

# Longitudinal molecular trajectories of diffuse glioma in adults

GLASS Consortium; Watts, Colin

DOI:

[10.1038/s41586-019-1775-1](https://doi.org/10.1038/s41586-019-1775-1)

License:

None: All rights reserved

*Document Version*

Peer reviewed version

*Citation for published version (Harvard):*

GLASS Consortium & Watts, C 2019, 'Longitudinal molecular trajectories of diffuse glioma in adults', *Nature*, vol. 576, no. 7785, pp. 112-120. <https://doi.org/10.1038/s41586-019-1775-1>

[Link to publication on Research at Birmingham portal](#)

## General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

## Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# 1 LONGITUDINAL MOLECULAR TRAJECTORIES OF DIFFUSE GLIOMA IN ADULTS

2

## 3 **The Glioma Longitudinal Analysis (GLASS) Consortium**

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

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

25 <sup>2</sup>Amsterdam UMC, Vrije Universiteit Amsterdam, Departments of Neurology, Neurosurgery,  
26 Pathology, Brain Tumor Center Amsterdam, de Boelelaan 1117, Amsterdam, Netherlands

27 <sup>3</sup>Leeds Institute of Medical Research at St James's, University of Leeds, LS9 7TF, UK.

28 <sup>4</sup>DKFZ Division of Translational Neurooncology at the West German Cancer Center, German  
29 Cancer Consortium Partner Site & Department of Neurosurgery, University Hospital Essen,  
30 45147 Essen, Germany

31 <sup>5</sup>National Cancer Institute, Bethesda, MD 20892, USA

32 <sup>6</sup>Departments of Neuro-Oncology, Neurosurgery, Pathology, Translational Molecular Pathology,  
33 The University of Texas MD Anderson Cancer Center, Houston, Texas 77031, USA

34 <sup>7</sup>CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, 1090  
35 Vienna, Austria

36 <sup>8</sup>1<sup>st</sup> Department of Pathology and Experimental Cancer Research, Semmelweis University, 1085  
37 Budapest, Hungary

38 <sup>9</sup>Preston Robert Tisch Brain Tumor Center at Duke, Duke University Medical Center, Durham,  
39 North Carolina 27710, USA

40 <sup>10</sup>Dana-Farber Cancer Institute, 450 Brookline Ave, Boston, MA 02215, USA

41 <sup>11</sup>Broad Institute, 415 Main Street, Cambridge, MA 02142, USA

42 <sup>12</sup>Department of Population and Quantitative Health Sciences and Case Comprehensive Cancer  
43 Center, Case Western Reserve University School of Medicine, 2103 Cornell Rd, WRB 2-526,  
44 Cleveland, Ohio 44106, USA

45 <sup>13</sup>Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria

46 <sup>14</sup>Division of Neuro-Oncology, Massachusetts General Hospital, Boston, MA 02114, USA  
47 <sup>15</sup>Department of Pathology, Northwestern University Feinberg School of Medicine, Chicago IL,  
48 60611, USA  
49 <sup>16</sup>University of Liverpool & Walton Centre NHS Trust, Liverpool, L9 7LJ, UK  
50 <sup>17</sup>Division of Neurosurgery, The University of Connecticut Health Center, Farmington, CT, USA  
51 <sup>18</sup>Leeds Teaching Hospital NHS Trust, St James's University Hospital, Leeds, LS9 7TF, UK  
52 <sup>19</sup>Department of Radiation Oncology, Arthur G. James Hospital/Ohio State Comprehensive  
53 Cancer Center, Columbus, OH, 43210, USA  
54 <sup>20</sup>UConn Health, Department of Genetics and Genome Sciences, Farmington, CT, 06030, USA  
55 <sup>21</sup>Yale University School of Public Health, New Haven, CT, 06511, USA  
56 <sup>22</sup>Department of Neurosurgery, Brigham and Women's Hospital, Boston, MA, USA  
57 <sup>23</sup>Departments of Neurology, Pathology, Radiology and Biomedical Engineering, Medical  
58 College of Wisconsin, Milwaukee, WI USA  
59 <sup>24</sup>Department of Neurosurgery, University of California San Francisco, CA 94143, USA  
60 <sup>25</sup>Fondazione IRCCS Istituto Neurologico Besta, Milano, Italy  
61 <sup>26</sup>Division of Molecular Genetics, Heidelberg Center for Personalized Oncology, German Cancer  
62 Research Consortium, German Cancer Research Center (DKFZ), Heidelberg, Germany  
63 <sup>27</sup>Departments of Neurology, Radiology and Nuclear Medicine, Erasmus MC - University  
64 Medical Center Rotterdam, PO Box 2040, 3000 CA Rotterdam, the Netherlands  
65 <sup>28</sup>Olivia Newton-John Cancer Research Institute, Austin Health, Melbourne, Australia  
66 <sup>29</sup>La Trobe University School of Cancer Medicine, Heidelberg, Victoria, Australia;  
67 <sup>30</sup>Department of Medicine, University of Melbourne, Heidelberg, Victoria, Australia  
68 <sup>31</sup>Neuro-oncology Branch, National Institutes of Health, Bethesda, Maryland, 20892, USA  
69 <sup>32</sup>Anatomic Pathology Service, Hôpital de l'Enfant-Jésus, CHU de Québec-Université Laval,  
70 Québec QC G1J 1Z4, Canada  
71 <sup>33</sup>Departments of Neurology, Pathology, Cell Biology, Systems Biology and Biomedical  
72 Informatics, Institute for Cancer Genetics, Columbia University Medical Center, New York, New  
73 York 10032, USA  
74 <sup>34</sup>Cooperative Trials Group for Neuro-Oncology (COGNO) NHMRC Clinical Trials Centre, The  
75 University of Sydney, New South Wales, Australia  
76 <sup>35</sup>Departments of Neurosurgery, Public Health Sciences, Henry Ford Health System, Henry  
77 Ford Cancer Institute, Detroit, MI 48202, USA  
78 <sup>36</sup>Cure Brain Cancer Biomarkers and Translational Research Group, Prince of Wales Clinical  
79 School, UNSW Sydney, Australia  
80 <sup>37</sup>Department of Neurosurgery, Sungkyunkwan University School of Medicine, Samsung  
81 Medical Center, Seoul, Korea  
82 <sup>38</sup>Department of Anatomical and Cellular Pathology, The Chinese University of Hong Kong, 1/F,  
83 Prince of Wales Hospital, Shatin, Hong Kong  
84 <sup>39</sup>Department of Oncology, Luxembourg Institute of Health, Luxembourg  
85 <sup>40</sup>Department of Neurosurgery, University of Colorado School of Medicine, Aurora, Colorado,  
86 80045, USA  
87 <sup>41</sup>Department of Neurosurgery, Seoul National University College of Medicine, Seoul National  
88 University Hospital, Seoul, Korea  
89 <sup>42</sup>Institute of Neuropathology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany  
90 <sup>43</sup>Institute for Refractory Cancer Research, Samsung Medical Center, Seoul, Korea  
91 <sup>44</sup>Department of Neurosurgery, University Hospitals-Case Medical Center, Seidman Cancer  
92 Center, and the Case Comprehensive Cancer Center, Cleveland, Ohio 44106, USA  
93 <sup>45</sup>The Hospital for Sick Children, Toronto, ON, M5G1X8, Canada  
94 <sup>46</sup>Interdisciplinary Division of Neuro-Oncology, Hertie Institute for Clinical Brain Research, DKTK  
95 Partner Site Tübingen, Eberhard Karls University Tübingen, Germany

96 <sup>47</sup>Department of Neurosurgery, School of Medicine and Winship Cancer Institute of Emory U  
97 University; 1365C Clifton Rd. NE, Atlanta, GA30084, USA  
98 <sup>48</sup>Institute of Cancer Genome Sciences, Department of Neurosurgery, University of Birmingham,  
99 UK  
100 <sup>49</sup>Department of Neurology, University Hospital Zurich, Zurich, Switzerland  
101 <sup>50</sup>Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands  
102 <sup>51</sup>Department of Neurosurgery, Medical University of Vienna, 1090 Vienna, Austria  
103 <sup>52</sup>Institute of Neurology, Medical University of Vienna, 1090 Vienna, Austria  
104 <sup>53</sup>Division of Neurosurgery, Department of Surgery, University Health Network, Toronto, C  
105 Canada.  
106 <sup>54</sup>These authors contributed equally.  
107  
108 † Correspondence to [roel.verhaak@jax.org](mailto:roel.verhaak@jax.org).  
109

## 110 **ABSTRACT**

111 The evolutionary processes that drive universal therapeutic resistance in adult patients with  
112 diffuse glioma remain unclear<sup>1,2</sup>. Here, we analyzed temporally separated DNA sequencing data  
113 and matched clinical annotation from 222 patients with glioma. Through mutational and copy  
114 number analyses across the three major subtypes of diffuse glioma, we observed that driver  
115 genes detected at initial disease were retained at recurrence, while there was little evidence of  
116 recurrence-specific gene alterations. Treatment with alkylating-agents resulted in a  
117 hypermutator phenotype at different rates across glioma subtypes, and hypermutation was not  
118 associated with differences in survival. Acquired aneuploidy was frequently detected in recurrent  
119 gliomas characterized by presence of an IDH mutation but without 1p/19q codeletion and further  
120 converged with acquired cell cycle alterations and poor outcomes. We show that the clonal  
121 architecture of each tumor remains similar over time and that absence of clonal selection was  
122 associated with increased survival. Finally, we did not observe differences in immunoediting  
123 levels between initial and recurrent glioma. Our results collectively argue that the strongest  
124 selective pressures occur early during glioma development and that current therapies shape this  
125 evolution in a largely stochastic manner.

