

Structural variants shape driver combinations and outcomes in pediatric high-grade glioma

Frank Dubois (✉ fdubois@broadinstitute.org)

Dana-Farber Cancer Institute <https://orcid.org/0000-0002-7654-6208>

Frank Dubois

Dana-Farber Cancer Institute

Ofer Shapira

The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02124, USA

Noah Greenwald

Stanford <https://orcid.org/0000-0002-7836-4379>

Travis Zack

Dana Farber Cancer Institute

Jeremiah Wala

Dana-Farber Cancer Institute

Jessica Tsai

Dana-Farber Cancer Institute

Djihad Hadjadj

McGill University

Alexander Crane

Dana-Farber Cancer Institute

Ashot Harutyunyan

McGill University <https://orcid.org/0000-0002-9050-9315>

Kiran Kumar

Dana-Farber Cancer Institute

Mirjam Blattner-Johnson

German Cancer Research Center

Jayne Vogelzang

Dana-Farber Cancer Institute

Cecila Sousa

Dana-Farber Cancer Institute

Kyung Shin Kang

Dana-Farber Cancer Institute

Claire Sinai

Dana-Farber Cancer Institute

Dayle Wang

Dana-Farber Cancer Institute

Prasidda Khadka

Dana-Farber Cancer Institute

Kathleen Lewis

Broad Institute of MIT and Harvard

Lan Nguyen

Broad Institute of MIT and Harvard

Hayley Malkin

Dana-Farber Cancer Institute

Patricia Ho

Columbia University <https://orcid.org/0000-0002-4622-8188>

Ryan O'Rourke

Dana-Farber Cancer Institute

Rose Gold

Dana-Farber Cancer Institute

Davy Deng

Dana-Farber Cancer Institute

Jonathan Serrano

NYU Langone Medical Center

Matija Snuderl

New York University Medical Center

Karen Wright

Dana-Farber Cancer Institute

Susan Chi

Boston Children's Hospital, Harvard Medical School

Jacques Grill

Institut Gustave Roussy

Claudia Kleiman

Jewish General Hospital <https://orcid.org/0000-0002-5158-7126>

Liliana Goumnerova

Boston Children's Hospital and Harvard Medical School <https://orcid.org/0000-0002-2662-719X>

Nada Jabado

McGill University and The Research Institute of the McGill University Health Centre

<https://orcid.org/0000-0003-2485-3692>

David Jones

Hopp Children's Cancer Center Heidelberg (KiTZ) <https://orcid.org/0000-0002-2036-5141>

Mark Kieran

Dana-Farber Cancer Institute

Keith Ligon

Dana-Farber Cancer Institute

Rameen Beroukhim

Department of Cancer Biology, Dana-Farber Cancer Institute and Harvard Medical School

<https://orcid.org/0000-0001-6303-3609>

Pratiti Bandopadhyay

Dana-Farber Cancer Institute

Biological Sciences - Article

Keywords: Pediatric high-grade gliomas (pHGGs), diffuse midline gliomas (DMGs), pediatric high-grade glioma

Posted Date: April 14th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-389596/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Cancer on July 4th, 2022. See the published version at <https://doi.org/10.1038/s43018-022-00403-z>.

Structural variants shape driver combinations and outcomes in pediatric high-grade glioma

2

Frank PB Dubois^{1,2}, Ofer Shapira^{1,2}, Noah F Greenwald^{1,2}, Travis Zack^{1,2}, Jeremiah Wala^{1,2}, Jessica W
4 Tsai^{2,3,4}, Jihad Hadjadj⁵, Alexander Crane^{1,2}, Ashot S Harutyunyan⁵, Kiran H Kumar^{1,2}, Mirjam Blattner-
Johnson⁶, Jayne Vogelzang⁷, Cecila Sousa⁷, Kyung Shin Kang^{1,2}, Claire Sinai⁷, Dayle K Wang^{2,4}, Prasidda
6 Khadka^{1,2}, Kathleen Lewis², Lan Nguyen², Hayley Malkin⁷, Patricia Ho^{1,2}, Ryan O'Rourke^{1,2}, Rose Gold^{1,2},
Davy Deng^{1,2}, Jonathan Serrano⁸, Matija Snuderl⁸, Karen D Wright^{3,4}, Susan N Chi^{3,4}, Jacques Grill⁹,
8 Claudia L Kleinman¹⁰, Liliana C Goumnerova^{11#}, Nada Jabado¹², David T. W Jones⁶, Mark W Kieran^{2,3,4,#},
Keith L. Ligon^{2,7,13*}, Rameen Beroukhir^{1,2,14*}, Pratiti Bandopadhyay^{2,3,4*}

10

¹Departments of Cancer Biology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, 02215, USA

²Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

³Department of Pediatrics, Harvard Medical School, Boston, MA, 02215, USA

⁴Department of Pediatric Oncology, Dana-Farber Boston Children's Cancer and Blood Disorders Center, Boston, MA, 02215, USA

⁵Department of Human Genetics, McGill University, Montreal, QC H3A 0C7, Canada

⁶Hopp Children's Cancer Center Heidelberg (KITZ); Pediatric Glioma Research Group, German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany.

⁷Department of Oncologic Pathology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA

⁸New York University Medical Center, New York, NY 10016, USA.

⁹ Department of Pediatric and Adolescent Oncology and INSERM Unit 981, Gustave Roussy Institute and University of Paris Saclay, Villejuif, France

¹⁰Department of Human Genetics, McGill University, Montreal, QC H3A 0C7, Canada; Lady Davis Research Institute, Jewish General Hospital, Montreal, QC H3T 1E2, Canada

¹¹Department of Neurosurgery, Boston Children's Hospital; Dana Farber/Boston Children's Cancer and Blood Disorders Center, Boston, MA, 02215 USA.

¹²Department of Human Genetics, McGill University, Montreal, QC H3A 0C7, Canada; Division of Experimental Medicine, Department of Medicine; Department of Pediatrics, McGill University, and The Research Institute of the McGill University Health Centre, Montreal, QC H4A 3J1, Canada.

¹³Department of Pathology, Brigham & Women's Hospital, Boston; Center for Patient Derived Models, Dana-Farber Cancer Institute; Department of Pathology, Boston Children's Hospital and Harvard Medical School, Boston, MA, 02215, USA

¹⁴Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, 02215, USA

Current affiliation: LCG: Tromboprotea, MWK: Bristol-Myers Squibb, Boston, Devens, MA, 01434, USA; and Lawrenceville, NJ, 08648, USA.

12 ***Corresponding authors** (these authors jointly supervised this work):

Keith Ligon: keith_ligon@dfci.harvard.edu

14 Rameen Beroukhir: rameen_beroukhir@dfci.harvard.edu

Pratiti Bandopadhyay: pratiti_bandopadhyay@dfci.harvard.edu

16

Abstract

18
20
22
24
26
28
30
32
34
36

38

Pediatric high-grade gliomas (pHGGs), encompassing hemispheric and diffuse midline gliomas (DMGs), remain a devastating disease. The last decade has revealed oncogenic drivers including single nucleotide variants (SNVs) in histones. However, the contribution of structural variants (SVs) to gliomagenesis has not been systematically explored due to limitations in early SV analysis approaches. Using SV algorithms, we recently created, we analyzed SVs in whole-genome sequences of 179 pHGGs including a novel cohort of treatment naïve samples—the largest WGS cohort assembled in adult or pediatric glioma. The most recurrent SVs targeted MYC isoforms and receptor tyrosine kinases, including a novel SV amplifying a MYC enhancer in the lncRNA *CCDC26* in 12% of DMGs and revealing a more central role for MYC in these cancers than previously known. Applying *de novo* SV signature discovery, we identified five signatures including three (SVsig1-3) involving primarily simple SVs, and two (SVsig4-5) involving complex, clustered SVs. These SV signatures associated with genetic variants that differed from what was observed for SV signatures in other cancers, suggesting different links to underlying biology. Tumors with simple SV signatures were *TP53* wild-type but were enriched with alterations in TP53 pathway members *PPM1D* and *MDM4*. Complex signatures were associated with direct aberrations in *TP53*, *CDKN2A*, and *RB1* early in tumor evolution, and with extrachromosomal amplicons that likely occurred later. All pHGGs exhibited at least one simple SV signature but complex SV signatures were primarily restricted to subsets of H3.3^{K27M} DMGs and hemispheric pHGGs. Importantly, DMGs with the complex SV signatures SVsig4-5 were associated with shorter overall survival independent of histone type and *TP53* status. These data inform the role and impact of SVs in gliomagenesis and mechanisms that shape them.

Introduction

40

Pediatric high-grade gliomas (pHGGs), encompassing diffuse midline gliomas (DMGs) and hemispheric tumors, represent the most common cause of cancer-related deaths in children age 0-14 years¹. Over the last decade, several landmark papers have revealed recurrent single nucleotide variants (SNVs) in the core histones H3.3 and H3.1 including H3.1^{K27M}, H3.3^{K27M}, and H3.3^{G34R} mutations, co-occurring with alterations in the TP53 signaling pathway and receptor tyrosine kinases (RTKs)²⁻⁹. Such information has been generally derived from exome or targeted sequencing panels, limiting the ability to characterize structural variants (SVs). As such the role of such events in pHGG is underexplored.

48

SVs represent connections, or rearrangements, between distant genomic loci and are critical drivers in other cancers. They underlie all somatic copy-number alterations (SCNAs) except whole chromosome gains and losses, thereby altering more of the genome than any other genetic alteration. A single SV can have long-range effects on dozens to hundreds of genes. In some cases, SVs result in extrachromosomal amplifications (otherwise known as double minutes)¹⁰ that can lead to hundreds of copies of an oncogene per cell¹¹⁻¹⁵. Moreover, the effects of SVs on cellular fitness often result from changes in chromatin structure such as disruption of topologically associated domains (TADs) and gene-enhancer interactions¹⁶⁻²¹. Therefore, unlike SNVs, the SVs with the largest effects on selection are often outside of the exome and require whole-genome sequencing (WGS) for their characterization. Moreover, these distant effects raise challenges to interpreting the effects of individual SVs.

58

Both the frequency with which SVs recur at individual loci, and the mechanisms by which they are generated, can vary widely across cancers. Recently, we and others have developed analytic tools to detect both the recurrent, likely oncogenic SVs, and patterns of SVs (SV signatures). These signatures can indicate the mechanisms by which SVs are generated²²⁻²⁶, and recent efforts have begun to characterize them in breast²⁶, ovarian^{27,28} and other cancer types²³. However, unlike SNVs, whose signatures have been characterized across tens of thousands of exomes, no consensus catalog of SV signatures exists, and the relationships between currently described SV signatures and other cancer types has not been well-explored. For example, high rates of tandem duplications have been associated with deficiencies in homologous recombination (HR) only in tumors with very high SV burdens^{23,29-31}. It remains unclear if these associations translate to other tumor types, including pHGGs, and which other SV signatures and associated variant-generating processes exist.

70

The differences across lineages indicate a substantial role for epigenetics in shaping the SVs that are observed in cancer^{22,23}. Mutations in core histones were first reported in pHGGs^{8,9}, highlighting the role of epigenetic dysregulation in these tumors. pHGGs therefore offer a unique perspective on the relationships between patterns of SVs and different alterations in chromatin. Associations between patterns of SVs and other molecular and clinical characteristics of these tumors are also largely unknown.

76

One obstacle to the genomic characterization of DMGs has been a lack of pre-treatment tissue. Due to their location, most often in the pons, biopsies were traditionally avoided. A concern in characterizing post-treatment samples is that treatment—most often involving ionizing radiation—might itself alter the SV patterns detected in these tumors. Recently, however, pre-treatment biopsies have been performed more widely due to landmark clinical trials showing their safety³²⁻³⁵.

82

We leveraged pre-treatment biopsies from these studies, as well as novel and published data from both pre-treatment biopsy and autopsy samples of DMGs and hemispheric tumors, to assemble the largest WGS cohort to

84 date from pHGGs. We then applied recently developed tools to sensitively detect SVs, identify recurrent and likely
driver events, and stratify pHGGs based upon mechanistically informative SV signatures. We further detected
86 genetic events and differences in clinical outcome associated with patterns of SVs across the genome.

88

Results

90

Significantly recurrent SVs

92

In order to characterize SVs in pHGG including samples whose genomes had not been altered by treatment, we
94 assembled a cohort of pHGG WGS data from 179 children including 61 hemispheric tumors and 118 diffuse midline
gliomas, or DMGs. Of these, 61 are novel to this study (Supp. Table 1). Among the DMGs, 84 (71%) were from pre-
96 treatment biopsies including 33 obtained from the first multi-institutional North American clinical trial to
incorporate diagnostic biopsies³⁵. The tumor purity of the pre-treatment biopsies was comparable to autopsy
98 samples (median: 0.8 vs. 0.78, $p = 0.5$) (**Ext. Data Fig. 1A**). This represents the largest cohort of pretreatment
DMGs with whole genome sequencing to date.

100

In analyses of somatic copy-number alterations (SCNAs), we were able to detect a new recurrent amplification in
102 8q24.21, 2 MB telomeric to *MYC*. In total, 28 tumors (16%) exhibited absolute copy-numbers (CNs) > 3.7 . In more
than half of these (15/28), the amplicon excluded *MYC* itself, such that they defined a peak region of amplification
104 that did not encompass *MYC* (**Ext. Data Fig. 1C**). All but one of these were DMGs, representing a significant
enrichment ($p = 0.0016$). A separate peak was also detected that did encompass *MYC* without the telomeric
106 region, due to two tumors with extrachromosomal amplicons that linked *MYC* to other loci. The new amplification
peak may not have been detected in prior array- and exome-based studies because it encompasses only 200 kb
108 and lies entirely outside of the exome³. Aside from this amplicon, the results of SCNA and SNV recurrence analyses
largely agreed with those detected in prior studies³⁻⁷. These include high frequencies of H3.3^{K27M} (50%) or H3.1^{K27M}
110 (12%) mutations, reflecting the enrichment for midline glioma in this study, and other known drivers of pHGG
(**Ext. Data Fig. 1B**).

112

We also made the novel finding that several previously-described regions were recurrently amplified together to
114 high levels, such as 2p25.1 and the MYCN locus at 2p24.3. This pattern of apparently linked SCNAs in distinct
genomic loci suggests underlying recurrent SVs. We therefore comprehensively catalogued SVs using an assembly-
116 based method³⁶ that is capable of detecting complex and short SVs that are not easily detected using standard
alignment-based methods. In total, we detected 15,485 SVs, averaging 87 per tumor, including 1482 (10%) that
118 were 10-300 bp in span, a “blind spot” in prior analyses³⁶.

To distinguish recurrent SVs, we took two approaches based upon methods that we recently developed³⁷. In the
120 first, we conducted a “1D” analysis to detect loci with significantly recurrent SV breakpoints (SRBs). In the second,
122 we conducted a “2D” analysis to detect pairs of loci that are recurrently juxtaposed by SVs (“significantly recurrent
juxtapositions”, or SRJs). SRBs might not appear within SRJs if they connect to a dispersed set of loci. Conversely,
124 breakpoints within SRJs may not be SRBs; the finding of recurrent juxtapositions between two loci increases our
sensitivity for SRJs whose individual breakpoints might be insufficiently frequent to be noted as SRBs.

126

We identified SRBs in five TADs (Supp. Table 3 or **Fig. 1A**, and **Ext. Data Fig. 1D**) and two SRJs (**Fig. 2A**; **Ext. Data Fig. 3A**). The most significant SRB was within the MYC TAD, in the long non-coding RNA (lncRNA) CCDC26. The same locus was also identified as a component of an SRJ connecting two loci at the telomeric end of the MYC TAD. The remaining SRBs corresponded to SVs within the TADs of the receptor tyrosine kinases (RTKs) MET ($q=0.0025$), EGFR ($q=0.029$), and PDGFRA ($q=0.032$), as well as an SV within the TAD of the transcription factor ID2. This latter SRB was also a component of the second SRJ, which connected the loci of ID2 and MYCN. We describe each of these in more detail below.

Rearrangements within the lncRNA CCDC26 within the MYC TAD activate MYC by amplifying an interacting enhancer

Our 1D and 2D analyses both nominated rearrangements within the MYC TAD as among the most significantly recurrent SVs in pHGG. A total of 28 SVs were identified within this TAD. Among these, six were complex rearrangements that amplified the MYC protein coding sequence, connecting the amplicon to locations outside of the MYC TAD. An additional seven rearrangements were dispersed deletions or low-quality SV calls, which we classified as potential passenger events or artifacts. The remaining 15 represented a tandem duplication centering on intron 1 of the lncRNA CCDC26, with a median span of 216 kbp and a minimal common region of amplification (MCR) of only 42 kb (**Fig. 1B**).

The 2 Mb region telomeric to MYC has been shown to contain MYC enhancers in hematopoietic lineages^{38,39}, neuroblastoma⁴⁰, and lung and endometrial carcinomas⁴¹, but with lineage specific enhancer locations^{38,41}. We therefore hypothesized that the CCDC26 amplicon promotes oncogenesis by amplifying an associated enhancer. We analyzed previously published H3K27ac enhancer tracks generated from H3^{K27M} mutant and wild-type pHGGs⁴² and adult glioblastomas⁴³ and observed H3K27ac enhancer peaks within the MCR of the CCDC26 amplicon (**Figure 1C**, **Ext. Data Fig. 2A**). We also validated the presence of the CCDC26 associated enhancer in an independent dataset that leveraged ATAC-seq to profile the chromatin landscape of an independent cohort of pHGGs⁴⁴. This enhancer also appears to be present in normal neural tissue available through ENCODE^{45,46}. In contrast, enhancer maps from cells derived from hematopoietic and lung tissues did not show peaks at this location (**Ext. Data Fig. 2A**). We conclude that the CCDC26 amplicon centers on a neural lineage-specific enhancer peak.

Although this enhancer is present across neural tissues, histone mutant DMGs exhibit 31% more H3K27ac binding at this locus than H3^{WT} gliomas, constituting the fourth most differential super enhancer between these groups (q -value = 0.05; $p=0.001$)⁴². Intriguingly, the CCDC26 amplicon within our dataset also occurred almost exclusively in H3^{K27M} gliomas (present in 14/97 H3^{K27M} vs. 1/82 H3^{G34R} mutant or H3^{WT} tumors; $q = 0.0018$). These data indicate that the SRB in CCDC26 amplifies an enhancer which is present in the whole neural lineage but shows a quantitative enrichment specifically in H3^{K27M} gliomas.

