, A.N. & Tarsounas, M. Targeting DNA damage response pathways in cancer. *Nat Rev Cancer* (2022).
75. Brady, S.W., Gout, A.M. & Zhang, J. Therapeutic and prognostic insights from the analysis of cancer mutational signatures. *Trends Genet* **38**, 194-208 (2022).
76. Zhao, E.Y., *et al.* Homologous Recombination Deficiency and Platinum-Based Therapy Outcomes in Advanced Breast Cancer. *Clin Cancer Res* **23**, 7521-7530 (2017).
77. Fong, P.C., *et al.* Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med* **361**, 123-134 (2009).
78. Davies, H., *et al.* HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat Med* **23**, 517-525 (2017).
79. Staaf, J., *et al.* Whole-genome sequencing of triple-negative breast cancers in a population-based clinical study. *Nat Med* **25**, 1526-1533 (2019).
80. Nguyen, L., J, W.M.M., Van Hoeck, A. & Cuppen, E. Pan-cancer landscape of homologous recombination deficiency. *Nat Commun* **11**, 5584 (2020).
81. Campbell, B.B., *et al.* Comprehensive Analysis of Hypermutation in Human Cancer. *Cell* **171**, 1042-1056 e1010 (2017).
82. Le, D.T., *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409-413 (2017).

## Mutational signatures of genotoxic cancer therapies

83. Mandal, R., *et al.* Genetic diversity of tumors with mismatch repair deficiency influences anti-PD-1 immunotherapy response. *Science* **364**, 485-491 (2019).
84. Le, D.T., *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* **372**, 2509-2520 (2015).
85. Lemery, S., Keegan, P. & Pazdur, R. First FDA Approval Agnostic of Cancer Site - When a Biomarker Defines the Indication. *N Engl J Med* **377**, 1409-1412 (2017).
86. Samstein, R.M., *et al.* Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet* **51**, 202-206 (2019).
87. McGrail, D.J., *et al.* High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Ann Oncol* **32**, 661-672 (2021).
88. Litchfield, K., *et al.* Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* **184**, 596-614 e514 (2021).
89. Cuppen, E., *et al.* Implementation of Whole-Genome and Transcriptome Sequencing Into Clinical Cancer Care. *JCO Precis Oncol* **6**, e2200245 (2022).
90. Duncavage, E.J., *et al.* Genome Sequencing as an Alternative to Cytogenetic Analysis in Myeloid Cancers. *N Engl J Med* **384**, 924-935 (2021).
91. Wong, M., *et al.* Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer. *Nat Med* **26**, 1742-1753 (2020).
92. van der Velden, D.L., *et al.* The Drug Rediscovery protocol facilitates the expanded use of existing anticancer drugs. *Nature* **574**, 127-131 (2019).
93. Tamborero, D., *et al.* Support systems to guide clinical decision-making in precision oncology: The Cancer Core Europe Molecular Tumor Board Portal. *Nat Med* **26**, 992-994 (2020).
94. Patel, A.P., *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396-1401 (2014).
95. Tirosh, I., *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309-313 (2016).
96. Venteicher, A.S., *et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**(2017).
97. Minata, M., *et al.* Phenotypic Plasticity of Invasive Edge Glioma Stem-like Cells in Response to Ionizing Radiation. *Cell Rep* **26**, 1893-1905 e1897 (2019).
98. Neftel, C., *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849 e821 (2019).
99. Chaligne, R., *et al.* Epigenetic encoding, heritability and plasticity of glioma transcriptional cell states. *Nat Genet* **53**, 1469-1479 (2021).
100. Johnson, K.C., *et al.* Single-cell multimodal glioma analyses identify epigenetic regulators of cellular plasticity and environmental stress response. *Nat Genet* **53**, 1456-1468 (2021).
101. Marine, J.C., Dawson, S.J. & Dawson, M.A. Non-genetic mechanisms of therapeutic resistance in cancer. *Nat Rev Cancer* **20**, 743-756 (2020).
102. Milanovic, M., *et al.* Senescence-associated reprogramming promotes cancer stemness. *Nature* **553**, 96-100 (2018).
103. Duy, C., *et al.* Chemotherapy Induces Senescence-Like Resilient Cells Capable of Initiating AML Recurrence. *Cancer Discov* **11**, 1542-1561 (2021).
104. Shlush, L.I., *et al.* Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature* **547**, 104-108 (2017).