126

## 127 **INTRODUCTION**

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

133 Intratumoral heterogeneity is a well-recognized characteristic of gliomas and results from  
134 selective pressures such as a limited availability of nutrients, clonal competition, and  
135 treatment<sup>12-15</sup>. Tumors are thought to circumvent these growth bottlenecks via dynamic  
136 competition of subclones resulting in the most favorable environment for tumor sustenance<sup>1</sup>.  
137 Recent studies have suggested that stochastic changes in clone frequency (i.e. neutral  
138 evolution) and immunogenic surveillance may further contribute to the observed intratumoral  
139 heterogeneity<sup>16,17</sup>. An understanding of evolutionary dynamics at multiple time points is needed  
140 to develop strategies aimed at delaying or preventing the onset of tumor progression.

141 To investigate clonal dynamics over time and in response to therapeutic pressures, we  
142 established the Glioma Longitudinal Analysis (GLASS) Consortium. GLASS is a community-  
143 driven effort that seeks to overcome the logistical challenges in constructing adequately  
144 powered longitudinal genomic glioma datasets by pooling datasets from patients treated at  
145 institutions worldwide<sup>18</sup>. We have analyzed longitudinal profiles across the three molecular  
146 glioma subtypes to identify the molecular processes active at initial and recurrent time points.  
147 These analyses identified few common features of glioma evolution across subtypes, and  
148 instead pointed toward highly variable and patient-specific trajectories of genomic alterations.

## 149 **RESULTS**

### 150 **GLASS cohort**

151 We pooled existing and newly generated longitudinal DNA sequencing datasets from 288  
152 patients treated at 35 hospitals (Supplementary Table 1, Extended Data Fig. 1). After applying  
153 quality filters, tumor samples from 222 patients with high-quality data in at least two time points  
154 were classified according to molecular markers into three major glioma subtypes: 1. IDH-mutant  
155 and chromosome 1p/19q co-deleted (IDHmutant-codel; n = 25) 2. IDH-mutant without  
156 chromosome 1p/19q codeletion (IDHmutant-noncodel; n = 63) and 3. IDH wild type (IDHwt; n =  
157 134), in alignment with the World Health Organization classification of Central Nervous System  
158 tumors<sup>10,11</sup>. For each patient we selected two time-separated tumor samples, henceforth initial  
159 and recurrence, for further analysis.

## 160 **Mutational burdens and processes over time**

161 We first evaluated temporal changes in mutational burden and processes to understand general  
162 patterns of glioma evolution. Mutation burdens in initial tumors were comparable with previously  
163 reported rates<sup>6,7,19</sup>. 2.20 mutations (single-nucleotide variants and small insertions/deletions)  
164 per Megabase (Mutations/Mb) for IDHmutant-codels; 2.52 Mutations/Mb for IDHmutant-  
165 noncodels; and 2.85 Mutations/Mb for IDHwt glioma (Fig. 1a; Extended Data Fig. 2a). Excluding  
166 DNA hypermutation cases (> 10 Mutations/Mb, n = 35), the mutation burden increased after  
167 recurrence in 70% of the cohort (Extended Data Fig. 2a). To study changes during tumor  
168 progression, we separated mutations into three fractions: initial only, recurrence only, or shared.  
169 Interestingly, private fraction but not shared fraction mutation burdens were comparable  
170 between subtypes (Extended Data Fig. 2b). Patient age at diagnosis was significantly  
171 associated with the shared mutational burden and to a lesser extent the mutation burden private  
172 to the initial tumor (Extended Data Fig. 2c). On average, tumors with longer time to recurrence  
173 had slightly higher mutation burdens (Extended Data Fig. 2d).

174 These fraction-specific differences in mutation burden suggested that the activity of  
175 distinct mutational processes may also be time-dependent. We therefore classified mutations in  
176 each fraction according to the Catalogue of Somatic Mutations in Cancer (COSMIC) signature  
177 database<sup>20</sup>. As expected, signature activity was closely related to subtype and fraction (Fig. 1b,  
178 Extended Data Fig. 3a). Signature 1 (aging) was nearly always the dominant signature amongst  
179 shared mutations in IDHwt tumors, whereas the shared fraction in IDHmut-noncodel and  
180 IDHmut-codel tumors - tumor subtypes associated with a younger age of diagnosis - additionally  
181 showed a strong presence of signature 16 (unknown etiology). Signatures 3 (double strand  
182 break repair) and 15 (mismatch repair) along with signature 8 (unknown etiology) were mostly  
183 confined to the private fractions, suggesting that these processes were of lesser importance to  
184 tumor maintenance than those associated with aging.

185 Treatment of glioma includes alkylating agents that can induce post-treatment  
186 hypermutation<sup>21-23</sup>. We observed enrichment of the associated signature 11 in recurrent tumors  
187 with a mutational load exceeding 10 Mutations/Mb and treated with alkylating agents (Fig. 1a,  
188 Extended Data Fig. 3b). Treatment-associated hypermutation occurred most frequently among  
189 IDHmutant-noncodels (47%), followed by IDHmutant-codels (25%), and IDHwt gliomas (16%)  
190 (Fig. 1c). The difference in the proportion of hypermutation events was significantly different  
191 between the three glioma subtypes (Fisher's exact-test  $P = 2.0e-03$ ), suggesting that IDHmutant  
192 noncodels are most sensitive to developing a hypermutator phenotype<sup>24</sup>.

193 Treatment-induced hypermutation has been associated with disease progression<sup>23</sup>. We  
194 did not find overall survival differences between alkylating agent-treated hypermutators and  
195 alkylating agent-treated non-hypermutators independent of age, subtype, and *MGMT*  
196 methylation status (Fig. 1d, Supplementary Table 2a-b). In order to further assess the  
197 pathogenicity of acquired mutations, we studied their clonality<sup>25</sup>. Newly acquired clonal  
198 mutations have penetrated most of the tumor (i.e., a selective sweep) between initial and  
199 recurrence and mark clonal expansion<sup>26</sup>. Conversely, acquired subclonal mutations are less  
200 prevalent, and therefore less likely to drive disease progression. Previous reports have  
201 suggested that alkylating agent-associated mutations hypermutation are frequently clonal<sup>27</sup>. We  
202 found that in 48% of hypermutated tumors a majority of the recurrence-only mutations were  
203 clonal, potentially reflecting cases where a selective sweep occurred (Extended Data Fig. 4a).  
204 However, IDHmut-noncodeI hypermutators with predominantly clonal mutations did not show  
205 differences in survival compared with those harboring predominantly subclonal mutations (log-  
206 rank test  $P = 0.38$ , Extended Data Fig. 4b). Alkylating agents such as temozolomide prolong  
207 survival of adult patients with glioma<sup>28,29</sup>. Our results show that treatment-induced  
208 hypermutation is common across subtypes and does not associate with a reduced overall  
209 survival supporting the noted benefit of alkylating agent therapy.

### 210 **Selective pressures during glioma evolution**

211 Environmental and treatment-induced pressures may drive changes in clonal architecture at  
212 recurrence. To evaluate selection over time we clustered copy number changes and mutations  
213 based on their cancer cell fraction (CCF). CCF values represent the fraction of cancer cells  
214 harboring a given alteration and reflect the relative timing of events, since alterations that are  
215 present in a subset of cancer cells likely occurred later than events present in all cancer cells  
216 (Fig. 2a). Most tumors (84%) demonstrated a mutational cluster with  $CCF > 50\%$  that persisted  
217 from the initial tumor into recurrence, likely reflecting the tumor trunk and harboring the tumor-  
218 initiating driver mutations (Fig. 2b, Extended Data Fig. 5a)<sup>30</sup>. To determine changes in clonal  
219 dominance over time we ranked clusters within each sample by their CCF and found similarities  
220 in clonal architecture throughout the course of disease (Kendall rank correlation,  $\tau = 0.20$ ,  $P =$   
221  $3.76E-24$ , Fig. 2b, Extended Data Fig. 5b-d). These results suggested that the clonal structure  
222 at initial disease mostly persisted into recurrence.

223 To deepen our assessment of selective pressures, we evaluated selection in initial and  
224 recurrent tumors by determining the normalized ratio between non-synonymous and  
225 synonymous mutations (dNdScv). Higher ratios ( $> 1$ ) suggest positive selection, and ratios less  
226 than one suggest negative selection. We found evidence for positive selection at both time

227 points despite differences between subtypes (Fig. 2c). Separating mutations into mutational  
228 fractions demonstrated that shared but not private mutations showed positive dN/dS ratios in all  
229 three glioma subtypes indicating that only shared mutations (including truncal mutations) are  
230 likely subject to positive selection (Fig. 2c). The dN/dS ratio of initial-only mutations showed that  
231 these are neither positively nor negatively selected for, while recurrence-only mutations were  
232 subject to negative selection in IDHwt.

233 To verify the reduced selective pressure in the private mutations we used an orthogonal  
234 method to test for evidence of selection (neutralitytestr)<sup>31</sup>. The method uses variant allele  
235 frequency distributions and estimated mutation rates to detect whether profiles significantly  
236 deviate from a model of neutral evolution (i.e. as depicted by a linear relationship in Fig. 2d). In  
237 accordance with dNdScv results, private mutations demonstrated dynamics consistent with  
238 neutral evolution (Fig. 2d). Shared subclonal mutations deviated from linearity and were  
239 consistent with selection both in non-hypermotators and hypermutators (Fig. 2d, Extended Data  
240 Fig. 6a-b), providing additional evidence that the strongest selective forces occur early in  
241 gliomagenesis.