This lineage-specific enhancer also appears to interact directly with *MYC*. We interrogated Hi-C data from an H3.3^{K27M} mutant pHGG, two patient-derived H3.3^{K27M} mutant cell lines, and iPSC-derived neural progenitors, and found a prominent interaction peak between the MCR of the *CCDC26* amplicon and the *MYC* promoter in all of them (**Fig. 1D**; **Ext. Data Fig. 2B**). We also observed increased *MYC* expression in pHGGs with this amplicon relative to samples without any SCNAs in the MYC TAD (**Fig. 1E**; $p=0.04$), supporting the hypothesis that the amplicon targets a *MYC* enhancer. Moreover, these CCDC26-SV samples had similar *MYC* expression as samples with amplicons encompassing the *MYC* coding region ($p=0.85$). To functionally validate enhancer activity of the

172 H3K27ac peak in the *CCDC26* amplicon, we generated two enhancer reporter systems (E1 and E2), each
encompassing slightly more than one half of the overall enhancer sequence, with a small region of overlap. The
174 E1 subset showed significant enhancer activity ($p=0.019$, $n=3$, T-test) compared to the reporter construct without
insert. (Fig. 1F-G).

176

These results indicate that the most significantly recurrent SV in histone mutant pHGGs amplifies a neural-lineage
178 *MYC* enhancer that is augmented in the H3^{K27M} context, thereby increasing *MYC* expression.

180

Activation of MYCN through enhancer amplification and hijacking

182

Somatic enhancer amplification also seemed to play a role in the activation of *MYCN* in pHGG. The SRJ connecting
184 *ID2* with *MYCN* represents a set of highly complex SVs in tumors with high-level amplifications within this region
on chromosome 2 (Fig. 2B). *ID2* is a transcription factor that is well-recognized to play essential roles in regulating
186 neural differentiation^{47,48}. Accordingly, we observe the *ID2* locus to be associated with an H3K27ac enhancer that
is present across all analyzed pHGG tumor samples (Ext. Data Fig. 3B). The MCR of the *MYCN-ID2* amplification
188 contains both the *ID2*-associated enhancer peak and the coding sequence of *MYCN* (Ext. Data Fig. 3B). The SVs
result in juxtaposition of the enhancer in *ID2* with *MYCN*, reducing the distance between the two from the normal
190 7 Mbp to less than 700 kbp (Fig. 2C). These data suggest these SVs hijack the *ID2* enhancer to activate *MYCN*.

192 We also identified four pHGGs with *MYCN* amplifications that did not connect to *ID2*. However, these latter
“localized *MYCN*” amplicons always encompassed more of the immediate neighborhood to *MYCN* than the
194 complex *MYCN-ID2* amplicons above. In contrast to *MYCN-ID2* amplicons, which only contained a small fraction
of the *MYCN* TAD (23% on average), localized *MYCN* amplicons contained most of this TAD (60% on average,
196 $p=0.03$, T-test), including several enhancers that are not in the *MYCN-ID2* amplicons (Fig. 2F).

198 The high-level *MYCN* amplicons show typical characteristics of extrachromosomal amplicons (EC-amps), reaching
copy numbers of 50 – 300 per cell. Other oncogenes with absolute copy numbers of more than 10 have been
200 shown to be amplified on double minutes in various cancer types^{14,49}. While projections on the linear reference
genome resulted in complex patterns that are typical of EC-amps^{15,50}, it was possible to construct simple circular
202 amplicons containing *MYCN* and, in the *MYCN-ID2* cases, *ID2* (Fig. 2D). Indeed, these circular amplicons represent
an optimal solution to explain the joint copy number and SV profiles in this region. The *ID2-MYCN* amplicons
204 appeared to reflect multiple subclonal amplicons including some containing additional oncogenes such as *MDM4*
(Fig. 2E). In contrast, pHGGs with localized *MYCN* amplicons showed significantly fewer SVs between the *MYCN*
206 TAD and other TADs (Fig. 2G and Ext. Data Fig. 3C).

208 We conclude that *MYCN-ID2* rearrangements are likely to be an example of enhancer hijacking, bringing a strong
enhancer in *ID2* next to *MYCN* on amplicons without the endogenous elements of the *MYCN* TAD, whereas
210 amplifications of *MYCN* without *ID2* also amplify enhancers within the *MYCN* TAD.

212

Significantly recurrent breakpoints around RTKs also suggest extrachromosomal amplicons

214

The remaining SRBs all involved receptor tyrosine kinases (RTKs) that are known to be amplified and oncogenic in
pHGG: *PDGFRA*, *EGFR*, and *MET*^{3-7,51}. These loci, along with *MYC* and *MYCN*, were also the only regions with high-

216

level amplifications recurrent in at least three patients (**Fig. 3A**). Similar to our findings from the MYC isoforms, the RTK SRBs also showed both simple-SVs that presumably amplify local enhancers (**Fig. 3B, Ext. Data Fig. 4C**) and SVs that appeared to reflect complex extrachromosomal amplicons that integrate distant sites and reach as many as 200 copies (**Fig. 3C and E**)

Overall, 35/179 tumors showed at least one amplicon (>50 kbp in size) with an absolute copy number greater than ten. Among 34 of these 35 tumors, the high-level amplicons contained at least one well-known oncogene, and apart from the coding sequence of the oncogene, they recurrently incorporated the same genomic loci around the oncogene (exemplified in **Ext. Data Fig. 4A-B** for *PDGFR* and *EGFR*). The high-level amplicons in several pHGGs also contained multiple oncogenes from distant locations (**Fig. 3D-E**). In some cases, we detected SVs connecting these oncogenes with supporting reads that allowed for the reconstruction of circular extrachromosomal amplicons containing oncogenes from different chromosomes (**Fig. 3E**). This might suggest either synchronous chromothripsis of two chromosomes or translocation of one oncogene next to another with subsequent circularization. These data suggest that only a subset of pHGGs develops high-level amplicons, which recurrently contain the same (presumably regulatory) sequences in addition to the target oncogene.

To further understand the structure of these amplicons, we first focused on high-level amplicons containing *PDGFRA*. These amplicons span more than 2.5 Mbp and are super-imposed upon low-level amplicons of the surrounding region, often starting from the centromere. The amplicons included *KIT* in 80% of cases and reached *KDR* in 60%. All but one (14/15, 93%) of amplicons in the *PDGFRA* TAD amplified the *PDGFRA* coding sequence itself, often to the highest copy numbers reached in the region (**Ext. Data Fig. 4A**). The sole exception harbored amplification of a short sequence centromeric to *PDGFRA* that contains H3K27ac enhancers (**Fig. 3B**) that have been shown to interact with the *PDGFRA* promoter in pHGGs¹⁹ and adult GBMs,⁵² suggesting use of enhancer amplification to activate *PDGFRA*. Indeed, this region was included in the *PDGFRA* amplicon in nearly all tumors (14/15, 93%) (**Ext. Data Fig. 4A**). The exception appeared to contain an in-frame fusion of *NMU* and *PDGFRA*, suggesting an alternative mode of *PDGFRA* activation. These data suggest that SVs in pHGG recurrently incorporate an upstream enhancer-rich region into high-level *PDGFRA* amplicons.

The high-level *EGFR* and *MET* amplicons also extended beyond the RTK coding sequence to recurrently involve associated enhancer elements (**Ext. Data Fig. 4B-D**). The *EGFR* amplicons showed a clear skew towards the enhancer region in *SEC61G*, which drives *EGFR* expression in extrachromosomal amplicons in adult GBM²¹. Both *EGFR* and *MET* amplicons showed subclonal SVs within the coding sequence, potentially allowing for expression of alternate transcripts. In two tumors, these resulted in the *EGFRVIII* variant (one is exemplified in **Fig. 3E**). Similar to the *MYCN-ID2* SVs, *EGFR* amplicons tended to have a highly complex subclonal structure with incorporation of distant oncogenes (**Fig. 3E**). *MET* amplicons also skewed towards a region including enhancer elements in *CAPZA2* (**Ext. Data Fig. 4D**). However, *MET* amplicons exhibited less subclonal diversity, with simpler circular amplicons containing only *MET* and its associated enhancers. We conclude that the RTK amplicons are shaped by the epigenetic machinery necessary to drive expression of these oncogenes.

SV signatures relate to genetic and epigenetic tumor states

To identify the mechanisms underlying the SVs in these tumors, we applied a Bayesian NMF-based analysis²⁴ to detect SV signatures, based upon their spans, complexity (abundance of neighboring SVs), and rearrangement type (deletion, duplication or inversion). We identified only five such signatures (**Fig. 4A**), broadly corresponding

262 to simple short (<30kbp) duplications (SVsig1: “Simple Short Dups”); mixed simple-SVs (all sizes and types except
small duplications, SVsig2: “mixed simple”); simple interchromosomal rearrangements (SVsig3: “Simple
264 Interchromosomal”); complex interchromosomal events (SVsig4: “Complex Interchrom.”) and complex large SVs
(>=1 Mbp, all types, SVsig5: “Complex Large”).

266 We next looked for possible causes and consequences of the pHGG SVsig signatures by testing for associations
268 between the activity of each SVsig and presence of recurrent and known oncogenic⁵³ genetic events (Supp. Table
2). The Complex-SV signatures (SVsigs4-5) were most closely associated with *TP53* disruption and oncogenic
270 amplifications encompassing *MYC*, *MYCN*, *PDGFRA*, and *EGFR*—which often themselves represented complex-SVs
– and losses of *CDKN2A/B* and *PTEN*. These signatures were also anticorrelated with oncogenic mutations in
272 *PPM1D* and *ACVR1* (Supp. Table 4 and **Fig. 4B**). Simple Interchrom. was associated with *TERT* promoter mutations.
Notably, we observed complex-SV signatures in pHGGs with high SV counts ($\rho = 0.68$, $p < 2.2e-16$, Spearman's)
274 and disrupted DNA damage response (DDR), and found simple tandem duplications to be enriched in pHGGs with
lower SV counts. This contrasts with prior results from several adult cancers, where tandem duplications were
276 shown to be the dominant signature in highly rearranged cancers with disrupted DDR pathways^{23,26,30,31}, including
homologous recombination (HR)/BRCA^{23,54}. Indeed, none of the genes previously implicated in tandem duplication
278 signatures or loss of HR in adults^{23,54} reached a significant level of association with any signature in this cohort.

280 The complexity of SVsigs appeared to relate to the epigenetic status of the tumors in which they occurred, and
specifically which histone underwent modification. H3.1^{K27M} tumors almost exclusively exhibited the simple SVsigs
282 (SVsigs1-3), harboring significantly fewer gliomas with SVsigs4-5 than H3.3^{K27M} tumors ($q = 10^{-4}$). H3.3^{G34R} pHGGs
tended to exhibit the complex SVsigs while histone wild-type pHGGs had varying SV complexity, but these did not
284 rise to statistically significant levels of enrichment.

286 Taken together, we conclude that SV signatures associate with distinct genetic variants in pHGG.

288

Associations between SV and SNV signatures

290

We next asked whether pHGGs separate into subsets with different DNA damage and damage response
292 characteristics, based on patterns of both SVs and SNVs. Evaluation of SNVs in isolation revealed 14 SNV
signatures, including signatures similar to known aging, APOBEC, HR-deficiency and hypermutation SNV
294 signatures²⁵ (**Supp. Figs. 1-3**). Two of these SNV signatures were correlated with SV signatures (**Fig. 5A**). An SNV
signature resembling COSMIC signature 13 (SBS13), thought to be caused by APOBEC mutagenesis especially in
296 single-stranded DNA around replication forks and sites of DNA damage and repair^{55,56}, was associated with
Complex-SV signatures ($q=2 \times 10^{-16}$ for Complex Large and 10^{-13} for Complex Interchrom. signatures). A second SNV
298 signature, resembling COSMIC signature SBS3, was most closely associated with the Mixed Simple-SV signature
($q=10^{-13}$) and also with Complex-SV signatures ($q=2 \times 10^{-7}$ for Complex Large and 10^{-8} for Complex Interchrom.).
300 SBS3 has been associated with HR deficiency²⁵.

302 The entire pHGG cohort separated into three groups reflecting different amplitudes of the five SV and 14 SNV
signatures (**Fig. 5B-C**). The largest cluster (“Complex-SV”) was dominated by Complex-SV signatures ($q < 4.1 \times 10^{-20}$,
304 **Fig. 4C**) and was enriched for *TP53* inactivation ($q < 0.1$, **Fig 5D**), high-level RTK and *MYC/MYCN* amplicons ($q < 0.2$)
and depleted for *ACVR1*, *PPM1D*, and *H3.1^{K27M}* mutations ($q < 0.1$). In contrast, the second largest cluster (“Simple-
306 SV”) was dominated by the Simple-SV signatures ($q=0.03$), lacked *TP53* disruption ($q < 0.1$), and was enriched for

308 *PPM1D* mutations ($q < 0.1$). This cluster also contained the hypermutant samples with their distinct SNV signatures, lacked most SCNAs, and seemed to be driven instead by a combination of SNVs including *ACVR1*, *PPM1D*, H3.1^{K27M}, and *PIK3CA* mutations. The smallest cluster (“Intermediate”) was enriched with the HR-deficiency SNV signature
310 SBS3 ($q = 6.9 \times 10^{-5}$) and the Mixed Simple-SV signature ($q = 4.1 \times 10^{-7}$) and harbored an intermediate number of SVs compared to the other two clusters (**Fig. 5E**). Despite being *TP53* mutant in many cases, tumors in this cluster
312 showed fewer copy-number variants overall with the exception of high-level *PDGFRA* amplifications. All three clusters included both hemispheric and midline gliomas, although the Intermediate cluster was modestly enriched
314 for the latter ($q = 0.13$, Supplementary Table 1). These data suggest that distinct variant generating processes shape pHGG genomes.

316

318 *Genomic contexts of recurrent SVs and SV signatures*

320 Different histone mutations are known to be associated with distinct recurrent SNVs and SCNAs; we confirmed these known relationships and expanded them to include two novel associations with SVs^{3,57} (**Fig. 6A and B**). The
322 significantly recurrent SV in *CCDC26* resulting in *MYC* enhancer amplification was enriched in H3.3^{K27M} gliomas ($q = 0.008$) and H3.1^{K27M} pHGGs were enriched for a focal deletion of *CDKN2C* with breakpoints in the adjacent gene
324 *FAF1* ($q = 0.04$).

326 We also found that H3.3^{K27M} pHGGs divided into two groups based on SV complexity. Most exhibited high Complex-SV signature activity, but 42% did not. These two H3.3^{K27M} groups also exhibited distinct combinations
328 of genetic alterations in known cancer-related genes. We evaluated the statistical significance of these differences by calculating Jaccard distances between genetic profiles across 369 variants in genes from the Cancer Gene
330 Census⁵³ for every pair of tumors. H3.3^{K27M} Simple-SV pHGGs exhibited short distances from each other, as did H3.3^{K27M} Complex-SV pHGGs; however, the distances between the two groups was much greater ($q < 6 \times 10^{-8}$; **Fig.**
332 **5C**). Indeed, these H3.3^{K27M} Simple-SV pHGGs were as different from the H3.3^{K27M} Complex-SV pHGGs as the H3.1^{K27M} Simple-SV pHGGs were (Ext. Data **Fig. 5A**). The H3.3^{K27M} Simple-SV pHGGs showed a variant pattern
334 resembling H3.1^{K27M} DMGs (Ext. Data **Fig. 5B**). Mutations in *PPM1D*, *ACVR1*, and *PIK3CA*, gains of 1q encompassing *MDM4*, and amplifications in *CCDC26*, encompassing *MYC* enhancers, were enriched in these tumors relative to
336 H3.3^{K27M} Complex-SV pHGGs (all $q < 0.09$ except for *CCDC26* amplifications, where $q = 0.11$). In contrast, loss of *TP53* (by SNV or SCNA) and amplifications of *PDGFRA* and *MYC* were depleted (all $q < 0.05$). These data suggest
338 that the propensity of pHGGs to develop Complex-SVs influences the combination of driver alterations they accrue, even within groups defined by their histone mutations.

340

Across the DMGs with more than 20% Complex-SV signature activity (“H3^{K27M} Complex-SV”, including H3.3^{K27M}
342 [n=43] and H3.1^{K27M} [n=3] DMGs), the TP53 pathway was inactivated almost universally through direct disruption of *TP53* (44/46 cases, 96%; **Fig 6A**). In contrast, the majority of DMGs with less complex signature activity (“H3^{K27M}
344 simple-SV”, H3.3^{K27M} [n=30] and H3.1^{K27M} [n=21]) lacked such direct *TP53* disruption (37/51, 73% TP53^{WT}; $q = 2.3 \times 10^{-5}$), but appeared to suppress the TP53 pathway through other mechanisms. Mutations in *PPM1D* were
346 more prevalent in this group, though still in a minority (7/30 H3.3^{K27M}, 2/21 H3.1^{K27M}, 20% in total; vs 1/46 H3^{K27M}-complex tumors; $q = 0.008$). It is possible that gains of 1q, encompassing *MDM4*, also served to suppress the TP53
348 pathway in these tumors. Although 1q spans approximately 2580 genes, we observed two sources of evidence that their prevalence in Simple-SV DMGs related to *MDM4* and TP53 pathway suppression. First, *MDM4* was
350 significantly overexpressed in 1q-amplified pHGGs of all types in our cohort ($q = 0.004$; Ext. Data **Fig. 5C**). Second, these 1q gains were the only arm-level SCNAs that anti-correlated with disruption of *TP53* ($q = 2.8 \times 10^{-6}$ in H3^{K27M}-

352 DMGs; $q=0.0003$ across all pHGGs, **Ext. Data Fig. 5C-D**) apart from gains of chromosome 2 ($q < 0.0025$). In contrast,
354 seven of the other thirteen significantly recurrent arm-level SCNAs were positively correlated with TP53 disruption
355 in H3^{K27M}-DMGs (all $q < 0.018$), presumably due to the role of *TP53* in generating aneuploidies^{58,59}. Indeed, across
356 TP53^{WT} H3^{K27M} DMGs, gains of 1q were among the most common genetic events, observed in 85% of tumors (16/20
357 H3.3^{K27M}, 17/19 H3.1^{K27M}), as opposed to 31% of *TP53*-disrupted H3^{K27M} DMGs ($p= 1.4 \times 10^{-6}$). No pHGGs in our
358 cohort exhibited focal, high-level amplifications of *MDM2*. These data suggest that direct disruption of *TP53*
359 contributes to a different pattern of SVs compared to other mechanisms of TP53 pathway inactivation, which
360 occur primarily through alterations of *PPM1D* and *MDM4*.