## Mutational signatures of genotoxic cancer therapies

105. Fennell, K.A., *et al.* Non-genetic determinants of malignant clonal fitness at single-cell resolution. *Nature* **601**, 125-131 (2022).
106. Rehman, S.K., *et al.* Colorectal Cancer Cells Enter a Diapause-like DTP State to Survive Chemotherapy. *Cell* **184**, 226-242 e221 (2021).
107. Gupta, P.B., *et al.* Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**, 633-644 (2011).
108. Shaffer, S.M., *et al.* Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431-435 (2017).
109. Sanchez-Danes, A., *et al.* A slow-cycling LGR5 tumour population mediates basal cell carcinoma relapse after therapy. *Nature* **562**, 434-438 (2018).
110. Drews, R.M., *et al.* A pan-cancer compendium of chromosomal instability. *Nature* **606**, 976-983 (2022).
111. Steele, C.D., *et al.* Signatures of copy number alterations in human cancer. *Nature* **606**, 984-991 (2022).
112. Li, Y., *et al.* Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112-121 (2020).
113. Charpentier, M., Spada, S., Van Nest, S.J. & Demaria, S. Radiation therapy-induced remodeling of the tumor immune microenvironment. *Semin Cancer Biol* **86**, 737-747 (2022).
114. Thol, K., Pawlik, P. & McGranahan, N. Therapy sculpts the complex interplay between cancer and the immune system during tumour evolution. *Genome Med* **14**, 137 (2022).
115. Vazquez-Garcia, I., *et al.* Ovarian cancer mutational processes drive site-specific immune evasion. *Nature* (2022).
116. Jimenez-Sanchez, A., *et al.* Unraveling tumor-immune heterogeneity in advanced ovarian cancer uncovers immunogenic effect of chemotherapy. *Nat Genet* **52**, 582-593 (2020).
117. Crisafulli, G., *et al.* Temozolomide Treatment Alters Mismatch Repair and Boosts Mutational Burden in Tumor and Blood of Colorectal Cancer Patients. *Cancer Discov* **12**, 1656-1675 (2022).
118. Lussier, D.M., *et al.* Radiation-induced neoantigens broaden the immunotherapeutic window of cancers with low mutational loads. *Proc Natl Acad Sci U S A* **118**(2021).
119. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L.H. & Aerts, H. Artificial intelligence in radiology. *Nat Rev Cancer* **18**, 500-510 (2018).
120. Bera, K., Schalper, K.A., Rimm, D.L., Velcheti, V. & Madabhushi, A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol* **16**, 703-715 (2019).
121. Kather, J.N., *et al.* Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat Cancer* **1**, 789-799 (2020).
122. Fu, Y., *et al.* Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat Cancer* **1**, 800-810 (2020).
123. Boehm, K.M., *et al.* Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. *Nat Cancer* **3**, 723-733 (2022).
124. Boehm, K.M., Khosravi, P., Vanguri, R., Gao, J. & Shah, S.P. Harnessing multimodal data integration to advance precision oncology. *Nat Rev Cancer* **22**, 114-126 (2022).

## Mutational signatures of genotoxic cancer therapies

### Avoidance of self-plagiarism

In the context of this doctoral work, the following articles were published:

1. Barthel, F.P., Johnson, K.C., Varn, F.S., Moskalik, A.D., Tanner, G., Kocakavuk, E. *et al.* Longitudinal molecular trajectories of diffuse glioma in adults. *Nature* 576, 112–120 (2019). <https://doi.org/10.1038/s41586-019-1775-1>
2. Kocakavuk, E., Anderson, K.J., Varn, F.S. *et al.* Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer. *Nat Genet* 53, 1088–1096 (2021). <https://doi.org/10.1038/s41588-021-00874-3>
3. Varn, F.S., Johnson, K.C., Martinek J., Huse J.T., Nasrallah, M.P., Wesseling, P., Cooper, L.A.D., Malta, T.M., Wade, T.E., Sabedot, T.S., Brat, D., Gould, P.V., Wöehrer, A., Aldape, K., Ismail, A., Sivajothi, S.K., Barthel, F.P., Kim, H., Kocakavuk, E., *et al.* Glioma progression is shaped by genetic evolution and microenvironment interactions. *Cell* 185, 2184-2199 (2022). <https://doi.org/10.1016/j.cell.2022.04.038>
4. Amin, S.A., Anderson, K.J., Boudreau, C.E., Martinez-Ledesma, E., Kocakavuk, E., *et al.* Comparative Molecular Life History of Spontaneous Canine and Human Gliomas. *Cancer Cell* 37, 243-257.e7 (2020). <https://doi.org/10.1016/j.ccell.2020.01.004>