242 Cohort-level analysis of selection masks the heterogeneity that exists in individual  
243 evolutionary trajectories. To determine the selective effects at each tumor time point we used a  
244 Bayesian framework (SubClonalSelection) which simultaneously provides sample-specific  
245 probabilities for both selection and neutrality while modeling sources of noise in sequencing  
246 data. The classification of a sample as “selection” or “neutral” is determined by whichever model  
247 has the greater probability. Classification as “neutral” reflects the accumulation of random  
248 mutations that are not subject to selection. Given the stringent algorithm requirements, 183  
249 patients were included in this analysis with at least one time point, and 104 patients with both  
250 time points (16 IDHmutant-codons, 29 IDHmutant-noncodons, 59 IDHwt, Supplementary Table  
251 3). Neutral to neutral was the most common evolutionary trajectory across all three subtypes  
252 (52%), and IDHwt tumors displayed the highest observed selection at any time point with  
253 selection detected in 64% of tumors (Fisher’s exact test  $P = 0.01$ , Fig. 2e, Supplementary Table  
254 3). IDHwt gliomas with evidence for selection at recurrence had a shorter overall survival than  
255 IDHwt gliomas classified as neutral at recurrence ( $P = 2.7E-02$ ; log-rank statistic, Fig. 2f),  
256 suggesting that subclonal competition associates with more aggressive tumor behavior. To  
257 address the limitations of smaller sample sizes in the IDH-mutant subtypes, we performed a  
258 Cox proportional hazards model including age at first diagnosis, all three glioma subtypes, and  
259 mode of selection at recurrence. This analysis revealed that selection at recurrence was  
260 significantly associated with shorter survival across subtypes (HR = 1.53 95% CI 1.00-2.41,  $P =$



261 4.8E-02, Supplementary Table 4). We next investigated whether radiation and chemotherapy  
262 imposed a selective effect, by comparing the evolutionary status at recurrence with treatment  
263 and other clinical variables. We did not observe significant associations between subclonal  
264 selection and radiation therapy or chemotherapy (Fisher's exact-test  $P > 0.05$ , Supplementary  
265 Table 5), suggesting that standard therapeutic approaches for glioma have limited impact on the  
266 subclonal tumor architecture. While high-depth sequencing datasets may be required to detect  
267 subtle selective effects<sup>26</sup>, our analyses raise the possibility that the survival benefit derived from  
268 standard chemoradiation results from tumor cell elimination where treatment sensitivity of  
269 individual cells is not determined by genetic factors.

### 270 **Driver alteration frequencies across time**

271 We evaluated how stability, acquisition, and loss of mutation and copy number drivers<sup>6</sup> over  
272 time impact glioma evolution. We used dNdScv to nominate 12 candidate mutation driver genes  
273 at both time points ( $Q < 0.05$ , Fig. 3a, Extended Data Fig. 7a) and determined significant copy  
274 number alterations that recapitulated previously identified drivers (Extended Data Fig. 7b).  
275 Mutations in *IDH1* and co-occurring 1p/19q chromosome-arm loss have been suggested as  
276 glioma-initiating events<sup>1</sup>, which was corroborated by the observation that these events were  
277 never lost or acquired during the surgical interval (Fig. 3a, Extended Data Fig. 8a). Similarly, we  
278 observed that *TERT* promoter mutations were almost always shared in the IDHmutant-codel  
279 and IDHwt, though many samples lacked sufficient coverage in this GC-rich region.  
280 Chromosome 7 gains and chromosome 10 losses were present in a large majority of IDHwt  
281 initial tumors and persisted into recurrence.

282 Shifts in the fraction of cancer cells harboring an event may also indicate a time  
283 dependency of drivers. We determined changes in cellular prevalence of shared driver events  
284 by ordering events in each sample by their CCF (Extended Data Fig. 9). *ATRX* mutations in  
285 IDHmutant-noncodel initial tumors demonstrated lower CCFs than *TP53* ( $P = 0.03$ ) and *IDH1* ( $P$   
286  $= 0.10$ ) mutations, suggesting *IDH1* and *TP53* mutations precede *ATRX* inactivation<sup>1</sup>. There  
287 was no difference in CCF between *IDH1* and *TP53* amongst initial gliomas ( $P = 0.98$ ), however,  
288 *IDH1* mutations demonstrated significantly lower CCFs compared with *TP53* ( $P = 0.0018$ ) in  
289 recurrent gliomas. We did not observe any CCF differences among driver mutations detected in  
290 IDHwt tumors at either time point. Chromosome 10 deletion CCFs were higher compared to  
291 chromosome 7 amplifications ( $P = 0.0036$ ) implying that chromosome 10 deletions arise earlier  
292 <sup>32</sup>. Similarly, there was no difference in CCF between *CDKN2A* deletion and *EGFR* amplification  
293 ( $P = 0.70$ ). *EGFR* and chromosomal arm events significantly differed (i.e. 10p del vs *EGFR*  
294 amp,  $P = 0.0019$ ) but not *CDKN2A* deletion and chromosomal events (i.e. 10p del vs *CDKN2A*

295 del,  $P = 0.33$ ). The consistently high CCF for *EGFR* amplifications could indicate that these  
296 events precede even some larger chromosomal aberrations, while not excluding the possibility  
297 that high levels of extrachromosomal *EGFR*<sup>33</sup> artificially inflate CCF.

298 Longitudinal changes in CCF values provide additional insights into evolutionary  
299 dynamics. For instance, the CCF value may increase when a driver event is linked to clonal  
300 expansion, or conversely, decrease when a clone is outcompeted. Most individual drivers did  
301 not demonstrate significant consistent CCF changes between the initial tumor and recurrence  
302 (Extended Data Fig. 10a). A notable exception was the *TP53* mutation CCF that increased over  
303 time ( $P = 0.037$ ) in IDHmut-noncodels, but not IDHwt gliomas ( $P = 0.13$ , Extended Data Fig.  
304 10b). We did not observe any differences in *IDH1* CCF over time among IDHmut-noncode  
305 tumors, possibly because the general trend of these tumors to increase in CCF is counteracted  
306 by the biological loss of relevance of mutant *IDH1* over time (Extended Data Fig. 10c). Indeed, a  
307 gross comparison of all shared mutation CCFs revealed an increase in recurrent IDHmut-  
308 noncode  
309 l tumors ( $P < 0.0001$ ), which may reflect increased clonality and a reduction in  
310 intratumoral heterogeneity (Extended Data Fig. 10d). In contrast, shared CCFs decreased in  
311 IDHwt tumors, potentially indicating a general increase in intratumoral heterogeneity at  
312 recurrence ( $P < 0.0001$ , Extended Data Fig. 10d). We confirmed that IDHmutant-noncode  
313 l CCF  
314 increases and IDHwt decreases were not biased by patients with high mutation burden through  
315 the classification of patient-specific shared mutation CCF change (Extended Data Fig. 10e).

314 We next investigated whether specific somatic alterations were acquired or lost over  
315 time. Gene-specific enrichment of many recurrence-only mutations was found in hypermutated  
316 tumors, but there was no enrichment for somatic gene alterations in non-hypermutators  
317 suggesting that glioma recurrence is not directed by particular sets of mutations (Extended Data  
318 Fig. 8b). Within subtypes we detected an enrichment in *CDKN2A* homozygous deletions (Fig.  
319 3a, Extended Data Fig. 8a) in recurrent IDHmutant-noncodels, which was corroborated by  
320 additional cell cycle gene alterations (focal gain of *CCND2*, *CDK4*, *CDK6*, and mutation or  
321 homozygous loss of *RB1*). Mutations in cell cycle checkpoint control genes are associated with  
322 genomic instability<sup>34</sup>. Therefore, we analyzed aneuploidy levels by determining the proportion of  
323 the genome that had undergone aneuploidy events (Extended Data Fig. 11a-b). We observed  
324 that IDHmutant-noncode  
325 l tumors had a higher level of aneuploidy at recurrence (Wilcoxon rank  
326 sum test  $P = 1.4E-06$  total aneuploidy,  $p = 8.6E-03$  arm-level aneuploidy, Extended Data Fig.  
327 11c-d) with tumors carrying acquired cell cycle gene alterations displaying the largest increases  
328 in aneuploidy ( $P = 7.6E-06$ ; Wilcoxon rank sum test, Fig. 3b). We reasoned that *CDKN2A*  
deletions may precede aneuploidy. Homozygous *CDKN2A* deletions had significantly higher

329 CCFs compared to average CNV CCF across the genome (as a surrogate for aneuploidy  
330 related copy number changes), suggesting that *CDKN2A* loss occurred prior to aneuploidy (Fig.  
331 3c). These alterations may hasten disease progression as patients with either cell cycle  
332 alterations or the largest increases in aneuploidy at recurrence demonstrated significantly  
333 shorter survival than patients without these alterations (log-rank test  $P < 0.0001$ , Fig. 3d). Taken  
334 together, the persistence of drivers over time and the paucity of consistent change imply that  
335 therapy does not result in selection of specific sets of molecular changes.

### 336 **Immunoediting activity in glioma**

337 We next investigated how the immune microenvironment affects evolutionary trajectories. The  
338 immune system may prune tumor cells carrying immunogenic (neo-)antigens, resulting in the  
339 selection of subclones capable of evading the immune response. Evidence of this  
340 immunoediting process has been shown in several cancer types, including glioma<sup>35-38</sup>, and  
341 suggests active immunosurveillance that may be therapeutically exploited<sup>39</sup>. We  
342 computationally predicted neoantigen-causing mutations<sup>40</sup>. As expected, the neoantigen load  
343 across the GLASS cohort was strongly correlated with exonic mutation burden (Spearman's  
344  $Rho = 0.89$ ), with 42% of nonsynonymous exonic mutations giving rise to neoantigens on  
345 average. This fraction did not significantly differ by glioma subtype or between initial and  
346 recurrent tumors ( $P > 0.05$ , Wilcoxon rank-sum test; Fig. 4a). The most common neoantigen  
347 arose from the clonal R132H mutation in *IDH1* and was present in 22 out of 88 *IDH*-mutant  
348 initial and recurrent tumors. Beyond mutations in *IDH1*, no mutations gave rise to a neoantigen  
349 found in more than three tumors at a given timepoint (Supplementary Table 6). Across the  
350 dataset, neoantigens and non-immunogenic mutations exhibited similar changes in cancer cell  
351 fractions between initial and recurrent tumors indicating a lack of neoantigen-specific selection  
352 processes over time (Extended Data Fig. 12a).