361
362 Focusing on the tumors that harbored significantly recurrent SVs (SRSVs), we observed two groups. One group
363 contained the tumors with high-level amplicons of the coding sequences of the oncogenes *PDGFRA*, *EGFR*, *MET*,
364 *MYC*, and *MYCN* (“Onc-amp”). In contrast, the second group amplified presumed enhancer elements within the
365 TADs of these oncogenes without amplifying their coding sequences (“TAD-amp”). The Onc-amp pHGGs exhibited
366 significantly higher activity of the Complex-SV signatures ($p = 4 \times 10^{-7}$; **Fig. 7A-B**). Intriguingly, the two groups also
367 harbored inactivating alterations in different DNA damage response genes (**Fig. 7A**). Onc-amp pHGGs were
368 enriched for *TP53* SNVs (69% of Onc-amp vs. 18% of TAD-amp pHGGs, $q=0.01$) and *RB1* deletions (23% of Onc-
369 amp vs. 0% of TAD-amp pHGGs, $q=0.16$). In contrast, TAD-amp pHGGs were enriched with *PPM1D* SNVs (29% of
370 TAD-amp vs. 0% of Onc-amp pHGGs, $q=0.03$) and gains of 1q encompassing *MDM4* (71% of TAD-amp vs. 34% of
371 Onc-amp pHGGs, $q=0.16$). In sum, alterations in *TP53* and *RB1* associate with Complex-SV signatures and high-
372 level amplifications of oncogene coding sequences, while *PPM1D* SNVs and 1q gains more frequently occur with
373 Simple-SV signatures and amplifications of enhancer elements near oncogenes.

374 *Temporal evolution of genetic variants*

375
376 These data raise the possibility that alterations in DNA-damage responses (DDR) not only shape the processes that
377 generate SVs but also the types of driver alterations they exhibit in *MYC*, *MYCN* and RTKs. Specifically, we
378 considered two hypotheses in regards to tumor evolution. First, disruption of DDR could be an early event that
379 activates the Complex-SV generating process and culminates in the development of a specific class of genetic
380 events including the high-level RTK and *MYCN* amplicons described earlier. Alternatively, both disruption of the
381 DDR and the high-level oncogene amplification could happen later in tumor development as a consequence of the
382 complex SVs involving these cancer genes.

383
384 Notably, we observed no effects of therapy on SV patterns, suggesting that the SVs occurred during gliomagenesis.
385 Although radiation treatment has been shown to induce DNA breaks⁶⁰, we found no differences in the number of
386 SVs per sample (median 35 vs 42; $q= 0.6$) or in the activity of the Complex-SV signatures (median 24% vs 28%;
387 $q=0.7$) between pre-treatment biopsy and autopsy samples (**Supp. Figs. 4A-B**).

388
389 We therefore performed a timing analysis reflecting the relative ordering of mutations and SCNAs during
390 gliomagenesis⁶¹. We found that the focal amplification of the *MYC* enhancer in *CCDC26* is one of the earliest
391 variants in the TAD-amp samples (**Fig. 7C**), occurring earlier than alterations in *PPM1D* and 1q/*MDM4* gain. In
392 contrast, the amplification of the oncogenes’ protein coding sequences in the Onc-amp samples happened after
393 the loss of the tumor suppressor genes *TP53*, *RB1* and *CDKN2A/B* (**Fig. 7D**). These data suggest that simple tandem
394 duplications can arise in tumors without major disruptions of DDR, potentially contributing to tumor initiation,

396 whereas the creation of the complex high-level amplicons of MYC isoforms and RTKs requires prior direct genetic
disruption of *TP53*, *RB1*, or *CDKN2A/B*.

398

Prior studies have shown histone mutations to be the initiating event in the pHGGs in which they occur^{8,9,42,51,62-64}. However, the studies investigating pHGG evolution in human tumor tissues were limited to exomic alterations in fewer than 15 patients⁶⁵⁻⁶⁸. Our cohort of whole genome sequencing data from 179 patients increased our power to detect differences in timing of genetic events within each histone subgroup. Among SNVs, we confirmed the findings of these previous studies⁶⁵⁻⁶⁸ including that H3^{K27M} mutations are the earliest mutations, followed by SNVs in *ACVR1* and *TP53* in H3.1^{K27M} and H3.3^{K27M} gliomas, respectively. (**Ext. Data Fig. 6**). This large cohort of WGS samples also provided an unprecedented opportunity to time focal SCNAs based on the ratio of SNVs acquired before and after each change in each copy number⁶¹. We found that losses of *TP53*, *CDKN2A/B*, and *RB1* precede RTK amplifications across each of the H3.3^{K27M}, H3.3^{G34R} and H3^{WT} subgroups of pHGG. In H3.3^{K27M} DMGs, the simple amplifications of the *MYC* enhancer in *CCDC26* were early events while the complex amplifications of the *MYC* coding sequence occurred later in tumor development.

410

412 *Complex-SV DMGs are associated with shorter survival*

414 We suspected that the fundamental differences in SV-generating processes across DMGs could associate with clinical phenotypes including survival. First, we confirmed the known association^{57,69} between the H3.1^{K27M} mutation and longer overall survival (OS) compared to H3.3^{K27M} (9.3 vs 16.1 months; p=0.0004; **Fig. 6D** top), and the lack of association between *TP53* mutation and OS within H3.3^{K27M} pHGGs (p=0.72; **Fig. 6D** bottom). To address SV signatures specifically, we also investigated the correlations between the numeric values of the combined Complex-SV signature and OS. Across all DMGs, this Complex-SV signature was significantly anti-correlated with OS (**Fig. 8A**; p=0.001).

422 We also found the combined Complex-SV signature to be significantly associated with shorter survival in a multivariate Cox regression analysis of DMGs that controlled for the known predictors of survival^{3,57} (Histone-SNV and age) and for *TP53* status (**Fig. 8B**; p=0.038). This analysis confirmed a significantly increased hazard ratio for H3.3^{K27M} compared to both H3.1^{K27M} and H3^{WT} DMGs, and a lack of significant associations between *TP53* disruptions and OS in multi-variate analyses as previously described⁵⁷. However, associations with age did not reach significance, probably due to our low representation of the under-three and over-ten age groups. While all patients with DMGs in our study died from their disease, the combined effects of these factors caused survival differences of several months. For example, children with DMGs with at least 20% Complex-SV activity survived a median of 9.6 months, about 3 months less than the 12.3-month survival of children with less than 20% Complex-SV activity (**Fig. 8C**).

432

Discussion

434

Over the last decade, several landmark studies⁴⁻⁷ have revealed tremendous insights into the recurrent driver combinations of SNVs and SCNAs in pHGG³, including the sentinel discovery of the role of histone mutations as a founding alteration that shapes the genetic, epigenetic, and clinical features of these tumors^{8,70}. However, the genome-wide landscape of SVs has not been described across a large cohort of pHGGs, particularly treatment-naïve DMGs. We addressed this by leveraging novel analytical tools developed by our group to sensitively detect

440 both individual and recurrent SVs^{22,36} and a large cohort of treatment-naïve DMGs made available by the DIPG-
BATS trial³⁵, the first US multi-institutional trial to incorporate pontine biopsies at diagnosis. This analysis revealed
442 novel structural variant signatures that distinguish pHGGs with predominantly simple vs complex SVs, and that
these signatures, together with histone mutation status, determine the recurrent driver combinations and clinical
444 outcomes. Among these are novel recurrent SVs, including a tandem duplication in 12% of all DMGs that
colocalizes with a *MYC* enhancer within the lncRNA *CCDC26*.

446
Our analyses of structural variants in pHGG highlight a previously underrecognized role for *MYC* pathway
448 activation in this tumor type. *MYC* is the most frequently amplified gene in all of cancer, with focal amplifications
observed in 15% of tumors^{58,71}. In contrast, similar *MYC* amplifications are only present in 5% of all pHGGs³. The
450 discovery of recurrent amplification of the *MYC* enhancer in *CCDC26* in pHGGs, without amplification of the *MYC*
coding sequence itself, starts to address this discrepancy. The difference in how *MYC* is activated in pHGGs also
452 extends to the location of *CCDC26* duplication. While tissue-specific amplifications of *MYC* enhancers have been
previously reported in other cancers⁴¹, the *CCDC26* duplication is in a novel location apparently driven by
454 differences in enhancer landscapes across cell types. Altogether, when adding in high-level *MYCN* amplifications
as well, we observe 14% of all pHGGs to harbor SVs predicted to activate *MYC* pathways. Given this high rate, the
456 role of *MYC* activation in regulating cell-fate decisions and enhancing glioma formation in both histone-mutant
and wild-type gliomas requires further study.

458
Although both Simple- and Complex-SV pHGGs activate *MYC* signaling pathways, they do so in strikingly different
460 ways. While pHGGs with Simple-SV signatures amplify only the *MYC* enhancer, pHGGs with Complex-SV signatures
contain high-level amplicons of both the *MYC* coding sequence and the enhancer regions - often
462 extrachromosomal, as indicated by their circular topology. In this respect, *MYC* serves as an example for other
oncogenes, including *MYCN*, *PDGFRA*, *EGFR* and *MET*. Extrachromosomal amplicons that contain both oncogenes
464 and recurrent regulatory elements have also been observed in adult high-grade gliomas and other cancers^{13,21,72}.
However, their regulatory elements are different from pHGGs and appear to reflect the exquisite tissue specificity
466 of these regulatory loci²⁰.

468 Extrachromosomal amplicons (also known as double minutes, or DMs) have been shown to originate as
byproducts of chromothripsis¹¹. Our data suggest that, in pHGG, they often contain multiple oncogenes from
470 different chromosomes. Based on the proposed theory of their origin, an extrachromosomal amplicon with
oncogenes from different chromosomes would therefore either require simultaneous chromothripsis of two
472 chromosomes or need to develop sequentially by a less-clear mechanism. Our data also suggest that within a
pHGG, the extrachromosomal amplicons often have multiple variants, which could correlate to different
474 descendants of the initial DM as observed in recent mechanistic work¹¹ and a prior long-read sequencing study of
a single tumor from this study¹⁵. In cases where two oncogenes are integrated into a DM that is subsequently
476 amplified, the number of copies of each oncogene should be identical. However, we find that pHGGs often exhibit
different amplification levels of these oncogenes, suggesting sequential incorporation into the amplicon. The
478 exact mechanism for this remains elusive and could range from sequential chromothripsis events¹¹ to reversible
DM integration in proximity to oncogenes as observed in neuroblastoma¹², or possibly deletions within the DMs.
480 It is tempting to speculate that the rapid evolution and optimization of DMs^{14,50} could contribute to the rapid,
lethal growth of pHGGs and their poor response to any available therapies, similar to adult GBM⁷³. Developing
482 RTK inhibition is still a promising therapeutic goal in pHGG^{4,74,75}, but our study highlights that understanding how
these extrachromosomal amplicons evolve might inform how pHGGs become resistant.

484

486 As the largest study of WGS data within a glial lineage, we expanded on the SV signatures that have been detected
488 in other cohorts, allowing us to explore the contribution that glial lineages make to the repertoire of SV signatures.
490 SV-signature analysis approaches were recently published for breast²⁶, ovarian^{27,28} and other cancer types²³.
492 Intriguingly, we observe distinct SV patterns and associations with genetic variants compared to those other
494 lineages. While tandem duplications have previously been associated with homologous recombination (HR)
496 deficiency with alterations of *BRCA1/2*, *TP53*, *CDK12* and *CCNE1* across breast^{29,30}, ovarian and hepatocellular
cancers³¹, this did not extend to pHGGs. The tumors in our cohort that were enriched with tandem duplications –
the Simple-SV tumors – were associated with none of these, and in fact tended to lack mutations in *TP53* in favor
of alterations of *PPM1D* and *MDM4*. However, the original associations were found in adult cancer types with
markedly higher numbers of SVs²³ and different genetic drivers than pHGG. The contribution of lineage context
and pHGG specific variants including the histone mutations to the formation of tandem duplications requires
further study.

498 One association with histone mutations that we did observe was between H3.3^{K27M}, Complex-SV signatures, and
500 loss of *TP53*. While the association between *TP53* disruption and a higher SV burden has been described
502 previously^{76,77}, the reason for its association with H3.3^{K27M} mutations over H3.1^{K27M} is not understood. Similar to
504 pioneering work in medulloblastoma showing that *TP53* disruption precedes and might facilitate survival after
506 chromothripsis⁷⁸, we also found *TP53* disruption to be an early event in tumors with complex SVs. Notably, while
almost all DMGs with Complex-SV signatures were *TP53* disrupted, not all *TP53*-disrupted H3.3^{K27M} DMGs showed
Complex-SV signatures. Therefore, additional features might determine whether a pHGG acquires Complex-SV
signatures. Additionally, hemispheric pHGGs with Complex-SV signatures were frequently *TP53*^{WT} but tended to
harbor early loss of *CDKN2A/B*. This indicates that while *TP53* loss and H3.3^{K27M} are tightly correlated with
Complex-SV signatures they are neither necessary nor sufficient, either alone or in combination, for the generation
of the Complex-SV signatures in pHGG.

510 We found a reportable alteration of a known cancer gene in 98.3% (176/179) of patients, significantly expanding
512 the share of patients with identified potential driver variants compared to prior studies that were based largely
514 on exome sequencing. Many of the alterations that we observed were in non-coding regions of the genome,
516 targeting regulatory elements such as enhancers. Moreover, use of whole-genome sequencing data allowed us to
determine which patients had Complex- or Simple-SV signatures, which itself is associated with survival, even
when controlling for histone and *TP53* status. These findings indicate that both research and clinical sequencing
of these tumors should encompass the whole genome.

518 Data Availability

520 We are in the process of transferring all novel sequencing data to dbGaP. Once this process is complete the data
522 will be available at phs002380.v1.p1

524 Code Availability

526 Publicly available software was used as indicated in the methods. Custom code to connect and reformat the
528 outputs of the publicly available software as well as code to generate the figures is available upon request.

Acknowledgements

530 We thank and acknowledge the children and their families who made this research possible through their
532 participation in pHGG clinical trials and tissue banking protocols. We also thank the many clinicians and
534 researchers who led the DIPG BATS trial. We would also like to thank the members of the Ligon, Beroukhim and
Bandopadhyay Laboratories, in addition to Ruolin Liu, Justin Rhoades, Matthew Leventhal, Justin Hwang,
Jonathan So, David Mayhew, Jonathan Rennhack, Mehdi Touat, Andrew Cherniack, Beatrice Rondelli and Sophie
Polo for their support and helpful discussions.

536 The authors would like to thank and acknowledge the following funding sources:
The German Research Foundation (DFG) (research fellowship supporting FD), Fund for Innovative Cancer
538 Informatics (RB), the Gray Matters Brain Cancer Foundation (RB), The Bridge Project of MIT and Dana-
Farber/Harvard Cancer Center (RB, KLL), The Sontag Foundation (RB), The V Foundation for Cancer Research (PB),
540 Michael Mosier Defeat DIPG Foundation (PB), The ChadTough Foundation (PB), The St. Baldrick's Foundation (RB,
PB, JWT), Prayers from Maria Foundation (PB, MWK), Pediatric Brain Tumor Foundation (PB, RB), Friends of DFCI
542 (PB, RB), The Giving for Gabi Fund (PB), McKennaClaire Foundation (PB, MWK), We Love You Connie Foundation
(PB, RB, MWK, NJ), Imagine for Margo and Lemos Family (JG), The Pussycat Foundation Helen Gurley Brown
544 Presidential Initiative (JWT, PB), Canadian Institutes of Health Research (NJ, CLK), Zach Carson Fund (MWK, RB,
PB), Ellie Kavalieros DIPG Fund (MWK, RB, PB), Ryan Harvey Fund (MWK, RB, PB), Mikey Czech DIPG Foundation
546 (MWK, RB, PB), Caroline Cronk Fund (MWK, RB, PB), Markoff Art in Giving Foundation (MWK, RB, PB), Brock
Fleming Fund (MWK, RB, PB), Stop and Shop Pediatric Brain Tumor Program (MWK, RB, PB, KDW, SNC), the DIPG
548 Collaborative (MWK, RB, PB), The Cure Starts Now (RB)

550 Author Contributions

FD, MWK, KLL, PB, RB conceived the project. FD, OS, NG, TZ, JW, JT, DH, AC, AH, KK, MBJ, KSK, CS, DW, PK, HM,
552 PH, RR, RG, DD, JS, MS, KW, SC, JG, CK, LG, NJ, DJ, MWK, KLL, RB, PB acquired samples, generated and analyzed
the data. FD, RB, PB wrote the manuscript, all authors edited the manuscript. KLL, PB and RB supervised the overall
554 study.

Conflict of Interest

556 RB and PB receive grant funding from the Novartis Institute of Biomedical Research for an unrelated project and
RB receives grant funding from Merck. P.B consults for QED Therapeutics, R.B. consults for and owns equity in
558 Scorpion Therapeutics, reports consulting or advisory role for Novartis, Merck (I), Gilead Sciences (I), ViiV
Healthcare (I); research funding from Novartis; patents, royalties, other intellectual property—Prognostic Marker
560 for Endometrial Carcinoma (US patent application 13/911456, filed June 6, 2013), SF3B1 Suppression as a Therapy
for Tumors Harboring SF3B1 Copy Loss (international application No. WO/2017/177191, PCT/US2017/026693,
562 filed July 4, 2017), Compositions and Methods for Screening Pediatric Gliomas and Methods of Treatment Thereof
(international application No. WO/2017/132574, PCT/US2017/015448, filed 1/27/2017). KLL reports grants and
564 personal fees from BMS, grants from Amgen, personal fees and other from Travera LLC, personal fees from
Rarecyte, grants from Tragara, grants from Lilly, grants from Deciphera, grants from X4, grants from Charles River
566 Labs, grants from Specicare, outside the submitted work; has a patent US20160032359A1 pending. MWK is now
an employee of Bristol-Myers Squibb. LCG is the Co-Founder of Thromboprotea Inc. All the remaining authors
568 declare no competing interests.