**Mutational signatures of genotoxic cancer therapies**

**Statutory declarations**

Declarations required for the submission of the thesis

**Declaration:**

In accordance with § 6 (para. 2, clause g) of the Regulations Governing the Doctoral Proceedings of the Faculty of Biology for awarding the doctoral degree Dr. rer. nat., I hereby declare that I represent the field to which the topic “*Mutational signatures of genotoxic cancer therapies*” is assigned in research and teaching and that I support the application of *Dr. med. Emre Kocakavuk*.

Essen, 20.12.2022

Prof. Dr. C. Reinhardt

Name of the scientific  
supervisor/member of the  
University of Duisburg-Essen



Signature of the supervisor/  
member of the University of Duisburg-Essen

**Declaration:**

In accordance with § 7 (para. 2, clause d and f) of the Regulations Governing the Doctoral Proceedings of the Faculty of Biology for awarding the doctoral degree Dr. rer. nat., I hereby declare that I have written the herewith submitted dissertation independently using only the materials listed, and have cited all sources taken over verbatim or in content as such.

Essen, 20.12.2022



Signature of the doctoral candidate

**Declaration:**

In accordance with § 7 (para. 2, clause e and g) of the Regulations Governing the Doctoral Proceedings of the Faculty of Biology for awarding the doctoral degree Dr. rer. nat., I hereby declare that I have undertaken no previous attempts to attain a doctoral degree, that the current work has not been rejected by any other faculty, and that I am submitting the dissertation only in this procedure.

Essen, 20.12.2022



Signature of the doctoral candidate

**Annex to the Application for Admission to the Doctoral Examination**  
**Documentation of accreditation achieved during the qualifying phase in accordance with § 6a of the Regulations Governing the Doctoral Proceedings of the Faculty of Biology of the University of Duisburg-Essen dated 05.08.2013 (Gazette Year 11, 2013 p. 955/No. 121).**

<b>Credit points may be attained by (c.f. Doctoral Proceedings § 6a):</b>	<b>Accreditation attained during the qualifying phase</b>	<b>Credit points (CP)</b>
– Participation in courses for the acquisition of interdisciplinary qualifications (soft skills)	<ul style="list-style-type: none"> <li>- “Communicating Science” course, Banbury Center, Cold Spring Harbor, BY, USA (1 CP)</li> <li>- Education day at SNO 2019 (1 CP)</li> </ul>	2 CP
– Participation in specific modules of the Faculty’s Masters programme as doctoral project preparation	NA	
– Conduction of teaching courses or leading study groups	NA	
– Participation at conferences with presentation of own work	<ul style="list-style-type: none"> <li>- Talk at EANO 2022 (2 CP)</li> <li>- Poster at ESMO 2022 (2 CP)</li> <li>- Talk at EANO 2021 (2 CP)</li> <li>- Talk at ESMO 2021 (2 CP)</li> <li>- Talk at AACR Radiation Science and Medicine 2021 (2 CP)</li> <li>- Talk at SNO 2020 (2 CP)</li> <li>- Talk at MAP 2020 (2 CP)</li> </ul>	14 CP
– Other comparable achievements	<ul style="list-style-type: none"> <li>- First-author publication in Nature Genetics (3 CP)</li> <li>- Co-author publication in Nature (1 CP)</li> <li>- Co-author publication in Cell (1 CP)</li> <li>- Co-author publication in Cancer Cell (1 CP)</li> </ul>	6 CP

**The accreditation of more than 18 credit points during the qualifying phase has been achieved:**

**Dr. med. Emre Kocakavuk**

**Doctoral Candidate**

**Prof. Dr. med. C. Reinhardt**

**Doctoral Supervisor**

Essen, 20.12.2022



**Signature of the Doctoral Candidate**



**Signature of the Doctoral Supervisor**



## Declaration regarding the form of publication of the thesis

The publication has to take place by the free submission of:

- a) **40 copies** in case of own organized printing without distribution through bookstores
- b) **3 specimen copies** in case of publication in a scientific journal, series or as a monograph with distribution via book trade
- c) **3 copies** with a minimum print run of 150 copies via commercial publishers
- d) **2 bound copies** and an electronic version of the dissertation\*.
- e) **6 bound copies** in case of a cumulative dissertation without publication in digital form

\*In case of submission option d):

- The electronic version of the thesis may be published directly online.
- Publication of the thesis should be delayed by one year.