353 We then examined the extent to which immunoediting occurred by comparing each  
354 sample's observed neoantigen rate to an expected rate that was empirically derived from our  
355 dataset. The output of this approach is a normally distributed set of ratios centered at 1.  
356 Samples with an observed-to-expected neoantigen ratio  $< 1$  exhibit evidence of neoantigen  
357 depletion relative to the rest of the dataset, and thus are more likely to have been  
358 immunoedited. We found that none of the three glioma subtypes harbored observed-to-  
359 expected ratios that significantly differed from 1 ( $P > 0.05$ , one sample t-test), though *IDHwt*  
360 tumors exhibited significantly lower scores compared to *IDHmut-noncodels* (t-test,  $P = 0.04$ ; Fig.  
361 4b). We additionally did not observe an association between the observed-to-expected ratio and  
362 survival when adjusting for subtype and age (Wald test,  $P > 0.05$ ), nor was there a difference

363 between samples with neutral evolution dynamics compared to those exhibiting evidence of  
364 subclonal selection. When comparing samples longitudinally, we found that the observed-to-  
365 expected neoantigen ratio was strongly correlated between initial and recurrent tumors of each  
366 patient (Pearson's  $R = 0.73$ ,  $P = 5E-38$ ), suggesting that the neoantigen depletion level in the  
367 recurrence reflects that of the initial tumor (Fig. 4c).

368         Immunoediting is most likely to take place in the tumors with high cytolytic activity and  
369 low levels of immunosuppressive activity<sup>38</sup>. Hypermutators, which have high neoantigen loads,  
370 have previously been associated with highly cytolytic microenvironments<sup>37</sup>. However, we did  
371 not observe any differences in the observed-to-expected neoantigen ratio between  
372 hypermutated recurrent tumors and their initial counterparts, nor did we observe differences  
373 between hypermutated and non-hypermutated recurrent tumors, indicating that immunoediting  
374 activity is not related to the total number of mutations in a sample (Wilcoxon rank-sum test  $P >$   
375  $0.05$ ; Extended Data Fig. 12b). To more directly determine whether there were immunologic  
376 factors associated with neoantigen depletion, we analyzed CIBERSORT immune cell fractions  
377 from a subset of samples that had undergone expression profiling in a previous study ( $n = 84$   
378 from 42 tumor pairs)<sup>37,41</sup>. Initial tumors with an observed-to-expected neoantigen ratio  $>1$   
379 exhibited significantly higher levels of CD4+ T cells than those with a ratio  $< 1$ , while recurrent  
380 tumors with a ratio  $> 1$  exhibited significantly higher levels of macrophages, neutrophils, and  
381 significantly lower levels of plasma cells relative to those with ratio  $< 1$  ( $P < 0.05$ , Wilcoxon rank-  
382 sum test; Extended Data Fig. 12c).

383         While we did not detect many factors associated with the observed-to-expected  
384 neoantigen ratio, we did observe that the ratio was significantly associated with the total number  
385 of unique HLA loci in a patient (Spearman's  $Rho = 0.28$ ,  $P = 2E-9$ ), reflecting similar findings in  
386 lung cancer<sup>42</sup>. This may bias analyses comparing the ratio across patients. To determine  
387 whether immunoediting varies over time in a patient-agnostic manner, we compared the  
388 observed-to-expected neoantigen ratio derived from a sample's clonal mutations, which likely  
389 arose earlier in tumor evolution, to that derived from their subclonal mutations, which likely  
390 arose later. We did not observe a significant difference in the observed-to-expected neoantigen  
391 ratio of each patient's clonal and subclonal neoantigens, regardless of glioma subtype or  
392 whether the sample was an initial tumor or recurrence ( $P > 0.05$ , paired t-test; Fig. 4d).  
393 Together, these analyses suggest that neoantigens in glioma are not exposed to differing levels  
394 of selective pressure throughout their development.

## 395 **DISCUSSION**

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

405 The observation that treatment-induced hypermutation occurred across subtypes, but  
406 did not confer a detrimental effect on patient survival leaves the clinical significance of glioma  
407 hypermutation uncertain<sup>21-24,27</sup>. Future analyses that consider the number of therapy cycles and  
408 *MGMT* DNA methylation status will help to elucidate factors that predispose tumors to  
409 hypermutation and identify therapies that effectively exploit this phenotype's vulnerabilities (e.g.,  
410 high mutation burden). Acquired cell cycle alterations and aneuploidy in recurrent IDHmut-  
411 noncodel gliomas also provide a rationale to target these more aggressive phenotypes with  
412 CDK inhibitors<sup>43</sup> or with compounds that disrupt microtubule dynamics<sup>44</sup>. Finally, our analyses  
413 revealed that immunoediting activity does not vary in glioma over time, though we did observe  
414 variation between individual patients. Additional molecular and immunological data are needed  
415 to fully understand the impact this variability has on glioma evolution and to devise therapies  
416 directed at a glioma's immunogenicity<sup>17</sup>. To this end, we found that clonal neoantigens arising  
417 from the *IDH1* R132H mutation persisted from the initial tumor into the recurrence, justifying  
418 neoantigen vaccine approaches as treatments for initial and recurrent glioma<sup>45,46</sup>.

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

425

## 426 **ACKNOWLEDGEMENTS**

427 This work is dedicated to the memory of Simone Bischoff-Lardenoije and is made possible by  
428 the patients and their families whom generously contributed to this study. This work is supported  
429 by the National Brain Tumor Society, Oligo Research Fund; Cancer Center Support grants

430 P30CA16672 and P30CA034196; Cancer Prevention & Research Institute of Texas (CPRIT)  
431 grant number R140606; Agilent Technologies (R.G.W.V.); the National Institutes of Health-  
432 National Cancer institute for the following grants: NCI CA170278 (L.M.P., M.M.T., N.H.), NCI  
433 R01CA222146 (L.M.P, N.H.), NCI R01CA230031 (J.H.C., J.N.), NCI R01CA188288 (J.S.B.,  
434 R.B., P.B., K.L.L., A.C., A.E.S.), R01CA179044 (Antonio Iavarone), U54CA193313 (Antonio  
435 Iavarone). The National Brain Tumor Society (W.K.A.Y.; J.D.G). Brain Tumour Northwest tissue  
436 bank (including the Walton research tissue bank) is supported by the Sidney Driscoll  
437 Neuroscience Foundation and part of the Walton Centre and Lancashire Teaching Hospitals  
438 NHS Foundation Trusts (A.B., M.D.J.). This work was supported by a generous gift from the  
439 Dabbieri family (J.F.C.). Support is also provided by a Leeds Charitable Foundation grant  
440 (9R11/14-11 to LFS), University of Leeds Academic Fellowship (11001061) (L.F.S.) and  
441 Studentship (11061191) (G.T.) as well as Leeds Teaching Hospitals NHS Trust (Aruna  
442 Chakravarti, Azzam Ismail). The Leeds Multidisciplinary Research Tissue Bank staff was funded  
443 by the PPR Foundation and The University of Leeds (S.C.S.). Funds were received from The  
444 Brain Tumour Charity (C.W., Grants 10/136 & GN-000580, B.A.W., 200450). G.T. is funded by  
445 EKFS 2015\_Kolleg\_14. R01CA218144 (P.S.L, E.J.C, J.C. A.K.L.) and Strain for the Brain,  
446 Milwaukee, WI (P.S.L, E.J.C, J.C. A.K.L.). E.K is recipient of an MD-Fellowship by the  
447 Boehringer Ingelheim Fonds and is supported by the German National Academic Foundation.  
448 The Leeds Multidisciplinary Research Tissue Bank staff was funded by the PPR Foundation and  
449 part of the University of Leeds (S.C.S.). GLASS-Austria was funded by the Austrian Science  
450 Fund project KLI394 (A.W.). GLASS-Germany was funded by the German Ministry of Education  
451 and Research (BMBF) 031A425 (G.R., P.L.) and German Cancer Aid (DKH) 70-3163-Wi 3  
452 (M.W.). GLASS-NL receives support from KWF/Dutch Cancer Society project11026 (MCMK,  
453 PW, RGWV, PJF, JMN, MS, BAW). We thank the University of Colorado Denver Central  
454 Nervous System Biorepository (D.R.O.) for providing tissue samples. Sponsoring was also  
455 received from the National Institute of Neurological Disorders and Stroke (NINDS  
456 R01NS094615, R.G.), National Health and Medical Research Council project grant (A.M.D.).  
457 F.S.V. is supported by a postdoctoral fellowship from The Jane Coffin Childs Memorial Fund for  
458 Medical Research. F.P.B. is supported by the JAX Scholar program and the National Cancer  
459 Institute (K99 CA226387); K.C.J. is the recipient of an American Cancer Society Fellowship  
460 (130984-PF-17-141-01-DMC). We thank the Jackson Laboratory Clinical and Translation  
461 Support team for coordinating all data transfer agreements. We thank Matt Wimsatt for  
462 assistance in graphic design.

463 **CONFLICTS OF INTEREST**

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

#### 479 **CONTRIBUTIONS**

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

494

#### 495 **Methods**

496 **Data reporting** No statistical methods were used to predetermine sample size.