Methods

570

Sample acquisition

572 This study includes published⁴⁻⁷ data available under EGAS00001000575, EGAS00001001139, EGAS00001000572
574 and EGAS00001000192. Novel data was generated from samples obtained from the DIPG-BATs clinical trial, the
576 Dana-Farber Tissue Bank or collaborating institutions, under protocols approved by the institutional review board
578 of the Dana-Farber/Harvard Cancer Center with informed consent. DNA and RNA were extracted from single DMG
cores, pHGG biopsies, and autopsy samples using Qiagen AllPrep DNA/RNA extraction kits. Previously published
pHGG whole genome sequencing (WGS) data⁴⁻⁷ and, if present, paired RNA-seq were acquired from public
repositories.

Whole-genome sequencing

580 Library preparation for paired end whole genome sequencing (WGS) was performed as previously described¹⁷. In
582 brief, genomic DNA was fragmented and prepped for sequencing (to 60X depth for tumors and 30X depth for
584 normal samples) on an Illumina HiSeq 2000 instrument. Reads from both novel and published data were aligned
586 to the reference genome hg19/ GRCh37 with BWA⁷⁹, duplicate-marked, and indexed using SAMtools and Picard.
Base quality score was bias adjusted for flowcell, lane, dinucleotide context, and machine cycle and recalibrated,
and local realignment around insertions or deletions (indels) was achieved using the Genome Analysis Toolkit. All
paired samples underwent quality control testing to ensure accuracy of tumor-normal pairs.

588

SNV and SCNA analyses

590 SNVs were detected using Mutect2⁸⁰ and filtered for common sequencing artifacts, gnomad SNPs, and SNVs
592 present in a panel of whole genome-sequenced normal samples from blood from this cohort (n=175) and TCGA
(n=5). Significance of recurrent SNVs in non-hypermutable samples (SNV counts < 100,000/sample) was
594 determined with MutSigCV⁸¹. SCNAs were called using the GATK4 somatic CNV pipeline with normalization against
596 a panel of blood normals from 184 samples (174 from this cohort, 10 from TCGA) based on the Tangent method⁸².
Purity and ploidy were determined using ABSOLUTE⁸³. All somatic copy number alteration (SCNA) calls were
598 purity- and ploidy-adjusted as previously described⁸⁴. Significantly recurrent SCNAs were identified using GISTIC
2.0⁸⁵ with the following parameters: amp_thresh=0.5; del_thresh=0.7; arm_peel=0; broad_length_cutoff=0.5;
<https://firecloud.terra.bio> platform.

600

Structural variant detection and significance analysis

602 SvABA³⁶ was used to call SV in paired tumor normal mode with default parameters. In addition to filtering germline
604 variants against the paired normal, telomers and centromeres were blacklisted. The significance of recurrence
analysis was performed separately for breakpoints (1D) and juxtapositions (2D) analogous to the previously
published approach for pan-cancer²² and adapted to the pHGG context.

606 For the significance analysis of recurrent breakpoints (1D) as previously described²², the genome was binned into
50kbp bins with 500bp overlap. The bins were annotated by naming them after the TAD they were located in, with
608 the name of the TAD derived from the Cosmic Cancer Gene Census⁵³ gene in the TAD. Germinal zone TAD
boundaries from GSE77565⁸⁶ were used as the closest available normal neural progenitor. Eligible territory⁸⁷ was
610 defined by masking low-complexity genomic loci based on <https://github.com/lh3/sgdp-fermi/releases/download/v1/um35-hs37d5.bed.gz>. Only one SV per sample per bin was counted. Fishhook⁸⁸ was
612 used to calculate a background model for the likelihood of a breakpoint in each bin, based on six covariates:
replication timing (from http://mskilab.com/fishHook/hg19/RT_NHEK_Keratinocytes_Int92817591_hg19.rds), GC

614 content (from <http://hgdownload.cse.ucsc.edu/goldenPath/hg19/gc5Base/hg19.gc5Base.txt.gz>), presence of
SINE elements (<http://www.repeatmasker.org>), fragility (from
616 https://data.broadinstitute.org/pcawg6sv/1D_covariates/fragile_genes_smith.hg19fp.txt (adapted from ⁸⁹,
mappability (from <http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeMapability/>), and
618 heterochromatin status. ChromHMM for the caudate nucleus was chosen for the heterochromatin track based
epigenetic similarity to pHGG (Ext. Data Fig. 2A). These data were downloaded from
620 [https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/coreMarks/jointM
odel/final/E068_15_coreMarks_mnemonics.bed.gz](https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/coreMarks/jointModel/final/E068_15_coreMarks_mnemonics.bed.gz) and subset to ['8_ZNF/Rpts', '9_Het', '15_Quies',
622 '14_ReprPCWk'] to define heterochromatin. All these covariates have been repeatedly shown to be predictive of
the probability of a SV occurring in the bin^{22,23,88,90} and p-values for the influence of each individual covariate on
624 the model ranged from 2×10^{-16} to 0.21. p-values reflecting the significance of breakpoint enrichment were then
calculated and corrected for multiple hypothesis testing using the Benjamini–Hochberg procedure⁹¹. If several
626 bins within one TAD were significant, they were considered as one significant locus and linked to any COSMIC
Cancer Gene⁵³ gene in that TAD.

628
Recurrent juxtapositions (the 2D analysis; Ext. Data Fig. 3A) were detected using a background model determined
630 from 2658 cancers across several dozen cancer types²² and the same binning scheme (5583 bins with a median
span of 467kbp and interquartile range of 347kbp). Only one SV from each sample was allowed to contribute to
632 connections between any two bins (a “tile”). Only SVs with at least four supporting reads and a span of >1kbp
were included in this analysis. p-values reflecting the significance of enrichment of SVs within each tile were
634 corrected for multiple hypothesis testing using the Benjamini–Hochberg procedure⁹¹. Only significantly recurrent
juxtapositions which did not occur at the same nucleotide position, had a mean SvABA-assigned quality score of
636 larger than 20, included at least one SV detected with post-assembly (ASDIS or ASSMB) evidence, occurred in more
than two samples, and had a q-value smaller than 0.05 were considered for further analysis.

638

SV signature analysis

640 SV signature analysis followed previously published approaches^{23,26}, with slight modifications for overall lower SV
counts in pediatric tumors. In brief, SVs were stratified according to the span between the two breakpoints (0-
642 30kbp, 0.03-1Mbp, >1Mbp, interchromosomal); read orientation (deletion, duplication, inversion and
interchromosomal); and whether they were clustered or simple based on the comparison between local SV
644 density and the overall SV density in the sample, as determined by clusterSV²³. This resulted in a count matrix of
20 SV feature combinations x 179 pHGGs, which was analyzed with Bayesian NMF using SignatureAnalyzer^{24,25}
646 code to determine the optimal, most informative number of signatures and signature activities in each sample.

RNAseq analysis

648 RNAseq data were available for 112 of the 179 tumors (57 sequenced *de novo* and 55 previously published^{4,6}). For
650 *de novo* samples, cDNA libraries were prepped as previously described¹⁷ using the Tru-Seq Strand Specific Large-
Insert kit, and sequenced to a depth of 50 million paired ends using Illumina Hi-Seq machines. Reads from both
652 *de novo* and published data were aligned to the hg19 reference genome using STAR⁹² and quantified with RNA-
SeQC⁹³ following the GTEX analysis pipeline⁹⁴. Counts were normalized using the VST transform as implemented
654 in DESeq2⁹⁵ and batch-corrected with COMBAT⁹⁶ as implemented in sva⁹⁷.

ChIP-seq analysis

656 Previously published H3K27ac ChIP-seq sequencing data from primary DMGs were downloaded from
658 GSE128745⁴². Peaks were called using MACS2⁹⁸ callpeak with -B -SPMR to save the fragment pileup per million

reads track. The resulting bdg-files were used to calculate fold enrichment and q-value tracks with macs2 bdgcmp
660 which were transformed into bigwig files with rtracklayer and visualized together with copy number and SV calls
in gtrack. Additional bigwig files for adult GBM H3K27ac, pHGG ATAC seq and non-cancerous or non-brain tissues
662 were downloaded from GSE54792⁴³, GSE126319⁴⁴ and the Encode project^{45,46} respectively.

664 *Hi-C*

Library generation and sequencing

666 *In-situ* Hi-C libraries were generated from 5 million cultured H3.3^{K27M} glioblastoma cell lines HSJ-019 and HSJ-031
as well as from a H3.3^{K27M} primary tumor HSJ-031 following published protocols⁹⁹ with minor modifications.
668 Briefly, in situ Hi-C consists of 7 steps: (1) crosslinking cells with formaldehyde, (2) digesting the DNA using a 4-
cutter restriction enzyme (e.g., DpnII) within intact permeabilized nuclei, (3) filling in, biotinylating the resulting
670 5'-overhangs and ligating the blunt ends, (4) shearing the DNA, (5) pulling down the biotinylated ligation junctions
with streptavidin beads, (6) library amplification and (7) analyzing these fragments using paired end sequencing.
672 Quality control (QC) for efficient sonication was performed through the combination agarose DNA gel
electrophoresis and for appropriate size selection using the Agilent Bioanalyzer on final amplified libraries,
674 followed by low-pass sequencing on the Illumina HiSeq 2500 (~30M reads/sample) to assess quality of the libraries
using percent of reads passing filter, percent of chimeric reads, and percent of forward-reverse pairs.

676 Data processing

Additional Hi-C files for NPCs (neural progenitor cells) were downloaded
678 from www.synapse.org/#!Synapse:syn12979101 (registration required; Data Download - Study "iPSC-HiC")¹⁰⁰.
Analysis of Hi-C generated and downloaded fastq files was performed using Juicer and the associated Juicer
680 Tools¹⁰¹. Contact maps were generated using Juicer with the following parameters: -s DpnII -g hg19. Map
resolution was determined by using Juicer's "calculate_map_resolution.sh" script. Hi-C contact maps and
682 associated annotations were visualized using Juicebox¹⁰¹. The HIFI algorithm¹⁰² was used to process 5-kb
resolution Hi-C data to obtain higher accuracy estimates of interaction frequencies, using the following
684 parameters: bandSize=1000, outputNormalized, boundaryKS=1000.

Luciferase reporter

686 Cell lines.

The tumor tissue derived pHGG cell line DIPG13¹⁰³ was a gift from the Michelle Monje lab. Cell lines were grown
688 in ultra-low attachment flasks in a 1:1 ratio of Neurobasal A (Gibco) and DMEM/F-12 (Gibco) and one percent of
each HEPES Buffer Solution 1M, Sodium Pyruvate Solution 100nM, MEM Non-Essential Amino Acids Solution
690 10mM, Glutamax-I Supplement, and Penicillin/Streptomycin solution. The culture medium was supplemented
with epidermal and fibroblast growth factor (H-EGF & H-FGF; StemCell Tech, Inc.) at 20 ng/mL, platelet derived
692 growth factors (H-PDGF-AA & H-PDGF-BB StemCell Tech, Inc.) at 10 ng/mL, Heparin Solution (0.2%; StemCell Tech,
Inc.) at 2 ug/mL, and 50X B-27 Minus Vitamin A (Invitrogen). Cells were regularly passaged every 2-4 days to avoid
694 confluence and neurospheres were dissociated into single cells at the time of passage using Accutase (StemCell
Tech, Inc.) to prevent overly tight neurospheres.

696

Cell line authenticity and mycoplasma surveillance

698 SNP-based fingerprinting assays were performed at regular intervals to ensure authenticity of cell-lines. All cell
lines were routinely monitored for mycoplasma infection using the MycoAlert Mycoplasma Detection Kit (Lonza),
700 following the manufacturer's protocol.

702 Luciferase reporter construction

703 A lentiviral firefly luciferase reporter system was constructed from pGL4.26 (Promega) and the pLKO.1 backbone
704 via Gibson Assembly. The pLKO.1 backbone was digested with FastDigest® KflI and EcoRI. The minimal promoter
705 firefly reporter cassette was PCR amplified from pGL4.26 using the lucminP primer set (Supp. Table 6) using NEB
706 Q5 polymerase. These two fragments were assembled into the lentiviral firefly luciferase reporter using the
707 NEBuilder HiFi DNA Assembly Cloning Kit according to manufacturer's instructions. The DNA sequence in the
708 H3K27ac peak in the consensus CCDC26-SV amplicon was split into two fragments (E1/E2) and PCR amplified from
709 DIPG13 genomic DNA with the primers listed in Supp. Table 6 using NEB Q5 polymerase. The resulting E1 and E2
710 fragments were cloned into the vector, upstream of the minimal promoter in the lentiviral firefly luciferase
711 reporter system using KPN1 and NHE1 restriction sites. The lentiviral constitutively active pLX313-Renilla construct
712 was obtained from addgene (Plasmid # 118016) to serve as intrinsic control. All reporter experiments were
713 conducted in the DIPG13 cell line constitutively expressing Renilla luciferase from this pLX313-Renilla construct.

714

Viral Production

715 HEK-293T cells were cultured in T75 tissue culture treated flasks in DMEM (Gibco) supplemented with 10% FBS
716 (Gemini Bio). Lipofectamine-3000 (Invitrogen) was used to transfect with plasmid of interest in addition to
717 packaging plasmids VSV-G and psPAX2, according to manufacturer's protocol. Media was replaced with DMEM
718 supplemented with 20% FBS six hours after transfection. Media was harvested 24-48 hours post-transfection and
719 virus concentrated (20x) using Lenti-X Concentrator (Takara Bio) per manufacturer's protocol.

722 Lentiviral infection

723 Cells were dissociated and plated in a 12-well tissue culture plate at a density of 1.5 million cells/mL. Concentrated
724 virus was added to the media and the cells were centrifuged for 120 minutes at 2000 RPM and 30 C. Following
725 centrifugation, cells were harvested and placed into T-75 ultra-low attachment flasks. Selection (1µg/ml
726 puromycin for the firefly reporter, 300µg/ml for pLX313-Renilla) was added the following day to achieve survival
727 of 40-80% in the infected conditions.

728

Luciferase Reporter Readout

730 The Dual-Glo Luciferase Assay System (Promega) was used following manufacturers protocols for all
731 measurements. In brief, the firefly reading for each reporter condition was first normalized to the constitutive
732 Renilla luciferase signal. The final reporter values are gives as fold change of the E1 or E2 normalized firefly
733 luciferase signal compared to the empty lentiviral luciferase reporter four days post spinfection (two days post
734 puro selection).

736 *Visualization and reconstruction of complex MYCN and RTK amplicons*

737 JaBbA⁹⁰ was used to generate cancer genome graphs using SvABA SV, GATK CNV and absolute purity and ploidy
738 as inputs. The cancer genome graphs combined with ChIP-seq tracks were visualized in gGnome
739 (<https://github.com/mskilab/gGnome>) / gTrack (<https://github.com/mskilab/gTrack>), which were also used to
740 calculate distances between loci in the cancer genome. To infer the structure of possible extrachromosomal
741 amplicons the genome graphs were subsetted to only segments with CN > 20 and all circular paths in the genome
742 graphs were reconstructed with the gGnome walks() function.

744 To determine which genomic loci were recurrently incorporated into the amplicons, the distribution of amplicons
745 around the oncogene was evaluated using methods analogous to previously published approaches^{20,21}. First, the
746 TADs adjacent to the amplified oncogene were divided into 10kbp windows. The average copy number per 10kbp

748 window was calculated in all tumors with an amplicon of CN >5 anywhere in the TAD of the oncogene (using the
germinal zone TAD boundaries from GSE77565⁸⁶). Likewise, among tumors with an amplification of CN>5
750 anywhere within the TAD of the oncogene, the fraction of tumors with an amplification in each 10kbp window
was determined. If the fitness advantage provided by the amplicons resulted only from their inclusion of the
752 oncogene, one would expect a symmetric normal distribution with the peak at the location of the coding sequence
of the oncogene for both mean amplification level and fraction of tumors with amplifications in the 10kbp
754 windows around the oncogene. The location of likely enhancer elements, necessary to drive expression of the
amplified oncogene, was inferred from the direction of the skew of the observed distribution compared to the
expected symmetric normal distribution.

756

SNV signature analysis

758 *De novo* SNV signature extraction was performed using Bayesian NMF in SignatureAnalyzer, as previously
described²⁴. The resulting SNV signatures were compared to the COSMICv3 SBS signatures using cosine similarity
760 to annotate known etiologies and signature names.

Signature integration and definition of signature clusters with similar variant generating processes

762 To better understand the information contained in each of the five SV and 14 SNV signatures, consensus clustering
was applied to the tumor x signature proportion matrix, comprising the 19 values representing the proportion of
764 each of the SV and SNV signatures out of all SV and SNV signatures in each tumor. The proportions for each
signature were median-centered across all tumors before consensus clustering with the ConsensusClusterPlus R
766 package¹⁰⁴ using the parameters: reps=1000, pltem=0.9, pFeature=0.9, clusterAlg="hc", distance="spearman". The
resulting three most stable and informative clusters were named "Complex-SV", "Intermediate" and "Simple-SV",
768 after the signature with the highest enrichment in the cluster.

770

Comut plots and variant combination matrix

772 An annotated list of all SNVs, SCNAs and SVs was created for all tumors in this cohort. SNVs were annotated using
Oncotator¹⁰⁵. SVs were annotated and linked to a gene based on if the SV breakpoints were located in exons of
774 the gene (named 'coding SV'), intronic ('intron SV') or in the TAD of the gene ('flank SV'). Absolute purity- and
ploidy-adjusted copy number (CN) was determined for every gene using the width-weighted mean CN from all
776 segments overlapping the gene.

778 To create the variant combination matrix, this list of all variants in all genes from all tumors was subsetted to only
contain variants in genes from the Cancer Gene Census⁵³, in addition to genes which showed significantly
780 recurrent variants in this cohort. For SNVs, the variant classification was simplified to truncating_snvs =
("Nonsense_Mutation", "Frame_Shift_Del", "Frame_Shift_Ins", "Splice_Site", 'Start_Codon_SNP',
782 'START_CODON_SNP', 'Translation_Start_Site') and missense.snvs = ("Missense_Mutation", "In_Frame_Del",
"Stop_Codon_Del", 'DE_NOVO_START_IN_FRAME', 'DE_NOVO_START_OUT_FRAME', 'Nonstop_Mutation',
784 'In_Frame_Ins', 'START_CODON_INS'). SCNAs of genes with a ploidy- and purity-adjusted copy number of <0.2
were annotated as 'homdel', CN > 5.4 as amp, CN > 10 as ExChr_amp and amplifications covering only parts of a
786 gene with a CN > 3.1 as 'part.amp' based on the CN histogram across all tumors defining recurrent CN states. Using
the R data.table dcast function on the gene variants and samples, binary matrix of samples x variants was created.
788 A genetic variant had to recur in at least three samples (excluding the hypermutant samples for SNVs) to be kept
in the variant x sample matrix. For SV in the TAD of a gene ('flank SV') this threshold was increased to at least ten
790 because of their higher occurrence probability based on the large size of TADs compared to genes. The presence
or absence of the GISTIC peaks in each sample was used to incorporate the SCNAs of lower amplitude.