A blocking form with justification is enclosed.

Note: If it is foreseeable that patent or protection law reasons will last considerably longer than 1 year, an alternative submission option without submission of an electronic version is preferable (see submission option a or e).

Essen, 20.12.2022

Dr. med. Emre Kocakavuk

Name of doctoral candidate



Signature of doctoral candidate

Essen, 20.12.2022

Prof. Dr. med. C. Reinhardt

Name of supervisor



Signature of supervisor

### **Cumulative Thesis/Extent of Contribution**

Cumulative thesis of Dr. med. Emre Kocakavuk

#### **Author contributions**

Title of the publication: *Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer*

Authors: *E Kocakavuk, KJ Anderson, FS Varn, KC Johnson, SB Amin, EP Sulman, MP Lolkema, FP Barthel and RGW Verhaak*

#### **Contributions:**

- Conception - 90%
- Conduction of computational work - 95%
- Data analysis - 100%
- Species identification – N/A
- Statistical analysis – 100%
- Writing the manuscript – 75%
- Revision of the manuscript – 75%



Signature of the Doctoral Candidate



Signature of the Doctoral Supervisor

**Cumulative Thesis/Extent of Contribution**

Cumulative thesis of Dr. med. Emre Kocakavuk

**Author contributions**

Title of the publication: *Longitudinal molecular trajectories of diffuse glioma in adults*

Authors: *FP Barthel, KC Johnson, FS Varn, AD Moskalik, G Tanner, E Kocakavuk ... JT Huse, JF De Groot, LF Stead, RGW Verhaak & The GLASS Consortium*

**Contributions:**

- Conception - 20%
- Conduction of computational work - 25%
- Data analysis - 25%:
- Species identification – N/A
- Statistical analysis - 25%
- Writing the manuscript - 20%
- Revision of the manuscript - 25%



Signature of the Doctoral Candidate



---

Signature of the Doctoral Supervisor

**Cumulative Thesis/Extent of Contribution**

Cumulative thesis of Dr. med. Emre Kocakavuk

**Author contributions**

Title of the publication: *Comparative Molecular Life History of Spontaneous Canine and Human Gliomas*

Authors: *SA Amin, KJ Anderson, CE Boudreau, E Martinez-Ledesma, E Kocakavuk ... AB Heimberger, JM Levine and RGW Verhaak*

Contributions:

- Conception - 20%
- Conduction of computational work - 20%
- Data analysis - 30%
- Species identification - 5%
- Statistical analysis - 30%
- Writing the manuscript - 20%
- Revision of the manuscript - 20%



Signature of the Doctoral Candidate



Signature of the Doctoral Supervisor

## **Cumulative Thesis/Extent of Contribution**

Cumulative thesis of Dr. med. Emre Kocakavuk

### **Author contributions**

Title of the publication: *Glioma progression is shaped by genetic evolution and microenvironment interactions*


Authors: *Varn, F.S., Johnson, K.C., Martinek J., Huse J.T., Nasrallah, M.P., Wesseling, P., Cooper, L.A.D., Malta, T.M., Wade, T.E., Sabedot, T.S., Brat, D., Gould, P.V., Wöehrer, A., Aldape, K., Ismail, A., Sivajothi, S.K., Barthel, F.P., Kim, H., Kocakavuk, E., ..., Verhaak R.G.W.*

### **Contributions:**

- Conception - 15%
- Conduction of computational work - 20%
- Data analysis - 25%
- Species identification – N/A
- Statistical analysis - 20%
- Writing the manuscript - 10%
- Revision of the manuscript - 25%



Signature of the Doctoral Candidate



---

Signature of the Doctoral Supervisor