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

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

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

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



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

551  
552 **Computational pipelines** All pipelines were developed using snakemake 5.2.2<sup>47</sup>. Unless  
553 otherwise stated, all tools mentioned are part of the GATK 4 suite<sup>48</sup>. All data was collected at a  
554 central location (The Jackson Laboratory) and was analyzed using homogenous pipelines  
555 capable of processing both raw fastq files as well as re-process previously analyzed bam files.

556 **Alignment and pre-processing** Data pre-processing was conducted in accordance to the  
557 GATK Best Practices using GATK 4.0.10.1. Briefly, aligned BAM files were separated by read  
558 group, sanitized and stripped of alignments and attributes using 'RevertSam', giving one  
559 unaligned BAM (uBAM) file per readgroup. Uniform readgroups were assigned to uBAM files  
560 using 'AddOrReplaceReadgroups'. Similarly, unaligned fastq files were assigned uniformly  
561 designated readgroup attributes and converted to uBAM format using 'FastqToSam'. uBAM files  
562 underwent quality control using 'FastQC 0.11.7'. Sequencing adapters were marked using  
563 'MarkIlluminaAdapters'. uBAM files were finally reverted to interleaved fastq format using  
564 'SamToFastq', aligned to the b37 genome ('human\_g1k\_v37\_decoy') using 'BWA MEM 0.7.17',  
565 attributes were restored using 'MergeBamAlignment'. 'MarkDuplicates' was then used to merge  
566 aligned BAM files from multiple readgroups and to mark PCR and optical duplicates across  
567 identical sequencing libraries. Lastly, base recalibration was performed using 'BaseRecalibrator'  
568 followed by 'ApplyBQSR'. Coverage statistics were gathered using 'CollectWgsMetrics'.  
569 Alignment QC was performed running 'ValidateSamFile' on the final BAM file and QC results  
570 were inspected using 'MultiQC 1.6a0'<sup>49</sup>. A haplotype database for fingerprinting was generated  
571 using a modified version of the code on [https://github.com/naumanjaved/fingerprint\\_maps](https://github.com/naumanjaved/fingerprint_maps). The  
572 tool 'CrosscheckFingerprints' was used to confirm that all readgroups within a sample belong to  
573 the same individual, and that all samples from one individual match. Any mismatches were  
574 marked and excluded from further analysis.

575 **Variant detection** Variant detection was performed in accordance to the GATK Best practices  
576 using GATK 4.1.0.0. Germline variants were called from control samples using Mutect2 in  
577 artifact detection mode and pooled into a cohort-wide panel of normals. Somatic variants were  
578 subsequently called in individual tumor samples (single-sample mode) and in entire patients  
579 using GATK 4.1 Mutect2 in multi-sample mode. Mutect2 was given matched control samples,  
580 the aforementioned panel of normals and the gnomAD germline resource as additional controls.  
581 Cross-sample contamination was evaluated using 'GetPileupSummaries' and

582 'CalculateContamination' run for both tumor and matching control samples. Read orientation  
583 artifacts were evaluated using 'CollectF1R2Counts' and 'LearnReadOrientationModel'. Somatic  
584 likelihood, read orientation, sequence context, germline and contamination filters were applied  
585 using 'FilterMutectCalls'.

586 **Variant post-processing** BCFTools 1.9 was used to normalize, sort and index variants<sup>50</sup>. A  
587 consensus VCF was generated from all variants in the cohort, removing any duplicate variants.  
588 The consensus VCF file was annotated using GATK 4.1 Funcotator and the v1.6.20190124s  
589 annotation data source. Allele frequencies (AFs) from multi-sample Mutect2 were used to  
590 compare AFs between related samples. Multi-sample Mutect2 calls and filters mutations across  
591 a patient as a whole and does not determine mutation calls in a single samples. Single-sample  
592 mutation calls were overlaid on the multi-sample calls to infer whether variants were called in  
593 individual samples. Single-sample called variants that were not present in the multi-sample  
594 callset were discarded.

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

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

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

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

632 **Copy number calling** A copy number caller loosely based on GATK 'CallCopyRatioSegments'  
633 (which in turn is based off of ReCapSeg) and GISTIC was implemented to call both arm-level  
634 and high-level copy number changes, respectively<sup>51,52</sup>.

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

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

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

654 **dNdScv** The R package dNdScv<sup>53</sup> (<https://github.com/im3sanger/dndscv>) was run using the  
655 default and recommended parameters for all mutations in initial tumor samples, recurrent tumor  
656 samples, and for each mutational fraction (unique to initial, unique to recurrent and shared). All  
657 analyses were conducted separately within the three main tumor subtypes.

658 **Aneuploidy calculation** The most reductive metric of aneuploidy was computed by taking the  
659 size of all non-neutral segments divided by the size of all segments. The resulting aneuploidy  
660 value indicates the proportion of the segmented genome that is non-diploid.

661 In parallel, an arm-level aneuploidy score modeled after a previously described method was  
662 computed<sup>54</sup>. Briefly, adjacent segments with identical arm-level calls (-1, 0 or 1) were merged  
663 into a single segment with a single call. For each merged/reduced segment, the proportion of  
664 the chromosome arm it spans was calculated. Segments spanning greater than 80% of the arm  
665 length resulted in a call of either -1 (loss), 0 (neutral) or +1 (gain) to the entire arm, or NA if no  
666 contiguous segment spanned at least 80% of the arm's length. For each sample the number of  
667 arms with a non-neutral event was finally counted. The resulting aneuploidy score is a positive

668 integer with a minimum value of 0 (no chromosomal arm-level events detected) and a maximum  
669 value of 39 (total number of autosomal chromosome arms excluding the short arms for  
670 chromosomes 13, 14, 15, 21, and 22).

671 **Estimates of evolutionary pressures** Evolutionary pressures were evaluated both by variant  
672 status and glioma subtype using the neutralitytestr algorithm as previously described (R-  
673 package: neutralitytestr version: 0.0.2, <https://github.com/marcjwilliams1/neutralitytestr>)<sup>31</sup>.  
674 Individual variant allele frequency vectors were merged at the level of glioma subtype by variant  
675 status. Only mutations found in copy-neutral regions should be included in these analyses.  
676 For all else, default parameters were used. Merged VAF distributions were deemed to be  
677 selected when the neutral null hypothesis was rejected using several metrics. Tests for  
678 neutrality required that both  $R^2$  values  $< 0.98$  and the area between the two curves of 1) merged  
679 VAF data and 2) a normalized distribution expected under neutrality to be significantly different.

680 The SubclonalSelection algorithm was applied to GLASS mutation data to measure the  
681 selection strength in individual tumor samples (Julia package: SubclonalSelection,  
682 <https://github.com/marcjwilliams1/SubClonalSelection.jl>)<sup>16</sup>. Patients that had samples at both  
683 timepoints with a TITAN-defined purity estimate  $\geq 0.5$  and  $\geq 25$  subclonal mutations in non-  
684 diploid regions were included. Mean coverage across all mutations was used as the  
685 “read\_depth” input parameter and the model was run with the recommended  $10^6$  iterations and  
686 1000 particles. Samples were classified as neutral or selected based on the model that had the  
687 highest probability, in line with the prior applications to TCGA data<sup>16</sup>. Classification based on the  
688 highest model probability yielded stable results there was not a significant change in proportions  
689 when setting a higher classification probability threshold ( $P > 0.05$ , Pearson’s Chi-square test,  
690 for both probability thresholds of 0.6 and 0.7). At all three probability thresholds (0.5, 0.6, and  
691 0.7), Kaplan-Meier survival analyses between selection at recurrence and overall survival  
692 continued to indicate that patients with IDHwt tumors that were selected had a worse overall  
693 survival ( $P = 0.03$  (n=81),  $P = 0.01$  (n=66),  $P = 0.01$  (n=56) respectively).

694 **Mutation clonality** Each patient’s clonal architecture was inferred using PyClone (version  
695 0.13.1) by grouping SNVs into clonal clusters (<https://github.com/aroeth85/pyclone>)<sup>55</sup>. The  
696 patient-level input mutation matrix was reduced by limiting to sites with at least 30x coverage  
697 across all samples. PyClone was subsequently ran using a binomial density model, connected  
698 initiation, and 10000 iterations. Sample purities were provided for each patient and parental  
699 copy number (minor and major allele counts) from TITAN were given. PyClone results were  
700 post-processed using a burn-in of 1000, thin of 1, minimum cluster size of 2 and a maximum  
701 number of clusters per patient of 12. Individual mutations were determined to be clonal if the  
702 PyClone cancer cell fraction (CCF) values were  $\geq 0.5$ , subclonal for mutations with CCF  $\geq 0.1$   
703 and CCF  $< 0.5$ , mutations were considered non-clonal when CCF  $< 0.1$  as previously described  
704 <sup>56</sup>.

705 **CNV clonality** Allele specific copy number, tumor purity and ploidy estimates were derived  
706 using a probabilistic model (TITAN, version 1.19.1) for both whole genome and whole exome  
707 sequencing samples <sup>57</sup>. TITAN was supplied with the tumor denoised readcounts output by  
708 GATK DenoiseReadCounts and the tumor allelic counts at loci found to be heterozygous in  
709 control samples output by ModelSegments. An ‘alphaK’ (and ‘alphaKHigh’) parameter of 2500  
710 and 10000 was used for exomes and genomes, respectively. The patient sex was provided in

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

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

719 **Neoantigen analyses** Neoantigens in this analysis were defined as all 8-11-mer peptides that  
 720 arose from an exonic nonsynonymous SNV or indel and bound their respective patient's HLA  
 721 class I molecules at a binding affinity score (IC50) that was  $\leq 500$  nM and better than or equal to  
 722 the wild-type form of the peptide. Each patient's 4-digit HLA class I types were inferred using  
 723 OptiType (version 1.3.1, <https://github.com/FRED-2/OptiType>) run on each patient's matched  
 724 normal sample<sup>58</sup>. VCF files for each tumor sample were annotated using Variant Effect Predictor  
 725 (ensembl) with the Downstream and Wildtype plugins. Neoantigens from these VCFs were then  
 726 called using pVACseq (version 4.0.10, <https://github.com/griffithlab/pVAC-Seq>)<sup>40</sup> run using  
 727 netMHCpan (version 2.8, <http://www.cbs.dtu.dk/services/NetMHCpan-2.8/>)<sup>59</sup>. For each  
 728 pVACseq run, epitope length was set to 8, 9, 10, or 11, minimum binding affinity fold-change  
 729 was set to 1, and downstream sequence length was set to full, with default parameters used for  
 730 all other settings.