792

This procedure resulted in the binary variant combination matrix of 176 unique tumors by 353 variants (Three tumors showed no recurrent variants) which was used as the basis for all enrichment and variant clustering analyses. Additionally, 16 significantly recurrent arm level SCNAs were included for enrichment analyses but excluded from the variant clustering as they are highly correlated with the GISTIC peaks.

798 The cBioportal^{106,107} oncoprinter was used to visualize the variants in the tumor samples. The order of the columns representing the tumor samples within each subgroup was determined by hierarchical clustering (HC). HC with one minus Spearman rank correlation metric was applied on the signature activities in each tumor for Fig. 4B and 800 hc with one minus cosine similarity metric on the respective subsets of variant combination matrix for Fig. 5A and 802 Fig. 6A with average linkage in all cases. Genes of interest were manually selected based on the variants with the highest enrichment in the subgroups.

804

Variant timing analysis

806 Previously published approaches^{61,108} were leveraged to evaluate if a potential driver variant occurs early or late in tumor development. The palimpsest¹⁰⁹ R package was applied to determine single patient timing of SNVs, as 808 previously described¹⁰⁸. In brief, SNVs were classified into clonal vs. subclonal based on their cancer cell fraction (CCF, calculated from the variant allele fraction adjusted by local copy number and purity/ploidy for the respective 810 sample). SNVs overlapping with SCNA could further be timed into early or late depending on if they happened before or after the SCNA⁶¹. MutationTimer⁶¹ was used to determine the timing of SCNAs in individual patients, 812 including clone clusters as input. Mobster¹¹⁰ was used to define the clone clusters based on the distribution of the absolute⁸³ CCFs. The resulting molecular time for the SCNA segments was assigned to the GISTIC peaks present in 814 the respective sample with a width-weighted mean and categorized based on the timing quarters. For each subgroup, in addition to the timed GISTIC peaks, single patient timed SNVs in consensus cancer genes were tallied 816 into winning tables reflecting the frequency of this variant being an early event using published code¹⁰⁸. In the resulting winning tables each variant competes against all other variants in the table. For each comparison the 818 earlier variant wins. The BradleyTerryScalable R package was applied to estimate the winning probability across the subgroup for each variant, which in this setting is a measure for the probability of this variant being an early 820 event in the tested subgroup. The Bayesian maximum *a posteriori* probability (MAP) estimate was used to fit the model as previously described¹⁰⁸. To control for outlier samples the analysis was performed on 100 random 822 samples of 70% of each subgroup. The resulting distributions for the strength parameters (on log scale) were plotted for variants recurrent at least three times in the tested subgroup.

824

Survival analysis

826 Univariate correlations for differences in survival were analyzed using the Kaplan-Meier method and significance was determined with a log-rank test. Spearman rank correlation tests were used to determine correlations 828 between overall survival and the complex-SV signature activity. This was possible because all children with DMGs died within the observed period, resulting in an absence of censored data. Variables included in the multivariate 830 analysis (Cox-model) were Histone-SNV, age, TP53 status combined Complex-SV signature activity.

Statistical analysis

832 All statistical analyses were performed in R 3.6.3. Unless otherwise indicated, statistical comparisons were 834 performed using Fisher's exact tests or Wilcoxon tests, as appropriate. p-values < 0.05 were considered significant. Multiple testing was accounted for by using false discovery rate¹¹¹ q-values unless otherwise indicated. In all box

836 plots the boxes represent the range between the 25th and 75th percentiles and the central line indicates the median.

838

Figures and figure legends:

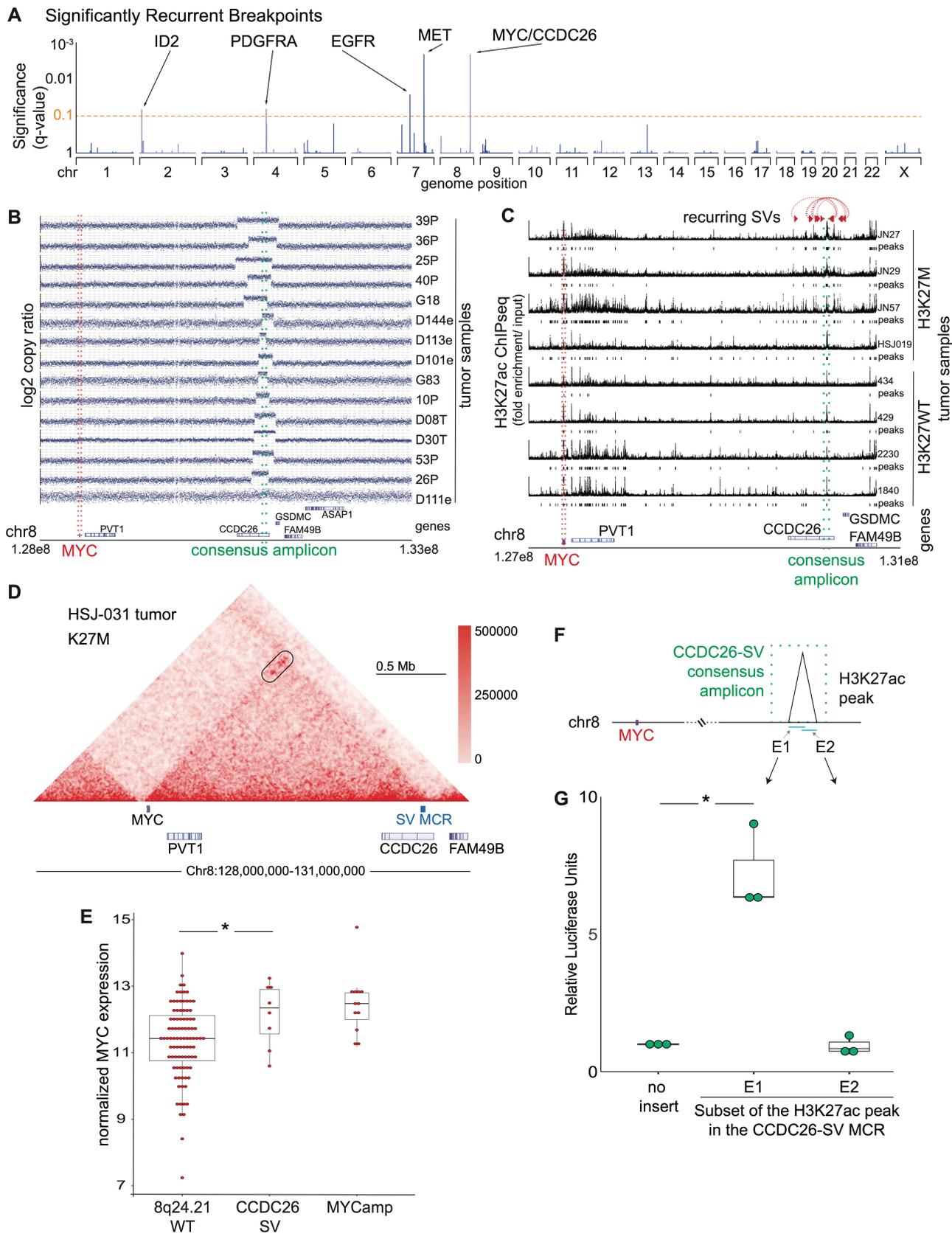


Figure 1

842 Figure 1

844 **Significantly recurrent breakpoints within CCDC26.** (A) Significance (multiple hypothesis corrected q-values, vertical axis) of recurrent breakpoints (genomic positions on the horizontal axis) across the 179 pHGG genomes. (B) Copy-number profiles across the MYC TAD for the 15 tumors with the recurrent CCDC26 SV. (C) H3K27ac ChIP-seq tracks within the TAD containing MYC (red lines) and CCDC26, for four H3^{K27M} and four H3K27^{WT} pHGGs tumors. Significantly enriched peaks (q-value < 0.01) are indicated below each H3K27ac ChIP-seq track. The CCDC26 amplicon boundaries for individual samples are indicated by the paired red arrows at the top. The consensus amplicon is indicated by the green dotted lines, and centers on an H3K27ac peak. (D) Hi-C heatmap across the MYC-CCDC26 locus from a midline glioma. Increasing interaction frequencies are indicated by brighter shades of red. The black oval outlines the frequent connections between MYC and the H3K27ac-enriched region within CCDC26. The minimal common region of amplification (SV MCR) is indicated at the bottom. (E) Normalized MYC expression in DMG samples with wild-type copy-number profiles at 8q24.21 (n= 92), CCDC26-SVs (n=8) or amplifications of the MYC coding sequence (n=12). *denotes a p<0.05 as determined by Wilcoxon rank sum test. (F) Schematic illustrating the luciferase reporter used to validate the enhancer in *CCDC26*, showing the positions of the E1 and E2 sequences with respect to the enhancer within *CCDC26*. (G) Normalized luciferase activity in DIPG13 cells following transduction of the E1 and E2 enhancer reporters or empty vector controls. Values represent the average of four technical replicates in each of three independent experiments. * denotes p=0.019, T-test (unpaired, two-sided).

860

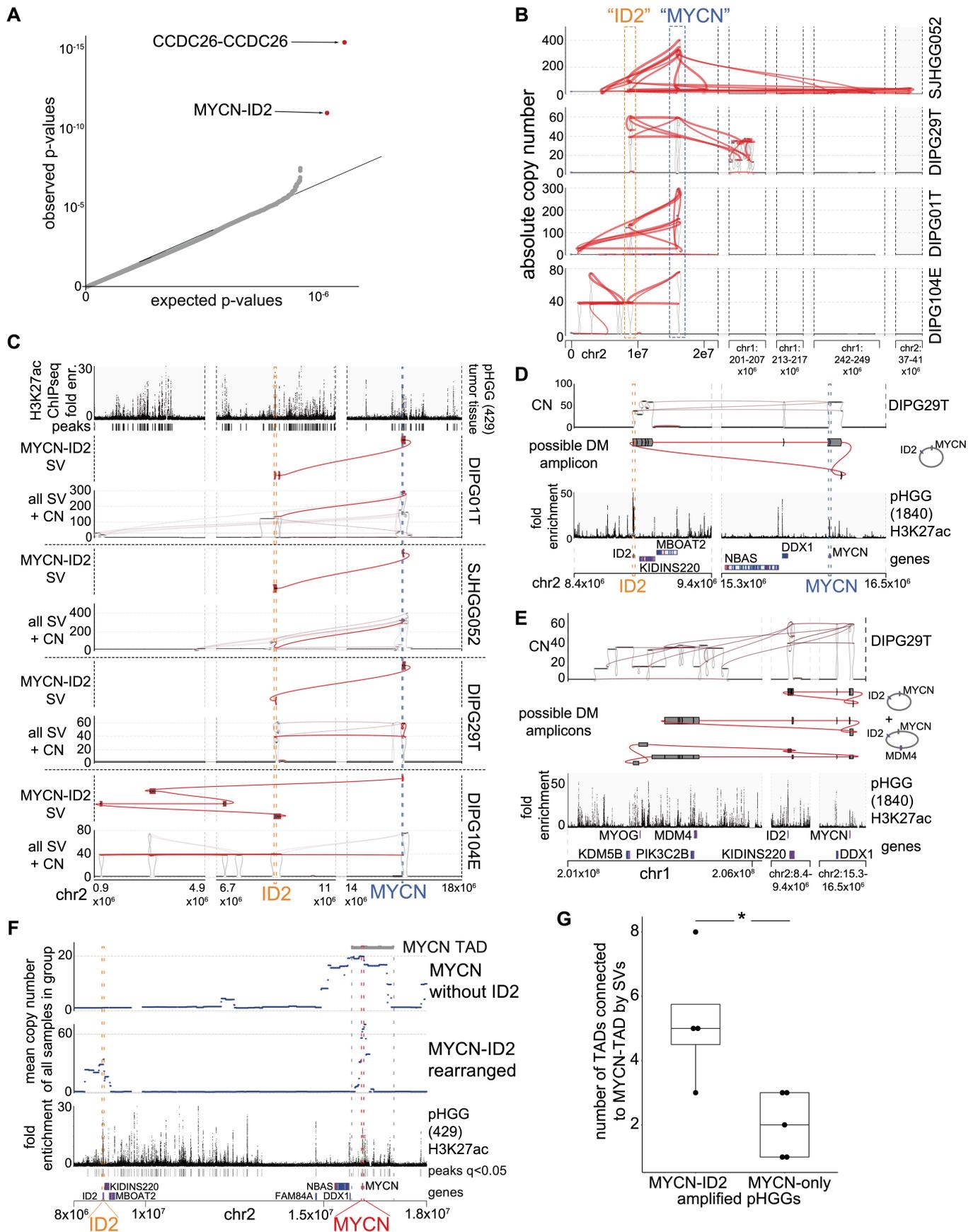
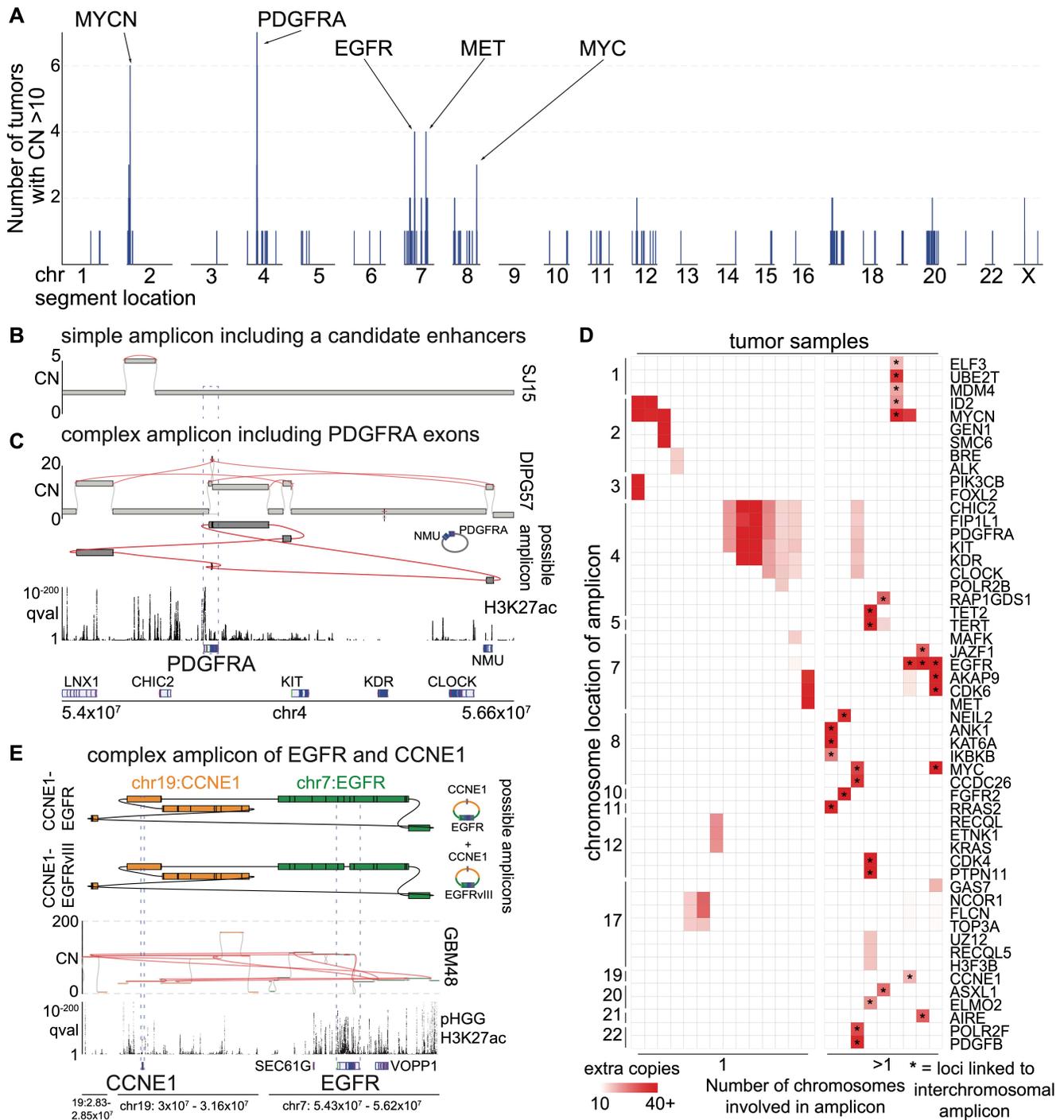


Figure 2

864 **Significantly recurrent juxtapositions: MYCN-ID2.** (A) Quantile-Quantile plot indicating the significance of
866 juxtapositions between pairs of genomic loci. SVs that reached statistical significance are depicted in red. (B) SVs
(red lines) involving MYCN and ID2 in samples with MYCN-ID2 rearrangements, and the number of copies at each
868 connected locus (vertical axis). The dashed boxes indicate loci encompassing ID2 and MYCN. (C) SV maps as in
panel B, with SVs juxtaposing ID2 and MYCN highlighted in red. The top track indicates H3K27ac marks in a pHGG
870 without a known MYCN-ID2 rearrangement, showing a strong enhancer within ID2. In each case, the MYCN-ID2
juxtaposition reduces the somatic distance MYCN and ID2 (each indicated by a dashed line) from 7 Mbp to less
872 than 700kbp. (D-E) Examples of possible reconstructions of circular extrachromosomal amplicons containing
MYCN and *ID2* from a single DMG. (D) The simplest possible circle connecting *MYCN* and *ID2*. The top track shows
874 copy-number and SVs, the middle track indicates the reconstructed topology, and the bottom track shows
H3K27ac binding at the indicated loci in a different pHGG. (E) The chr1 loci connected to the *MYCN-ID* complex in
876 this DMG. Short-read reconstructions allow for several extrachromosomal amplicons incorporating *MYCN-ID2* and
MDM4. The difference in the copy number could be explained either by a mix of amplicons containing respectively
MYCN-ID2 alone and *MYCN-ID2-MDM4* or by more complex amplicons incorporating multiple copies of *MYCN-*
878 *ID2* for each copy of *MDM4*. (E) MYCN amplicons in tumors without ID2 amplification (top track) incorporate a
larger fraction of the MYCN TAD (60%) relative to MYCN amplicons in tumors with MYCN-ID2-SV (second track;
880 23%, $p=0.03$, T-test). As a result, the former tend to include more loci with H3K27ac enrichment (third track) from
the MYCN-TAD; significantly enriched H3K27ac peaks (q -value < 0.01) are indicated with black bars between the
882 H3K27ac fold enrichment and the gene track. (F) MYCN-ID2 amplified pHGGs contain significantly more SVs
between the MYCN-TAD and other TADs than pHGGs amplifying only MYCN and its neighborhood. ($p = 0.03$,
884 Wilcoxon rank sum test).

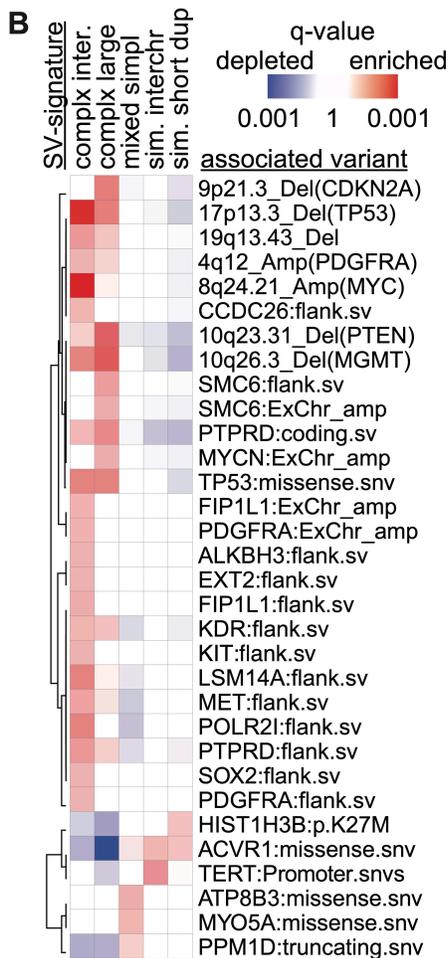
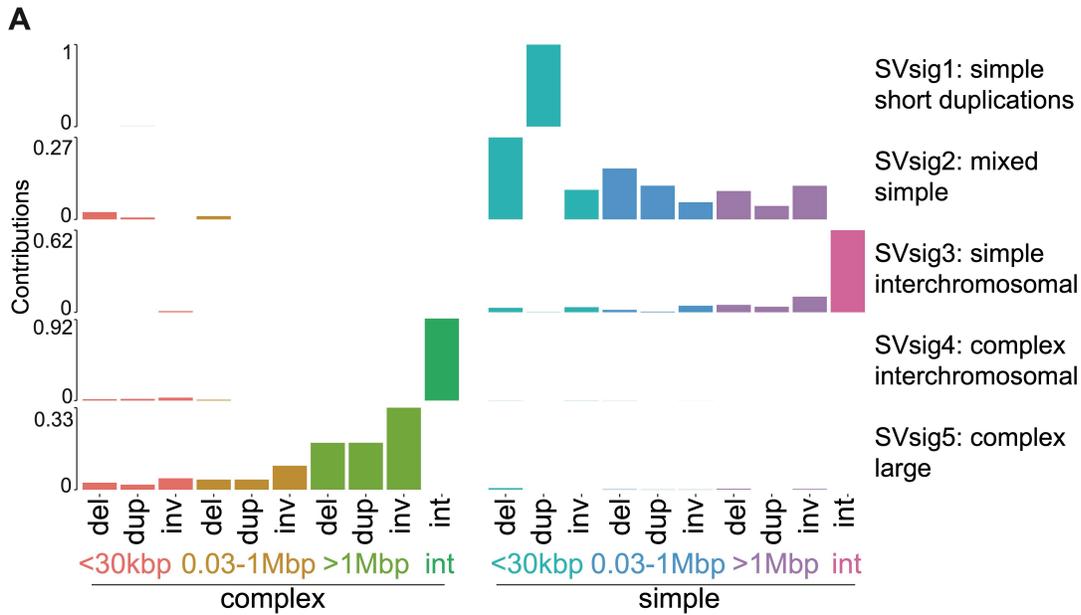
886



888 **Figure 3**

890 **High-level amplicons.** (A) For each genomic locus (horizontal axis), the number of tumors containing a high-level (CN>10) amplicon is indicated (vertical axis). SRBs are highlighted at the top. (B-C) Simple and complex-SVs exhibit distinct mechanisms to active PDGFRA. The top and bottom tracks indicate copy-numbers and the significance of H3K27ac enrichment (as calculated from eight pHGGs), respectively. SVs are highlighted in red. Selected gene loci are indicated on the bottom. (B) A simple amplicon of a region with known^{19,52} PDGFRA enhancers. (C) A complex high-level PDGFRA amplicon, displayed as in panel (B) with the addition of a track (second from top) indicating the topology of the amplicon. The complex-SV cluster around PDGFRA connects several segments on chr4, which are amplified to ten absolute copies. The SV calls support the reconstruction of an extrachromosomal amplicon incorporating PDGFRA exons and these segments. (D) Cancer genes involved in high-level amplicons (>10 copies)

898 within the cohort. 9/23 tumors (grouped on right) contain high-level amplicons encompassing loci from two or
more chromosomes. These linked loci are marked by *. The color of each cell represents the number of extra
900 copies due to the amplicon. (E) Example of a tumor with an extrachromosomal amplicon including two oncogenes
from different chromosomes. This tumor shows a cluster of SVs connecting the *EGFR* and *CCNE1* loci. The regions
902 of both oncogenes are amplified to different CNs but in both cases reach several dozen absolute copies. (top) The
complexity of the SVs allows for the reconstruction of several possible extrachromosomal amplicons. The CN
904 differences in the bulk profile (middle) could be explained by either a mix of different circles or by more complex
circles incorporating some segments repeatedly. The SV calls also reveal that a small fraction of the *EGFR*
906 amplicons in this patient already show the EGFRvIII variant. The bottom two tracks show the genes of interest at
the location and q-value H3K27ac track calculated from eight pHGG tumors.
908

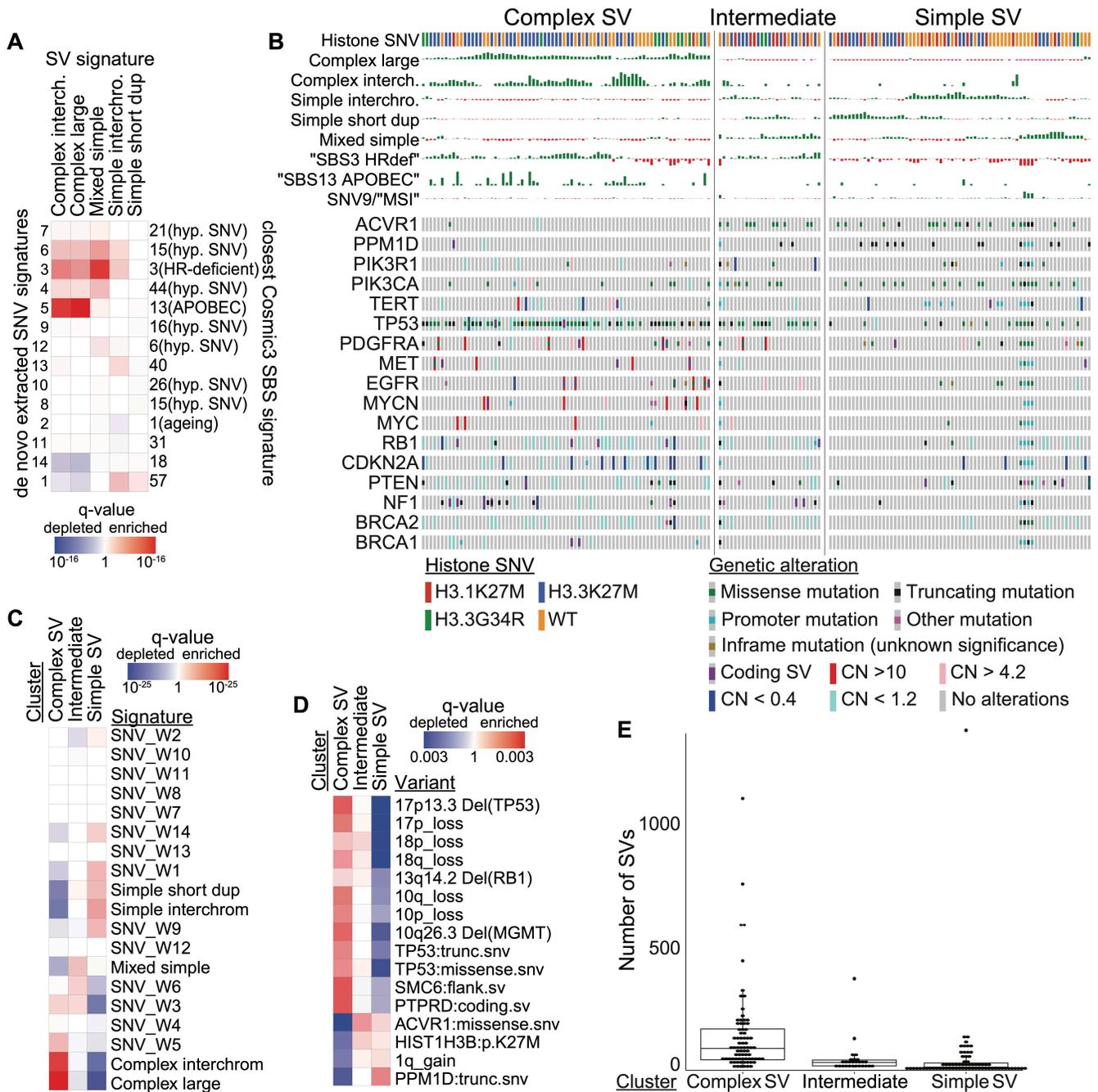


910 Figure 4

912 **SV signatures in pHGG.** (A) The horizontal axis indicates the size and type of SVs. Del stands for deletion, dup for duplication, inv for inversion, and int for interchromosomal rearrangement. The vertical axis indicates the fraction of SVs within each signature that are contributed by each SV type. (B) The statistical significance of

914 positive (enriched) and negative (depleted) associations between each SV signature and of all recurrently altered
somatic genetic alterations that are documented in the Cancer Gene Census⁵³.

916



918 **Figure 5**

920 **Integrated SV and SNV signature analyses reveal three pHGG subgroups with distinct variant generating processes.** (A) Correlations between SV and SNV signatures. Signature labels from this analysis are indicated on the left; the nearest COSMICv3 signatures are indicated on the right, with their proposed mechanisms in parentheses. Complex-SV signatures show a close correlation with APOBEC and homologous recombination deficiency SNV signatures (SBS3). q-values are based on Spearman rank correlations. (B) Consensus clustering of the normalized SNV and SV signature activities in each tumor sample (columns). Rows indicate signature activities (top) and potentially oncogenic variants (bottom). (C) Enrichment analysis for signature activities in each cluster from panel B. FDR q-values are based on Wilcoxon tests. (D) Significance of signature cluster associations for all variants with correlations reaching $q < 0.1$; q-values are based on Fisher's exact tests. Tumors in the complex-SV clusters are enriched for copy-number changes in cancer genes and SNVs in *TP53*, whereas simple-SV pHGGs tend

922
924
926
928

930 to exhibit SNVs in different cancer genes. (E) Number of SVs per tumor in each cluster. All differences are significant to $q < 0.003$ by Wilcoxon tests.

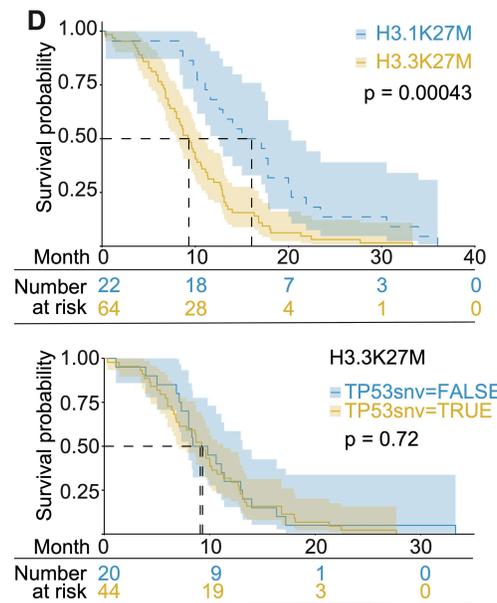
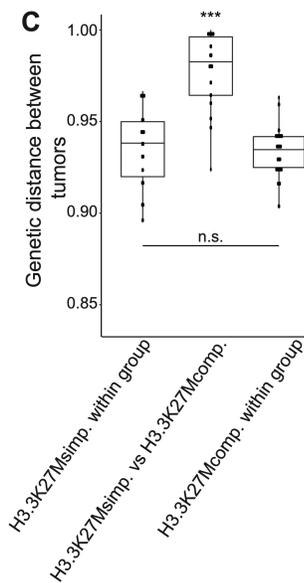
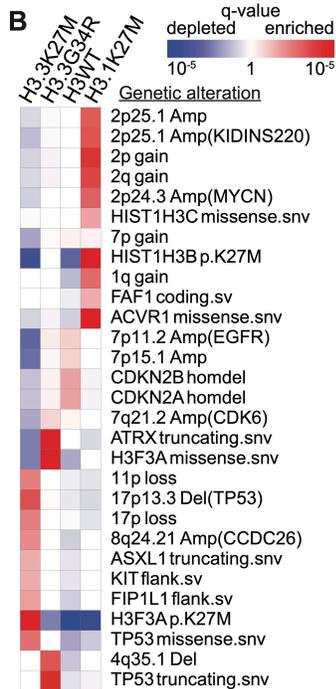
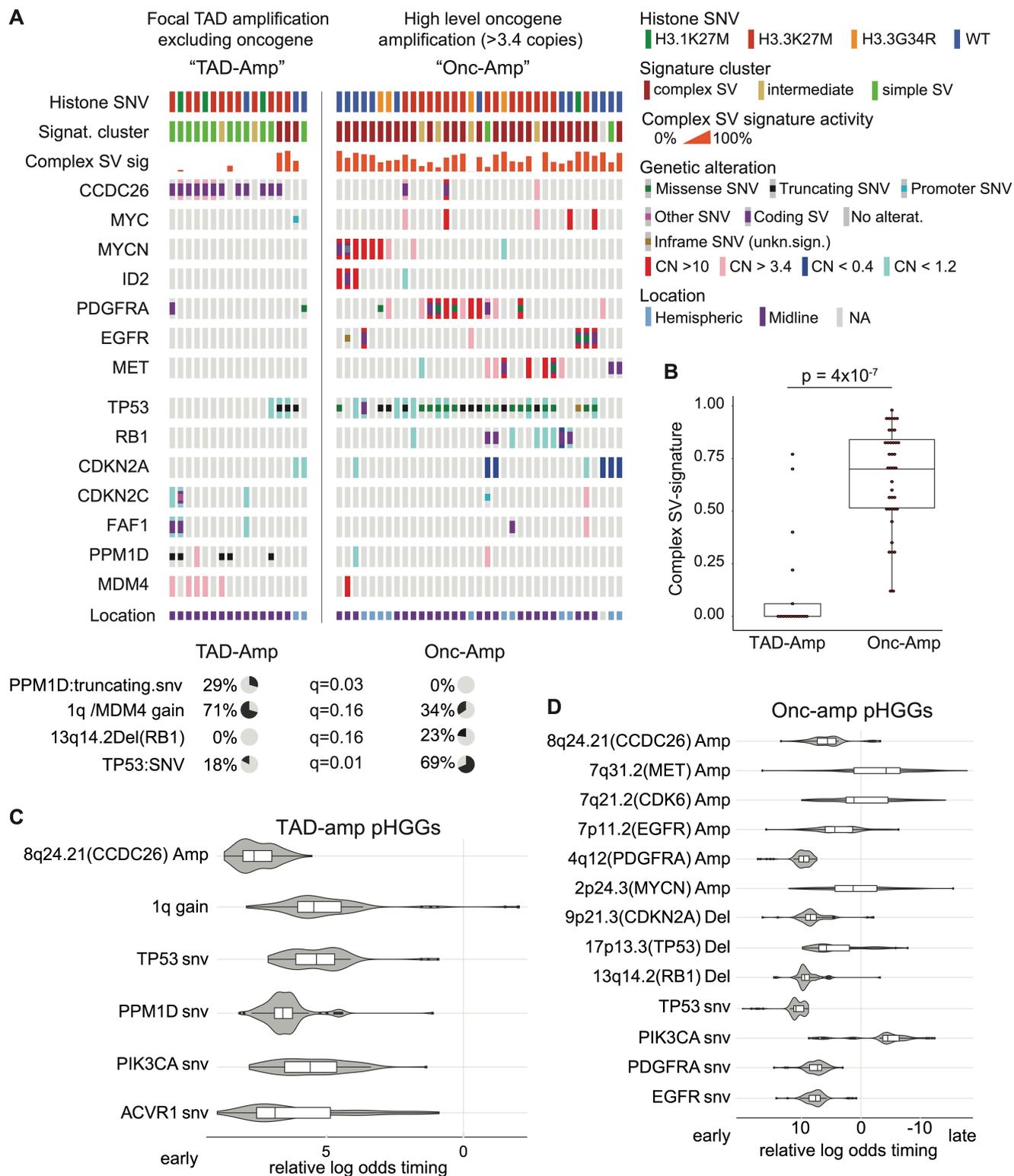


Figure 6

934 Figure 6

936 **Overview of somatic variants and associated features within histone mutation-defined subgroups. (A)** Co-mut
plot of the 176/179 (98.3%) tumors with somatic variants in at least one well-known oncogene. Columns represent
938 tumors, ordered within histone mutation-defined subgroups by hierarchical clustering of all potential driver
variants. The top two rows show signature metadata. **(B)** Significance of histone subgroup associations for all
940 variants with correlations reaching $q < 0.05$ (based on Fisher's exact tests) within any subgroup. **(C)** Jaccard
distances based upon the variants in each tumor (vertical axis), for pairs of tumors within the H3.3^{K27M} Simple-SV
942 group (left column), within the H3.3^{K27M} Complex-SV group (right column), or paired between these groups (middle
column). H3.3^{K27M} tumors were considered to be in the Complex-SV group if Complex-SV signatures comprised
more than 20% of its SV signature activity. *** denotes $q < 6 \times 10^{-8}$ relative to both other columns by a Wilcoxon
944 test. **(D)** Kaplan-Maier plot indicating overall survival for (top) H3.1^{K27M} and H3.3^{K27M} DMGs and (bottom) H3.3^{K27M}
DMG with and without *TP53* SNVs. p-values are from log-rank tests.