731 Downstream neoantigen analyses were performed using the pVACseq output linked to its  
 732 respective mutation information. Neoantigen-causing mutations were defined as all mutations  
 733 that gave rise to at least one neoantigen. The observed-to-expected neoantigen ratio was  
 734 calculated using a previously developed approach that compares each tumor's observed  
 735 neoantigen rate to an empirically derived expected rate that assumes no selection against  
 736 neoantigen-causing mutations<sup>38</sup>: From the gold set samples in the GLASS cohort ( $n = 222$ ),  
 737 define  $\bar{N}_s$  to be the expected number of nonsynonymous missense SNVs per synonymous SNV  
 738 with trinucleotide context  $s$ .  $\bar{B}_s$  is then defined as the expected number of neoantigen-generating  
 739 missense SNVs per nonsynonymous missense SNV with trinucleotide context  $s$ . For a given  
 740 sample  $i$ , define  $Y_i$  as the sample's set of synonymous SNVs and  $s(m)$  to be a synonymous  
 741 SNV with trinucleotide context  $m$ . The expected number of nonsynonymous missense SNVs,  
 742  $N_{pred}$ , and neoantigen-causing mutations,  $B_{pred}$ , can then be calculated as follows:

$$743 \quad N_{pred,i} = \sum_{m \in Y_i} \bar{N}_{s(m)}$$

$$744 \quad B_{pred,i} = \sum_{m \in Y_i} \bar{N}_{s(m)} \bar{B}_{s(m)}$$

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

749

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

750  
751  
752  
753  
754

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

755  
756

**Immune cell analyses** CIBERSORT relative immune cell fraction data used in downstream neoantigen analyses were downloaded from a previous publication<sup>37</sup>.

757  
758  
759

**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.

760  
761  
762

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

763  
764

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

765

## 766 References

767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789

- 1 Barthel, F. P., Wesseling, P. & Verhaak, R. G. W. Reconstructing the molecular life history of gliomas. *Acta Neuropathol* **135**, 649-670, doi:10.1007/s00401-018-1842-y (2018).
- 2 Sturm, D. *et al.* Paediatric and adult glioblastoma: multiform (epi)genomic culprits emerge. *Nat Rev Cancer* **14**, 92-107, doi:10.1038/nrc3655 (2014).
- 3 Bettgowda, C. *et al.* Mutations in CIC and FUBP1 contribute to human oligodendroglioma. *Science* **333**, 1453-1455, doi:10.1126/science.1210557 (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, doi:10.1101/gad.213686.113 (2013).
- 5 Cancer Genome Atlas Research, N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061-1068, doi:10.1038/nature07385 (2008).
- 6 Ceccarelli, M. *et al.* Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* **164**, 550-563, doi:10.1016/j.cell.2015.12.028 (2016).
- 7 TCGA\_Network *et al.* Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med* **372**, 2481-2498, doi:10.1056/NEJMoa1402121 (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, doi:10.1016/j.ccr.2009.12.020 (2010).
- 9 Yan, H. *et al.* IDH1 and IDH2 mutations in gliomas. *N Engl J Med* **360**, 765-773, doi:10.1056/NEJMoa0808710 (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, doi:10.1111/bpa.12171 (2014).

790 11 Louis, D. N. *et al.* The 2016 World Health Organization Classification of Tumors of the Central  
791 Nervous System: a summary. *Acta Neuropathol* **131**, 803-820, doi:10.1007/s00401-016-1545-1  
792 (2016).

793 12 Venteicher, A. S. *et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant  
794 gliomas by single-cell RNA-seq. *Science* **355**, doi:10.1126/science.aai8478 (2017).

795 13 Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary  
796 glioblastoma. *Science* **344**, 1396-1401, doi:10.1126/science.1254257 (2014).

797 14 Snuderl, M. *et al.* Mosaic amplification of multiple receptor tyrosine kinase genes in  
798 glioblastoma. *Cancer Cell* **20**, 810-817, doi:10.1016/j.ccr.2011.11.005 (2011).

799 15 Sottoriva, A. *et al.* Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary  
800 dynamics. *Proc Natl Acad Sci U S A* **110**, 4009-4014, doi:10.1073/pnas.1219747110 (2013).

801 16 Williams, M. J. *et al.* Quantification of subclonal selection in cancer from bulk sequencing data.  
802 *Nat Genet* **50**, 895-903, doi:10.1038/s41588-018-0128-6 (2018).

803 17 Nejo, T. *et al.* Reduced Neoantigen Expression Revealed by Longitudinal Multiomics as a Possible  
804 Immune Evasion Mechanism in Glioma. *Cancer Immunol Res*, doi:10.1158/2326-6066.CIR-18-  
805 0599 (2019).

806 18 Consortium, G. Glioma through the looking GLASS: molecular evolution of diffuse gliomas and  
807 the Glioma Longitudinal Analysis Consortium. *Neuro Oncol* **20**, 873-884,  
808 doi:10.1093/neuonc/noy020 (2018).

809 19 Hu, H. *et al.* Mutational Landscape of Secondary Glioblastoma Guides MET-Targeted Trial in  
810 Brain Tumor. *Cell* **175**, 1665-1678 e1618, doi:10.1016/j.cell.2018.09.038 (2018).

811 20 Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-  
812 421, doi:10.1038/nature12477 (2013).

813 21 Wang, J. *et al.* Clonal evolution of glioblastoma under therapy. *Nat Genet* **48**, 768-776,  
814 doi:10.1038/ng.3590 (2016).

815 22 Kim, H. *et al.* Whole-genome and multisector exome sequencing of primary and post-treatment  
816 glioblastoma reveals patterns of tumor evolution. *Genome Res* **25**, 316-327,  
817 doi:10.1101/gr.180612.114 (2015).

818 23 Johnson, B. E. *et al.* Mutational analysis reveals the origin and therapy-driven evolution of  
819 recurrent glioma. *Science* **343**, 189-193, doi:10.1126/science.1239947 (2014).

820 24 Hunter, C. *et al.* A hypermutation phenotype and somatic MSH6 mutations in recurrent human  
821 malignant gliomas after alkylator chemotherapy. *Cancer Res* **66**, 3987-3991, doi:10.1158/0008-  
822 5472.CAN-06-0127 (2006).

823 25 Jolly, C. & Van Loo, P. Timing somatic events in the evolution of cancer. *Genome Biol* **19**, 95,  
824 doi:10.1186/s13059-018-1476-3 (2018).

825 26 Turajlic, S., Sottoriva, A., Graham, T. & Swanton, C. Resolving genetic heterogeneity in cancer.  
826 *Nat Rev Genet*, doi:10.1038/s41576-019-0114-6 (2019).

827 27 Choi, S. *et al.* Temozolomide-associated hypermutation in gliomas. *Neuro Oncol* **20**, 1300-1309,  
828 doi:10.1093/neuonc/noy016 (2018).

829 28 Baumert, B. G. *et al.* Temozolomide chemotherapy versus radiotherapy in high-risk low-grade  
830 glioma (EORTC 22033-26033): a randomised, open-label, phase 3 intergroup study. *Lancet Oncol*  
831 **17**, 1521-1532, doi:10.1016/S1470-2045(16)30313-8 (2016).

832 29 Buckner, J. C. *et al.* Radiation plus Procarbazine, CCNU, and Vincristine in Low-Grade Glioma. *N*  
833 *Engl J Med* **374**, 1344-1355, doi:10.1056/NEJMoa1500925 (2016).

834 30 Yap, T. A., Gerlinger, M., Futreal, P. A., Pusztai, L. & Swanton, C. Intratumor heterogeneity:  
835 seeing the wood for the trees. *Sci Transl Med* **4**, 127ps110, doi:10.1126/scitranslmed.3003854  
836 (2012).

837 31 Williams, M. J., Werner, B., Barnes, C. P., Graham, T. A. & Sottoriva, A. Identification of neutral  
838 tumor evolution across cancer types. *Nat Genet* **48**, 238-244, doi:10.1038/ng.3489 (2016).

839 32 Korber, V. *et al.* Evolutionary Trajectories of IDH(WT) Glioblastomas Reveal a Common Path of  
840 Early Tumorigenesis Instigated Years ahead of Initial Diagnosis. *Cancer Cell* **35**, 692-704 e612,  
841 doi:10.1016/j.ccell.2019.02.007 (2019).

842 33 deCarvalho, A. C. *et al.* Discordant inheritance of chromosomal and extrachromosomal DNA  
843 elements contributes to dynamic disease evolution in glioblastoma. *Nat Genet* **50**, 708-717,  
844 doi:10.1038/s41588-018-0105-0 (2018).

845 34 Giam, M. & Rancati, G. Aneuploidy and chromosomal instability in cancer: a jackpot to chaos.  
846 *Cell Div* **10**, 3, doi:10.1186/s13008-015-0009-7 (2015).

847 35 Marty, R., Thompson, W. K., Salem, R. M., Zanetti, M. & Carter, H. Evolutionary Pressure against  
848 MHC Class II Binding Cancer Mutations. *Cell* **175**, 416-428 e413, doi:10.1016/j.cell.2018.08.048  
849 (2018).