946

Figure 7

948

Context of the significantly recurrent SVs. (SRSVs) (A) Co-mut plot for TAD-amp and Onc-amp SRSVs. TAD-amp SRSVs generate focal amplifications in the TAD of an oncogene without amplifying the protein coding sequence; Onc-amp SRSVs generate high-level (CN >3.4) amplifications or fusions of the coding sequence. The top three rows show associated metadata. The next seven rows indicate the genes affected by the SRSVs. The bottom seven rows show genes in DNA damage response pathways. Significant associations with the two groups are illustrated with pie charts below the plot, based on Fisher’s exact tests. (B) TAD-amp pHGGs show significantly lower combined

950

952

954 Complex-SV signature activity than Onc-amp pHGGs ($p = 4 \times 10^{-7}$, Wilcoxon). **(C-D)** Timing analysis of somatic variant
956 acquisition in TAD-amp **(C)** and Onc-amp **(D)** pHGGs based on a Bradley-Terry model. The horizontal axis shows
958 the log odds of the variant being an early event. The distributions indicate the results of 100 random subsamples
of the data. Only variants in the SRSV-affected genes and pathways (growth factor and MYC signaling) and in DNA
damage response genes altered in more than two samples are shown.

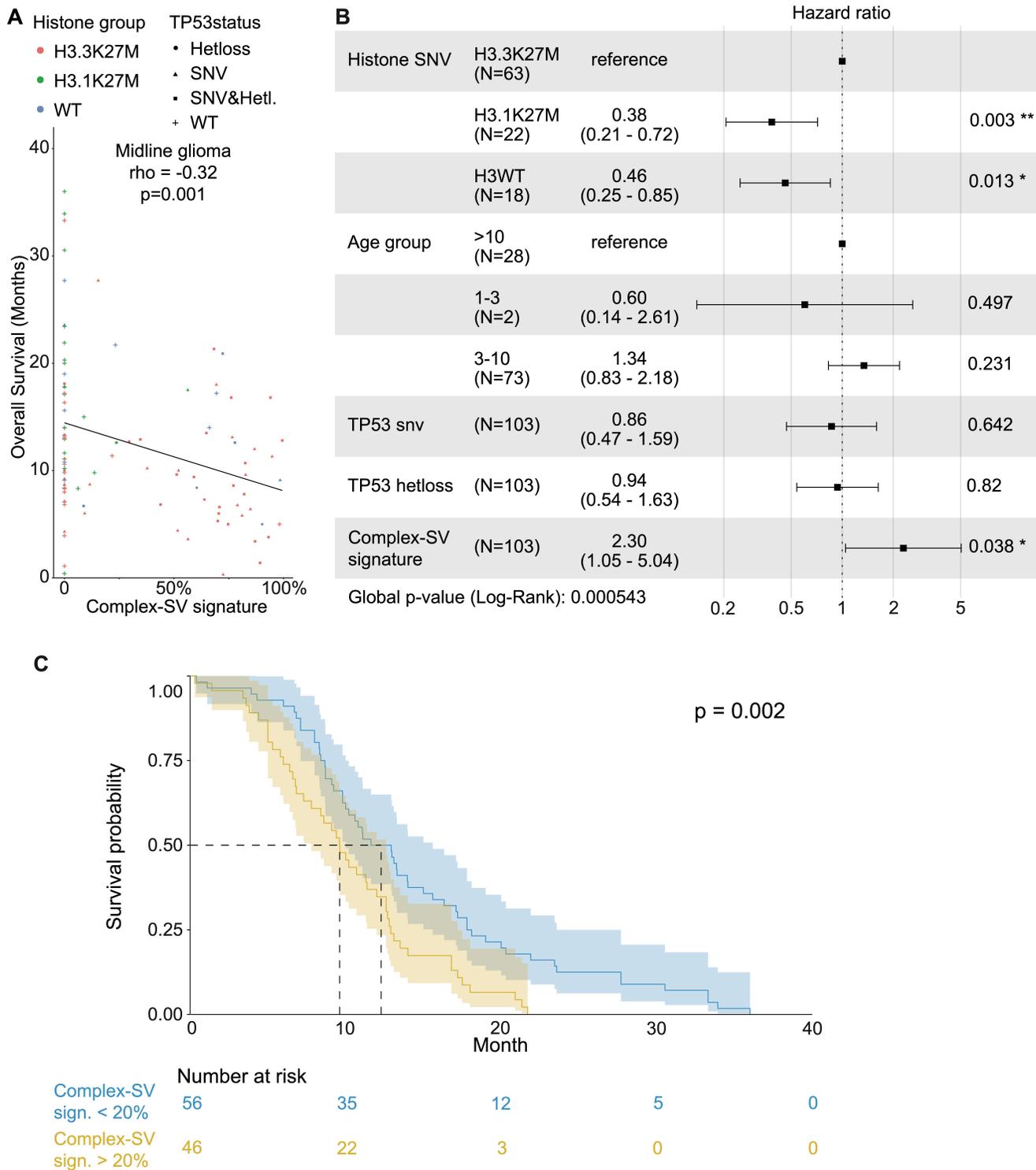


Figure 8

Association between the complex-SV signature and overall survival in midline gliomas. (A) Associations between the fraction of SV signature activity attributed to Complex-SV signatures and overall survival for all midline gliomas (n= 103; rho = -0.32 and p=0.001, Spearman correlation). (B) Cox proportional hazards analysis incorporating histone group, age, TP53 status and the combined Complex-SV signature. (C) Kaplan-Maier plot of a univariate analysis of the association between the Complex-SV signature and overall survival, in which tumors were classified as Complex-SV if Complex-SV signatures contributed more than 20% of all SV signature activity. The number of patients in each group is indicated below the plot. The p-value represents a log-rank test.

Extended Data Figure 1

970 **Sample characteristics and significantly recurrent variants.** (A) Purity of pre-treatment biopsy and autopsy
972 samples were not significantly different ($p = 0.5$, Wilcoxon). (B) Significantly recurrent SNVs in non-hypermutant
974 tumors. (C) Significantly recurrent SCNAs. All of these SCNAs have been noted³ except for a novel non-protein-
coding locus in 8q.24.21, near *MYC*—which is also within a separate recurrently amplified locus. (D) Q-Q plot for
the analysis of significantly recurrent SV breakpoints. The most significantly recurrent breakpoints are within the
long non-coding RNA *CCDC26*, within the TAD encompassing *MYC*.

976

Extended Data Figure 2

978 **Lineage specificity of the enhancer peak in *CCDC26*.** (A) ATAC-seq (top) and H3K27ac ChIP-seq (bottom)
980 enrichment (vertical axis) of samples from different lineages (indicated on right; “27y” indicates the sample was
obtained from a 27-year old person) across the TAD encompassing *MYC* (horizontal axis). The location of the *MYC*
coding sequence is highlighted in red. The *CCDC26* amplicon boundaries for the 15 samples with the amplicon are
982 indicated by the paired red arrows at the top. The consensus amplicon is indicated by the green dotted lines and
centers on an H3K27ac peak present only in glial samples. (B) Hi-C heatmaps depicting DNA interaction profiles (5
984 kb resolution) from iPSC-derived neural progenitor cells (top) and two cell lines harboring H3.3^{K27M} mutations
(middle and bottom). Red and white indicate high and low interaction frequencies, respectively. *MYC* interacts
986 more frequently with the H3K27ac peak within *CCDC26* (black oval) relative to neighboring loci. The minimal
common region of the *CCDC26* amplicon is indicated at the bottom of the heatmaps (SV MCR; blue rectangle).

988

Extended Data Figure 3

990 **Significantly recurrent juxtaposition between *MYCN* and *ID2*** (A) (top) Count matrix showing all possible
juxtapositions between pairs of genomic loci. (bottom) Illustration of the principle behind the analysis of recurrent
992 juxtapositions, as exemplified by the *MYCN-ID2* loci. First, we count the number of SVs connecting each pair of
genomic loci. Using a background model for the probability of juxtapositions generated from an analysis of 2658
994 cancers²², we then determine the probability of observing this number of SVs due to chance alone, corrected for
multiple hypothesis testing. This analysis revealed the *MYCN-ID2* juxtaposition as the only significantly recurrent
996 juxtaposition in the window shown. (B) Overlay of amplification frequencies on ChIP-seq data in the *ID2* and *MYCN*
loci. The top two tracks show, among tumors with *MYCN-ID2* rearrangements (top track) or *MYCN* amplifications
998 without *ID2* involvement (second track), the percentage of tumors with amplifications (y-axis) at each genomic
locus (x-axis). The bottom eight tracks indicate H3K27ac ChIP-seq profiles across these loci for four H3^{K27M} and
1000 four H3^{WT} pHGGs tumors. Coding sequences of *ID2* and *MYCN* are highlighted with yellow and red lines
respectively. Significantly enriched H3K27ac peaks (q -value < 0.01) are indicated below each ChIP-seq track. The
1002 small region at the *ID2* locus that is amplified in all *MYCN-ID2* pHGGs shows an H3K27ac peak in the ChIP tracks
from all six pHGG tumor samples. Tumors that amplify *MYCN* without *ID2* take in a much larger region of the
1004 *MYCN* TAD into the amplicon (60% vs. 23%, $p=0.03$, T-test). (C) G-track plots indicating copy-number profiles and
genome topology after consideration of local SVs, for two examples of pHGGs with focal *MYCN* amplicons without
1006 incorporation of *ID2*. For both tumors the copy number and SV profiles support several possible reconstructions
of extrachromosomal circular amplicons. All are limited to the neighborhood of *MYCN*, presumably incorporating
1008 endogenous enhancers from the *MYCN* TAD.

Extended Data Figure 4

Structures of recurrent RTK amplicons (A) Average amplicon profile for all pHGGs with amplifications in the
1012 *PDGFRA* TAD reaching a copy number (CN) of at least four ($n = 15$). The top track shows the percentage of those

tumors with amplifications (vertical axis) at each location (horizontal axis). The track below shows the average CN
1014 across all tumors with amplifications in the *PDGFRA* TAD. The segments included in the *PDGFRA* amplicon in 80%
of tumors are highlighted in the red box. Most amplicons range over several Mbp, often including *KIT*. (B) Average
1016 amplicon profile for all pHGGs with amplifications in the *EGFR* TAD reaching at least four copies (n = 7), displayed
as in (A). The segments included in all the *EGFR* amplicons are highlighted in the red box. The pHGG *EGFR*
1018 amplicons always include upstream enhancers elements around *SEC61G* known to activate *EGFR* in adult GBM²¹.
(C) Structure of a simple *EGFR*-TAD amplicon that encompasses enhancers that are also amplified in all tumors
1020 with *EGFR* coding sequence amplifications. (D) SVs, CN tracks and reconstructions for all pHGGs with high-level
MET amplifications. The observed high-level *MET* amplicons are a few 100kbp in size. Three out of four *MET*-
1022 amplified pHGGs incorporate a downstream region including an enhancer (see bottom H3K27ac track) into the
amplicon. For all four *MET*-amplified pHGGs possible reconstructions of the extrachromosomal amplicon are
1024 shown above the CN and SV track. (A-D) From bottom to top the tracks show: the genes of interest at the location,
a q-value H3K27ac track calculated from eight pHGG tumor samples, the CN and SV for the indicated tumor at the
1026 location and reconstructions of possible extrachromosomal amplicons if applicable.

1028 Extended Data Figure 5

Associations between SV-defined groups. (A) Jaccard distances between tumor pairs (vertical axis), calculated
1030 from the combination of variants in each tumor, across subgroups of H3^{K27M} DMGs. Tumor groups were
determined by whether mutations were present in H3.1 or H3.3 and whether the combined Complex-SV
1032 signatures contributed more or less than 20% to all SV signature activity. Tumors were paired within or between
these groups, as indicated on the horizontal axis. All differences were significant with $q < 0.005$ (FDR-corrected
1034 Wilcoxon test) unless indicated otherwise. (B) Volcano plot indicating the significance (vertical axis; FDR-corrected
Fisher's exact tests) of associations in H3.3^{K27M}Simple-SV relative to H3.3^{K27M}Complex-SV DMGs, for all genetic
1036 variants observed in at least 10% of tumors in one subgroup. Variants enriched or depleted to $q < 0.15$ are
highlighted in red on the right or in blue on the left, respectively. H3^{K27M} DMGs were considered to be in the
1038 Complex-SV group if Complex-SV signatures comprised more than 20% of its SV signature activity. (C) Association
between *MDM4* expression (vertical axis) and copy number (horizontal axis). *MDM4* gains universally represent
1040 arm-level gains of 1q. *** indicates adjusted $p = 0.004$, ANOVA with Tukey post-test. (D-E) Volcano plot indicating
the significance (vertical axis; FDR-corrected Fisher's exact tests) of associations between genetic variants and
1042 pHGG subgroups (horizontal axis), colored as in B. (E) Enrichment or depletion of arm-level SCNAs in *TP53*-
disrupted (n=97) vs *TP53*^{WT} (n = 77) pHGGs. We considered samples to have *TP53* disruption if the gene exhibited
1044 SNVs (n=88, often with copy loss as well) or copy loss alone (n=9). (F) Enrichment or depletion of arm-level SCNAs
in *TP53*-disrupted (n=56) vs *TP53*^{WT} (n = 39) H3^{K27M} mutant DMGs. *TP53* disruption represented SNVs (often with
1046 copy loss) in 53 samples or copy loss alone in 3 samples. Only significantly recurrent arm-level SCNAs are shown.

1048 Extended Data Figure 6

Timing analysis of somatic variant acquisition in histone mutation-defined pHGG subgroups. For each subgroup
1050 the individual (per-sample) timing of recurrent variants is fed into a Bradley-Terry model. This results in a strength
parameter for each variant which is indicated on the x-axis in log scale and can be interpreted as the relative log
1052 odds of the variant being an early event in this subgroup. Each distribution indicates the results of 100 random
subsamples of the respective subgroup. Only potential driver variants recurrent in more than two samples are
1054 shown. Subgroups: (A) H3.1^{K27M} (B) H3.3^{K27M} (C) H3.3G34R (D) H3^{WT} (E) hypermutant pHGGs.

1056

Supp. Fig. 1

1058 De-novo extracted SNV signatures.

1060 Supp. Fig. 2

1062 **SV signature activities across all samples in the cohort. (A)** SV signature activity in every tumor. Tumors with
higher SV counts on the left show complex-SV signature activity whereas tumors with lower SV counts on the right
show a mix of simple-SV signatures.

1064

Supp. Fig. 3

1066 **SNV signature activities across all samples in the cohort (A)** Cosine similarity between *de novo* extracted SNV
signatures and the COSMICv3 SBS-signatures. **(B)** SNV signature activity in every tumor. The hypermutant tumors
1068 on the left show signatures associated with hypermutation in COSMICv3. Signature 3, which is similar to the SBS3
homologous recombination deficiency signature, features prominently in many non-hypermutant tumors.

1070

Supp. Fig. 4

1072 **Comparison between pre-treatment biopsy and autopsy samples.** These groups exhibit no significant differences
in **(A)** the number of SVs per sample ($q=0.6$, Wilcoxon) or **(B)** the activity of the combined complex-SV signatures
1074 ($q=0.7$, Wilcoxon).

References

1076

- 1078 1. Ostrom QT, Gittleman H, Truitt G, Boscia A, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report:
1080 Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011–2015.
Neuro Oncol. 2018;20(suppl_4):iv1-iv86. doi:10.1093/neuonc/now131
- 1082 2. Puget S, Philippe C, Bax DA, et al. Mesenchymal Transition and PDGFRA Amplification/Mutation Are Key
Distinct Oncogenic Events in Pediatric Diffuse Intrinsic Pontine Gliomas. *PLoS One.* 2012;7(2):e30313.
<https://doi.org/10.1371/journal.pone.0030313>.
- 1084 3. Mackay A, Burford A, Carvalho D, et al. Integrated Molecular Meta-Analysis of 1,000 Pediatric High-Grade
and Diffuse Intrinsic Pontine Glioma. *Cancer Cell.* 2017;32(4):520-537.e5.
1086 doi:<https://doi.org/10.1016/j.ccell.2017.08.017>
- 1088 4. Project ICGCPT, Bender S, Gronych J, et al. Recurrent MET fusion genes represent a drug target in
pediatric glioblastoma. *Nat Med.* 2016;22:1314. <https://doi.org/10.1038/nm.4204>.
- 1090 5. Buczkowicz P, Hoeman C, Rakopoulos P, et al. Genomic analysis of diffuse intrinsic pontine gliomas
identifies three molecular subgroups and recurrent activating ACVR1 mutations. *Nat Genet.*
2014;46(5):451-456. doi:10.1038/ng.2936
- 1092 6. Wu G, Diaz AK, Paugh BS, et al. The genomic landscape of diffuse intrinsic pontine glioma and pediatric
non-brainstem high-grade glioma. *Nat Genet.* 2014;46(5):444-450. doi:10.1038/ng.2938
- 1094 7. Taylor KR, Mackay A, Truffaux N, et al. Recurrent activating ACVR1 mutations in diffuse intrinsic pontine
glioma. *Nat Genet.* 2014;46(5):457-461. doi:10.1038/ng.2925
- 1096 8. Project SJCRHUPCG, Wu G, Broniscer A, et al. Somatic histone H3 alterations in pediatric diffuse intrinsic
pontine gliomas and non-brainstem glioblastomas. *Nat Genet.* 2012;44:251.
1098 <https://doi.org/10.1038/ng.1102>.
- 1100 9. Schwartzenuber J, Korshunov A, Liu X-Y, et al. Driver mutations in histone H3.3 and chromatin
remodelling genes in paediatric glioblastoma. *Nature.* 2012;482:226.
<https://doi.org/10.1038/nature10833>.
- 1102 10. Bigner SH, Wong AJ, Mark J, et al. Relationship between gene amplification and chromosomal deviations
in malignant human gliomas. *Cancer Genet Cytogenet.* 1987;29(1):165-170. doi:10.1016/0165-
1104 4608(87)90045-8
- 1106 11. Shoshani O, Brunner SF, Yaeger R, et al. Chromothripsis drives the evolution of gene amplification in
cancer. *Nature.* 2020. doi:10.1038/s41586-020-03064-z
- 1108 12. Koche RP, Rodriguez-Fos E, Helmsauer K, et al. Extrachromosomal circular DNA drives oncogenic genome
remodeling in neuroblastoma. *Nat Genet.* 2020;52(1):29-34. doi:10.1038/s41588-019-0547-z
- 1110 13. Kim H, Nguyen N-P, Turner K, et al. Extrachromosomal DNA is associated with oncogene amplification
and poor outcome across multiple cancers. *Nat Genet.* 2020;52(9):891-897. doi:10.1038/s41588-020-
0678-2
- 1112 14. Turner KM, Deshpande V, Beyter D, et al. Extrachromosomal oncogene amplification drives tumour
evolution and genetic heterogeneity. *Nature.* 2017;543(7643):122-125. doi:10.1038/nature21356
- 1114 15. Xu K, Ding L, Chang T-C, et al. Structure and evolution of double minutes in diagnosis and relapse brain
tumors. *Acta Neuropathol.* 2019;137(1):123-137. doi:10.1007/s00401-018-1912-1
- 1116 16. Northcott PA, Lee C, Zichner T, et al. Enhancer hijacking activates GFI1 family oncogenes in
medulloblastoma. *Nature.* 2014;511(7510):428-434. doi:10.1038/nature13379
- 1118 17. Bandopadhyay P, Ramkissoon LA, Jain P, et al. MYB-QKI rearrangements in angiocentric glioma drive
tumorigenicity through a tripartite mechanism. *Nat Genet.* 2016;48:273.
1120 <http://dx.doi.org/10.1038/ng.3500>.
- 1122 18. Beroukhi R, Zhang X, Meyerson M. Copy number alterations unmasked as enhancer hijackers. *Nat
Genet.* 2017;49(1):5-6. doi:10.1038/ng.3754
- 1124 19. Chen CCL, Deshmukh S, Jessa S, et al. Histone H3.3G34-Mutant Interneuron Progenitors Co-opt PDGFRA
for Gliomagenesis. *Cell.* 2020. doi:<https://doi.org/10.1016/j.cell.2020.11.012>
20. Helmsauer K, Valieva ME, Ali S, et al. Enhancer hijacking determines extrachromosomal circular MYCN