850 36 McGranahan, N. *et al.* Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell*  
851 **171**, 1259-1271 e1211, doi:10.1016/j.cell.2017.10.001 (2017).

852 37 Wang, Q. *et al.* Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with  
853 Immunological Changes in the Microenvironment. *Cancer Cell* **32**, 42-56 e46,  
854 doi:10.1016/j.ccell.2017.06.003 (2017).

855 38 Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties  
856 of tumors associated with local immune cytolytic activity. *Cell* **160**, 48-61,  
857 doi:10.1016/j.cell.2014.12.033 (2015).

858 39 Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoediting: from  
859 immunosurveillance to tumor escape. *Nat Immunol* **3**, 991-998, doi:10.1038/ni1102-991 (2002).

860 40 Hundal, J. *et al.* pVAC-Seq: A genome-guided in silico approach to identifying tumor  
861 neoantigens. *Genome Med* **8**, 11, doi:10.1186/s13073-016-0264-5 (2016).

862 41 Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat*  
863 *Methods* **12**, 453-457, doi:10.1038/nmeth.3337 (2015).

864 42 Rosenthal, R. *et al.* Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**,  
865 479-485, doi:10.1038/s41586-019-1032-7 (2019).

866 43 Raub, T. J. *et al.* Brain Exposure of Two Selective Dual CDK4 and CDK6 Inhibitors and the  
867 Antitumor Activity of CDK4 and CDK6 Inhibition in Combination with Temozolomide in an  
868 Intracranial Glioblastoma Xenograft. *Drug Metab Dispos* **43**, 1360-1371,  
869 doi:10.1124/dmd.114.062745 (2015).

870 44 van den Bent, M. *et al.* Efficacy of depatuxizumab mafodotin (ABT-414) monotherapy in patients  
871 with EGFR-amplified, recurrent glioblastoma: results from a multi-center, international study.  
872 *Cancer Chemother Pharmacol* **80**, 1209-1217, doi:10.1007/s00280-017-3451-1 (2017).

873 45 Keskin, D. B. *et al.* Neoantigen vaccine generates intratumoral T cell responses in phase Ib  
874 glioblastoma trial. *Nature* **565**, 234-239, doi:10.1038/s41586-018-0792-9 (2019).

875 46 Schumacher, T. *et al.* A vaccine targeting mutant IDH1 induces antitumour immunity. *Nature*  
876 **512**, 324-327, doi:10.1038/nature13387 (2014).

877 47 Koster, J. & Rahmann, S. Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics*  
878 **34**, 3600, doi:10.1093/bioinformatics/bty350 (2018).

879 48 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome  
880 Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11 10 11-33,  
881 doi:10.1002/0471250953.bi1110s43 (2013).

882 49 Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for  
883 multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048,  
884 doi:10.1093/bioinformatics/btw354 (2016).



- 885 50 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079,  
886 doi:10.1093/bioinformatics/btp352 (2009).
- 887 51 Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of  
888 focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41, doi:10.1186/gb-  
889 2011-12-4-r41 (2011).
- 890 52 Beroukhi, R. *et al.* Assessing the significance of chromosomal aberrations in cancer:  
891 methodology and application to glioma. *Proc Natl Acad Sci U S A* **104**, 20007-20012,  
892 doi:10.1073/pnas.0710052104 (2007).
- 893 53 Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**,  
894 1029-1041 e1021, doi:10.1016/j.cell.2017.09.042 (2017).
- 895 54 Taylor, A. M. *et al.* Genomic and Functional Approaches to Understanding Cancer Aneuploidy.  
896 *Cancer Cell* **33**, 676-689 e673, doi:10.1016/j.ccell.2018.03.007 (2018).
- 897 55 Roth, A. *et al.* PyClone: statistical inference of clonal population structure in cancer. *Nat*  
898 *Methods* **11**, 396-398, doi:10.1038/nmeth.2883 (2014).
- 899 56 Turajlic, S. *et al.* Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx  
900 Renal. *Cell* **173**, 581-594 e512, doi:10.1016/j.cell.2018.03.057 (2018).
- 901 57 Ha, G. *et al.* TITAN: inference of copy number architectures in clonal cell populations from tumor  
902 whole-genome sequence data. *Genome Res* **24**, 1881-1893, doi:10.1101/gr.180281.114 (2014).
- 903 58 Szolek, A. *et al.* OptiType: precision HLA typing from next-generation sequencing data.  
904 *Bioinformatics* **30**, 3310-3316, doi:10.1093/bioinformatics/btu548 (2014).
- 905 59 Hoof, I. *et al.* NetMHCpan, a method for MHC class I binding prediction beyond humans.  
906 *Immunogenetics* **61**, 1-13, doi:10.1007/s00251-008-0341-z (2009).

907

## 908 Figure Legends

909 **Fig. 1 | Temporal changes in glioma mutational burden and processes.** **a.** Each column  
910 represents a single patient ( $n = 222$ ) at two separate timepoints grouped by glioma subtype and  
911 ordered left-to-right by decreasing mutation frequency at recurrence. Top, mutation frequency  
912 differences between initial and recurrent tumors. Blue dotted line indicates increased mutation  
913 frequency while a red dotted line indicates decreased mutational frequency. Stacked bar plot  
914 reflects the proportion of total mutations shared (mustard), private to initial (magenta), or private  
915 to recurrence (blue). Clinical information including hypermutation status, therapy, and grade  
916 changes. **b.** Stacked bar plot ( $n=219$ ) indicating the dominant mutational signature among initial,  
917 recurrent and shared mutation fractions stratified by glioma subtype. **c.** The proportion of glioma  
918 recurrences with alkylating agent-related hypermutation, grouped by glioma subtype. Fisher's  
919 exact test was used to compare proportions between subtypes. **d.** Kaplan-Meier curve depicting  
920 overall survival in hypermutant (red) versus non-hypermutant (blue) alkylating agent treated  
921 patients amongst IDHwt (left,  $n = 99$ ) and IDHmut-noncodel (right,  $n = 32$ ) tumors. Log-rank test  
922 P-values are shown.

923 **Fig. 2 | Quantifying selective pressures during glioma evolution.** **a.** Schematic depiction of  
924 cancer cell fraction (CCF) values during tumor evolution indicating clonality and associated  
925 relative timing. **b.** Comparison of PyClone clusters ranked by CCF in matched initial and recurrent  
926 tumors. **c. Left:** dN/dS ratio for all variants (i.e. global) in initial and recurrent tumors for each  
927 subtype. Hypermutators were not included ( $n = 187$ ). Dots represent the global dN/dS ratio with

928 associated Wald confidence intervals. *Right*: global dN/dS ratios for variant fractions per subtype.  
929 **d.** Cumulative distribution of subclonal mutations by their inverse variant allele frequency.  
930 Mutations were separated by timepoint, variant fraction, and glioma subtype. Deviation from a  
931 linear relationship, significant Kolmogorov-Smirnov P-values and  $R^2$  below 0.98 indicate  
932 selection. **e.** Sankey plot indicating the breakdown of SubClonalSelection evolutionary modes by  
933 subtype and therapy ( $n = 104$ ). The sizes of the bands reflect sample sizes and band colors  
934 highlight the glioma subtype. Gray coloring reflects instances when treatment information was not  
935 available. **f.** Kaplan-Meier curve showing survival differences between IDHwt recurrent tumors  
936 demonstrating selection ( $n = 39$ ) compared with neutrally evolving tumors ( $n = 44$ ). Log-rank P-  
937 value is indicated.

938 **Fig. 3 | Patterns of glioma driver frequencies over time.** **a.** Driver dynamics for SNVs  
939 nominated by the dNdScv and CNVs nominated by GISTIC ( $n = 222$ ). Each column represents a  
940 single patient at two separate time points stratified by subtype and ordered left-to-right by the  
941 number of driver alterations. The degree of aneuploidy difference (recurrence – initial) offers a  
942 summary metric for increases ( $> 0$ ) or decreases ( $< 0$ ) in aneuploidy at recurrence. Variants are  
943 marked and different shapes indicate whether a variant was shared or private. The variant type  
944 is depicted by its color. Stacked bar plots accompanying each gene/arm provide cohort-level  
945 proportions for whether the alteration was shared, lost, or acquired. **b.** Aneuploidy comparison in  
946 matching initial and recurrent IDHmut-noncodeI tumors. **c.** Within-sample CCF comparison of  
947 *CDKN2A* homozygous deletion (homdel) to genome-wide CCF as a proxy for aneuploidy. A  
948 relative higher CCF indicates temporal precedence. Wilcoxon signed-rank test P-value is  
949 indicated. **d.** Kaplan-Meier curve comparing survival in IDHmut-noncodeI tumors with an alteration  
950 in the cell cycle, acquired aneuploidy, or both (shades of red) versus unaltered IDHmut-noncodeI  
951 tumors (blue). Log-rank P-value is shown.

952 **Fig. 4 | Neoantigen selection during tumor progression.** **a.** Mean proportion of coding  
953 mutations giving rise to neoantigens (neoantigens/nonsynonymous) stratified by glioma subtype  
954 and timepoint ( $n = 222$ ). Error bars represent standard deviation. **b.** Boxplot depicting the  
955 distribution of observed to expected neoantigen ratios in the GLASS cohort stratified by glioma  
956 subtype. P-value was calculated using the Wilcoxon rank-sum test. Each box spans quartiles,  
957 with the lines representing the median ratio for each group. Whiskers represent absolute range,  
958 excluding outliers. **c.** Scatterplot depicting the association between the observed-to-expected  
959 neoantigen ratio in a patient's initial versus recurrent tumor. Each point represents a single  
960 patient. R represents Pearson correlation coefficient. Panels b and c only include samples with at  
961 least 3 neoantigens in the initial and recurrent tumors ( $n = 131, 63,$  and  $24$  for IDHwt, IDHmut-  
962 noncodeI, and IDHmut-codeI, respectively). **d.** Ladder plot depicting the difference in observed-  
963 to-expected neoantigen ratio between a tumor's clonal and subclonal neoantigens. Each set of  
964 points connected by a line represents one tumor. Tumors are stratified by whether they were a  
965 patient's initial or recurrent tumor. Lines are colored by each patient's glioma subtype. Panel d  
966 only includes samples with at least 3 clonal neoantigens and at least 3 subclonal neoantigens in  
967 both the initial and recurrent tumors ( $n = 35, 20$  and  $9$  for IDHwt, IDHmut-noncodeI, and IDHmut-  
968 codeI, respectively). P-value was calculated using a paired two-sided t-test. Colors in each panel  
969 represent the glioma subtype and are denoted at the bottom of the figure.

970 **Extended Data Fig. 1 | Sample Selection.** **a.** Quality control workflow steps identifying all  
971 GLASS samples available as a resource and the identification of the highest quality set of patient  
972 pairs ( $n = 222$ ) used for the presented mutational and copy number analyses. **b.** Additional  
973 available datasets.

974 **Extended Data Fig. 2 | Mutation burden by time point and subtype.** **a.** Boxplots and paired  
975 lines depicting coverage adjusted mutation frequencies in initial and matched recurrent samples  
976 across three subtypes. Wilcoxon signed-rank test P-values and sample sizes are indicated. **b.**  
977 Bee swarm plot depicting coverage adjusted mutation frequencies in fractions by subtype.  
978 Dashed line indicates the mean. One-way ANOVA P-values comparing three subtypes are  
979 indicated. **c.** Scatter plot showing the relationship between age at diagnosis and coverage  
980 adjusted mutation burdens by subtype and fraction. Linear model P-values are indicated and were  
981 adjusted by subtype. **d.** Similar to the analysis presented in **c**, but showing the relationship  
982 between time to recurrence and coverage adjusted mutation burdens.

983 **Extended Data Fig. 3 | Mutational signatures by fraction and subtype.** **a.** Correlation plot  
984 showing the Pearson's chi-squared ( $X^2$ ) residuals for each signature by fraction and subtype. A  
985  $X^2$  was performed for each subtype and P-values are indicated. Positive residuals (blue) indicate  
986 a positive correlation, whereas negative residuals (red) indicate an anticorrelation. The point size  
987 reflects the contribution to  $X^2$  estimate. **b.** The same ordered of patients as Fig. 1a along with  
988 relevant clinical information is provided alongside the fraction-specific mutational signatures.  
989 PyClone mutational clusters are also presented.

990 **Extended Data Fig. 4 | Hypermutator clonality.** **a.** Bar plots represent counts of recurrence-  
991 only mutations per hypermutator tumor that were known to receive alkylating agent therapy and  
992 were successfully run through the PyClone algorithm. Colors indicate mutation clonality and color  
993 intensity indicates whether the mutations resulted in coding changes. **b.** Kaplan-Meier curve  
994 comparing alkylating agent-treated patients with IDHmut-noncodel hypermutator tumors that were  
995 predominantly clonal ( $n = 8$ ), predominantly subclonal ( $n = 7$ ), versus IDHmut-noncodel non-  
996 hypermutators known to be treated with alkylating agents and had available PyClone data ( $n =$   
997 17). Log-rank P-value is shown.

998 **Extended Data Fig. 5 | Clonal structure evolution over time.** **a.** The minimum cancer cell  
999 fraction of the most persistent (shared between initial and recurrence) PyClone cluster. **b.**  
1000 Comparison of PyClone clusters ranked by CCF in matched initial and recurrent tumors, as Fig.  
1001 2b but separated by subtype. **c-d.** Examples of cluster CCF dynamics over time in three separate  
1002 samples, including **(c)** two multi-timepoint samples **(d)** and one multi-sector sample. These  
1003 additional data are available in the GLASS resource, but only two time-separated samples were  
1004 used throughout the manuscript to ensure clarity.

1005 **Extended Data Fig. 6 | Variant allele fraction distribution (a)** Non-hypermutator variant allele  
1006 fraction distributions for copy neutral variants in coding regions ( $n = 181$  patients). Variants are  
1007 separated by subtype, fraction, and also whether the variant was non-synonymous or  
1008 synonymous mutation in a coding region.  $R^2$  goodness-of-fit measure and associated P-values  
1009 are shown for both mutation types. Note that this data considers only the coding portion of genome  
1010 while Fig. 2d presents both coding and non-coding. **(b)** The cumulative distribution of the

1011 subclonal mutations in copy-neutral regions for hypermutators (n = 31 patients). For each variant  
1012 fraction and subtype, the R<sup>2</sup> goodness-of-fit measure and P-values are shown.

1013 **Extended Data Fig. 7 | Driver gene nomination.** **a.** Local (gene-wise) dNdScv estimates by  
1014 subtype (rows) and fraction (columns). Genes are sorted by Q-value and P-value. The Q-value is  
1015 shown in color, whereas the P-value is indicated in light gray. The Q-value threshold of 0.05 is  
1016 indicated by a horizontal red line. **b.** GISTIC significant amplification (red) and deletion (blue) plots  
1017 in initial (left) and recurrent tumors (right). Chromosomal locations are ordered on the y-axis, Q-  
1018 values are shown on the x-axis, and selected drivers are indicated by their chromosomal location  
1019 on the right.

1020 **Extended Data Fig. 8 | Driver acquisition over time** **a.** Tabulated numbers of SNV (top) and  
1021 CNV (bottom) driver events that were shared, initial-only, or recurrence-only. P-values were  
1022 obtained using a two-sided Fisher test comparing the initial-only fraction to the recurrence-only  
1023 fraction testing for acquisition. **b.** One-sided Fisher test comparing the initial-only fraction to the  
1024 recurrence-only fraction amongst previously implicated glioma drivers testing for driver  
1025 acquisition. P-values were adjusted for multiple testing using the FDR (x-axis). Hypermutators  
1026 (red) and non-hypermutators (black) were separately analyzed.

1027 **Extended Data Fig. 9 | Intra-tumor CCF comparison.** Ladder plots comparing the CCF of co-  
1028 occurring drivers in single tumor samples. The color of the lines and points indicates whether the  
1029 sample shown is an initial (brown) or recurrent (green) tumor. Two-sided Wilcoxon rank-sum test  
1030 P-values are shown for all initial samples, all recurrent samples, as well as all samples (black).

1031 **Extended Data Fig. 10 | Between time point intra-patient CCF comparison.** **a.** Driver-gene  
1032 CCF comparison between initial and matched recurrences. Lines are colored by variant  
1033 classification. Two-sided Wilcoxon rank-sum test P-values are shown. **b.** *TP53* CCF by subtype,  
1034 otherwise as in **(a)**. **c.** *IDH1* CCF by subtype, otherwise as in **(a)**. **d.** Ladder plot visualizing CCF  
1035 change across all SNVs between initial and recurrent tumors, separated by subtype. Wilcoxon  
1036 rank-sum test was used to test for differences between time points. **e.** Initial and recurrent  
1037 mutations in each patient were compared using a Wilcoxon rank-sum test. Bar plot with counts of  
1038 patients in each subtype are shown. Patients lacking significant change are shown in yellow,  
1039 those with a significant increase or decrease are shown in dark and light blue, respectively.

1040

1041 **Extended Data Fig. 11 | Aneuploidy calculation** **a.** Heatmap displaying the chromosomal arm-  
1042 level events (x-axis) with patients represented in each row. Patients are placed in the same order  
1043 for both the initial (left) and recurrence (right). White space was inserted as a break between the  
1044 three subtypes. **b.** Distribution of total aneuploidy difference. Acquired aneuploidy determination  
1045 (upper-quartile) indicated with a red line. **c.** Comparison of aneuploidy score between initial and  
1046 recurrent tumors separated by subtype **d.** As **(c)**, comparing aneuploidy value.

1047 **Extended Data Fig. 12 | Neoantigen evolution and cellular analysis** **a.** Bar plots representing  
1048 the number of shared mutations that give rise to neoantigens (top row, “immunogenic”) and those  
1049 that do not give rise to neoantigens (bottom row, “non-immunogenic”) stratified by longitudinal

1050 clonality (“(clonality in initial)-(clonality in recurrence)”) and further separated by subtype.  
1051 Percentage of longitudinal clonality per subtype and mutation immunogenicity are presented  
1052 above the respective bars. **b.** *Left:* Ladder plot depicting the difference in observed-to-expected  
1053 neoantigen ratio between the initial and recurrent tumors of patients with hypermutated tumors at  
1054 recurrence. Each set of points connected by a line represents one tumor ( $n = 70$ ). *Right:* Boxplot  
1055 depicting the distribution of observed to expected neoantigen ratios in recurrent tumors stratified  
1056 by hypermutator status ( $n = 35$  and  $183$  for hypermutators and non-hypermutators, respectively).  
1057 Each box spans quartiles, with the lines representing the median ratio for each group. Whiskers  
1058 represent absolute range, excluding outliers. P-values for panel b were calculated using a paired  
1059 and unpaired two-sided t-test, respectively. **c.** Stacked bar plots depicting the average relative  
1060 fraction of 11 CIBERSORT cell types in the neoantigen depleted ( $< 1$ ) and non-depleted ( $> 1$ )  
1061 initial and recurrent tumor subgroups. Asterisks to the right of each plot indicate a significant  
1062 difference ( $P < 0.05$ , Wilcoxon rank-sum test) between the depleted and non-depleted groups for  
1063 the noted cell type at that time.

1064