- 1126 amplicon architecture in neuroblastoma. *Nat Commun.* 2020;11(1):5823. doi:10.1038/s41467-020-19452-y
- 1128 21. Morton AR, Dogan-Artun N, Faber ZJ, et al. Functional Enhancers Shape Extrachromosomal Oncogene Amplifications. *Cell.* 2019;179(6):1330-1341.e13. doi:https://doi.org/10.1016/j.cell.2019.10.039
- 1130 22. Rheinbay E, Nielsen MM, Abascal F, et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature.* 2020. doi:10.1038/s41586-020-1965-x
- 1132 23. Li Y, Roberts ND, Wala JA, et al. Patterns of somatic structural variation in human cancer genomes. *Nature.* 2020;578(7793):112-121. doi:10.1038/s41586-019-1913-9
- 1134 24. Kim J, Mouw KW, Polak P, et al. Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. *Nat Genet.* 2016;48:600. http://dx.doi.org/10.1038/ng.3557.
- 1136 25. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. *Nature.* 2020;578(7793):94-101. doi:10.1038/s41586-020-1943-3
- 1138 26. Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature.* 2016;534:47. https://doi.org/10.1038/nature17676.
- 1140 27. Hillman RT, Chisholm GB, Lu KH, Futreal PA. Genomic Rearrangement Signatures and Clinical Outcomes in High-Grade Serous Ovarian Cancer. *JNCI J Natl Cancer Inst.* 2018;110(3):265-272. doi:10.1093/jnci/djx176
- 1142 28. Zhang AW, McPherson A, Milne K, et al. Interfaces of Malignant and Immunologic Clonal Dynamics in Ovarian Cancer. *Cell.* 2018;173(7):1755-1769.e22. doi:https://doi.org/10.1016/j.cell.2018.03.073
- 1144 29. Angus L, Smid M, Wilting SM, et al. The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. *Nat Genet.* 2019;51(10):1450-1458. doi:10.1038/s41588-019-0507-7
- 1146 30. Morganello S, Alexandrov LB, Glodzik D, et al. The topography of mutational processes in breast cancer genomes. *Nat Commun.* 2016;7(1):11383. doi:10.1038/ncomms11383
- 1148 31. Bayard Q, Meunier L, Peneau C, et al. Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. *Nat Commun.* 2018;9(1):5235. doi:10.1038/s41467-018-07552-9
- 1150 32. Puget S, Beccaria K, Blauwblomme T, et al. Biopsy in a series of 130 pediatric diffuse intrinsic Pontine gliomas. *Child's Nerv Syst.* 2015;31(10):1773-1780. doi:10.1007/s00381-015-2832-1
- 1154 33. Roujeau T, Machado G, Garnett MR, et al. Stereotactic biopsy of diffuse pontine lesions in children. *J Neurosurg Pediatr PED.* 107(1):1-4. doi:10.3171/PED-07/07/001
- 1156 34. Cage TA, Samagh SP, Mueller S, et al. Feasibility, safety, and indications for surgical biopsy of intrinsic brainstem tumors in children. *Child's Nerv Syst.* 2013;29(8):1313-1319. doi:10.1007/s00381-013-2101-0
- 1158 35. Gupta N, Goumnerova LC, Manley P, et al. Prospective feasibility and safety assessment of surgical biopsy for patients with newly diagnosed diffuse intrinsic pontine glioma. *Neuro Oncol.* 2018;20(11):1547-1555. doi:10.1093/neuonc/noy070
- 1160 36. Wala JA, Bandopadhyay P, Greenwald NF, et al. SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res.* 2018;28(4):581-591. doi:10.1101/gr.221028.117
- 1162 37. Wala JA, Shapira O, Li Y, et al. Selective and mechanistic sources of recurrent rearrangements across the cancer genome. *bioRxiv.* January 2017. http://biorxiv.org/content/early/2017/09/14/187609.abstract.
- 1164 38. Bahr C, von Paleske L, Uslu V V, et al. A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature.* 2018;553:515. https://doi.org/10.1038/nature25193.
- 1166 39. Fulco CP, Munschauer M, Anyoha R, et al. Systematic mapping of functional enhancer–promoter connections with CRISPR interference. *Science (80-).* 2016;354(6313):769 LP - 773. doi:10.1126/science.aag2445
- 1170 40. Zimmerman MW (Partners HD a S of H-RPN and IA through MIEH and FEA pd., Liu Y, He S, et al. MYC Drives a Subset of High-Risk Pediatric Neuroblastomas and Is Activated through Mechanisms Including Enhancer Hijacking and Focal Enhancer Amplification. *Cancer Discov.* 2018;8(3):320 LP - 335. doi:10.1158/2159-8290.CD-17-0993
- 1172 41. Zhang X, Choi PS, Francis JM, et al. Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat Genet.* 2015;48(2):176-182. doi:10.1038/ng.3470
- 1174 42. Krug B, De Jay N, Harutyunyan AS, et al. Pervasive H3K27 Acetylation Leads to ERV Expression and a Therapeutic Vulnerability in H3K27M Gliomas. *Cancer Cell.* 2019;35(5):782-797.e8.

- 1178 doi:10.1016/j.ccell.2019.04.004
- 1180 43. Suvà ML, Rheinbay E, Gillespie SM, et al. Reconstructing and Reprogramming the Tumor-Propagating
Potential of Glioblastoma Stem-like Cells. *Cell*. 2014;157(3):580-594.
doi:https://doi.org/10.1016/j.ccell.2014.02.030
- 1182 44. Nagaraja S, Quezada MA, Gillespie SM, et al. Histone Variant and Cell Context Determine H3K27M
Reprogramming of the Enhancer Landscape and Oncogenic State. *Mol Cell*. 2019;76(6):965-980.e12.
doi:https://doi.org/10.1016/j.molcel.2019.08.030
- 1184 45. Dunham I, Kundaje A, Aldred SF, et al. An integrated encyclopedia of DNA elements in the human
1186 genome. *Nature*. 2012;489(7414):57-74. doi:10.1038/nature11247
- 1188 46. Davis CA, Hitz BC, Sloan CA, et al. The Encyclopedia of DNA elements (ENCODE): data portal update.
Nucleic Acids Res. 2017;46(D1):D794-D801. doi:10.1093/nar/gkx1081
- 1190 47. Lee SB, Frattini V, Bansal M, et al. An ID2-dependent mechanism for VHL inactivation in cancer. *Nature*.
2016;529(7585):172-177. doi:10.1038/nature16475
- 1192 48. Paolella BR, Havrda MC, Mantani A, Wray CM, Zhang Z, Israel MA. p53 Directly Represses Id2 to Inhibit
the Proliferation of Neural Progenitor Cells. *Stem Cells*. 2011;29(7):1090-1101. doi:10.1002/stem.660
- 1194 49. Frankell AM, Jammula S, Li X, et al. The landscape of selection in 551 esophageal adenocarcinomas
defines genomic biomarkers for the clinic. *Nat Genet*. 2019;51(3):506-516. doi:10.1038/s41588-018-
0331-5
- 1196 50. Verhaak RGW, Bafna V, Mischel PS. Extrachromosomal oncogene amplification in tumour pathogenesis
and evolution. *Nat Rev Cancer*. 2019;19(5):283-288. doi:10.1038/s41568-019-0128-6
- 1198 51. Pathania M, De Jay N, Maestro N, et al. H3.3^{K27M} Cooperates with Trp53 Loss and PDGFRA
Gain in Mouse Embryonic Neural Progenitor Cells to Induce Invasive High-Grade Gliomas. *Cancer Cell*.
1200 2017;32(5):684-700.e9. doi:10.1016/j.ccell.2017.09.014
- 1202 52. Flavahan WA, Drier Y, Liao BB, et al. Insulator dysfunction and oncogene activation in IDH mutant
gliomas. *Nature*. 2016;529(7584):110-114. doi:10.1038/nature16490
- 1204 53. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census:
describing genetic dysfunction across all human cancers. *Nat Rev Cancer*. 2018;18(11):696-705.
doi:10.1038/s41568-018-0060-1
- 1206 54. Menghi F, Barthel FP, Yadav V, et al. The Tandem Duplicator Phenotype Is a Prevalent Genome-Wide
Cancer Configuration Driven by Distinct Gene Mutations. *Cancer Cell*. 2018;34(2):197-210.e5.
1208 doi:https://doi.org/10.1016/j.ccell.2018.06.008
- 1210 55. Hoopes JI, Cortez LM, Mertz TM, Malc EP, Mieczkowski PA, Roberts SA. APOBEC3A and APOBEC3B
Preferentially Deaminate the Lagging Strand Template during DNA Replication. *Cell Rep*.
2016;14(6):1273-1282. doi:10.1016/j.celrep.2016.01.021
- 1212 56. Haradhvala NJ, Polak P, Stojanov P, et al. Mutational Strand Asymmetries in Cancer Genomes Reveal
Mechanisms of DNA Damage and Repair. *Cell*. 2016;164(3):538-549.
1214 doi:https://doi.org/10.1016/j.ccell.2015.12.050
- 1216 57. Hoffman LM, Veldhuijzen van Zanten SEM, Colditz N, et al. Clinical, Radiologic, Pathologic, and Molecular
Characteristics of Long-Term Survivors of Diffuse Intrinsic Pontine Glioma (DIPG): A Collaborative Report
From the International and European Society for Pediatric Oncology DIPG Registries. *J Clin Oncol*.
1218 2018;36(19):1963-1972. doi:10.1200/JCO.2017.75.9308
- 1220 58. Zack TI, Schumacher SE, Carter SL, et al. Pan-cancer patterns of somatic copy number alteration. *Nat*
Genet. 2013;45:1134. http://dx.doi.org/10.1038/ng.2760.
- 1222 59. Taylor AM, Shih J, Ha G, et al. Genomic and Functional Approaches to Understanding Cancer Aneuploidy.
Cancer Cell. 2018;33(4):676-689.e3. doi:https://doi.org/10.1016/j.ccell.2018.03.007
- 1224 60. LETT JT, CALDWELL I, DEAN CJ, ALEXANDER P. Rejoining of X-ray Induced Breaks in the DNA of Leukaemia
Cells. *Nature*. 1967;214(5090):790-792. doi:10.1038/214790a0
- 1226 61. Gerstung M, Jolly C, Leshchiner I, et al. The evolutionary history of 2,658 cancers. *Nature*.
2020;578(7793):122-128. doi:10.1038/s41586-019-1907-7
- 1228 62. Larson JD, Kasper LH, Paugh BS, et al. Histone H3.3 K27M Accelerates Spontaneous Brainstem Glioma and
Drives Restricted Changes in Bivalent Gene Expression. *Cancer Cell*. 2019;35(1):140-155.e7.
doi:https://doi.org/10.1016/j.ccell.2018.11.015

- 1230 63. Cordero FJ, Huang Z, Grenier C, et al. Histone H3.K27M Represses &emph; to Accelerate Gliomagenesis in a Murine Model of DIPG. *Mol Cancer Res.* 2017;15(9):1243 LP - 1254.
1232 doi:10.1158/1541-7786.MCR-16-0389
- 1234 64. Funato K, Major T, Lewis PW, Allis CD, Tabar V. Use of human embryonic stem cells to model pediatric gliomas with H3.K27M histone mutation. *Science (80-)*. 2014;346(6216):1529 LP - 1533.
doi:10.1126/science.1253799
- 1236 65. Hoffman LM, DeWire M, Ryall S, et al. Spatial genomic heterogeneity in diffuse intrinsic pontine and midline high-grade glioma: implications for diagnostic biopsy and targeted therapeutics. *Acta Neuropathol Commun.* 2016;4(1):1. doi:10.1186/s40478-015-0269-0
1238
- 1240 66. Nikbakht H, Panditharatna E, Mikael LG, et al. Spatial and temporal homogeneity of driver mutations in diffuse intrinsic pontine glioma. *Nat Commun.* 2016;7(1):11185. doi:10.1038/ncomms11185
- 1242 67. Salloum R, McConechy MK, Mikael LG, et al. Characterizing temporal genomic heterogeneity in pediatric high-grade gliomas. *Acta Neuropathol Commun.* 2017;5(1):78. doi:10.1186/s40478-017-0479-8
- 1244 68. Vinci M, Burford A, Molinari V, et al. Functional diversity and cooperativity between subclonal populations of pediatric glioblastoma and diffuse intrinsic pontine glioma cells. *Nat Med.* 2018;24(8):1204-1215. doi:10.1038/s41591-018-0086-7
- 1246 69. Castel D, Philippe C, Calmon R, et al. Histone H3F3A and HIST1H3B K27M mutations define two subgroups of diffuse intrinsic pontine gliomas with different prognosis and phenotypes. *Acta Neuropathol.* 2015;130(6):815-827. doi:10.1007/s00401-015-1478-0
1248
- 1250 70. Khuong-Quang D-A, Buczkowicz P, Rakopoulos P, et al. K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas. *Acta Neuropathol.* 2012;124(3):439-447. doi:10.1007/s00401-012-0998-0
- 1252 71. Beroukhi R, Mermel C, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. *Nature.* 2010;463. doi:10.1038/nature08822
- 1254 72. Wu S, Turner KM, Nguyen N, et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature.* 2019;575(7784):699-703. doi:10.1038/s41586-019-1763-5
- 1256 73. Nathanson DA, Gini B, Mottahedeh J, et al. Targeted Therapy Resistance Mediated by Dynamic Regulation of Extrachromosomal Mutant EGFR DNA. *Science (80-)*. 2014;343(6166):72 LP - 76.
1258 doi:10.1126/science.1241328
- 1260 74. Miklja Z, Yadav VN, Cartaxo RT, et al. Everolimus improves the efficacy of dasatinib in PDGFR α -driven glioma. *J Clin Invest.* 2020;130(10):5313-5325. doi:10.1172/JCI133310
- 1262 75. Miklja Z, Pasternak A, Stallard S, et al. Molecular profiling and targeted therapy in pediatric gliomas: review and consensus recommendations. *Neuro Oncol.* 2019;21(8):968-980. doi:10.1093/neuonc/noz022
- 1264 76. Gröbner SN, Worst BC, Weischenfeldt J, et al. The landscape of genomic alterations across childhood cancers. *Nature.* 2018;555(7696):321-327. doi:10.1038/nature25480
- 1266 77. Zhang Y, Chen F, Donehower LA, Scheurer ME, Creighton CJ. A pediatric brain tumor atlas of genes deregulated by somatic genomic rearrangement. *Nat Commun.* 2021;12(1):937. doi:10.1038/s41467-021-21081-y
- 1268 78. Rausch T, Jones DTW, Zapatka M, et al. Genome Sequencing of Pediatric Medulloblastoma Links Catastrophic DNA Rearrangements with TP53 Mutations. *Cell.* 2012;148(1):59-71.
1270 doi:https://doi.org/10.1016/j.cell.2011.12.013
- 1272 79. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760. http://dx.doi.org/10.1093/bioinformatics/btp324.
- 1274 80. Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L. Calling Somatic SNVs and Indels with Mutect2. *bioRxiv.* January 2019:861054. doi:10.1101/861054
- 1276 81. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013;499:214. https://doi.org/10.1038/nature12213.
- 1278 82. Tabak B, Saksena G, Oh C, et al. The Tangent copy-number inference pipeline for cancer genome analyses. *bioRxiv.* January 2019:566505. doi:10.1101/566505
- 1280 83. Carter SL, Cibulskis K, Helman E, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol.* 2012;30:413. https://doi.org/10.1038/nbt.2203.
84. Viswanathan SR, Ha G, Hoff AM, et al. Structural Alterations Driving Castration-Resistant Prostate Cancer

- 1282 Revealed by Linked-Read Genome Sequencing. *Cell*. 2018;174(2):433-447.e19.
doi:<https://doi.org/10.1016/j.cell.2018.05.036>
- 1284 85. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and
confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome*
1286 *Biol*. 2011;12(4):R41. doi:10.1186/gb-2011-12-4-r41
86. Won H, de la Torre-Ubieta L, Stein JL, et al. Chromosome conformation elucidates regulatory
1288 relationships in developing human brain. *Nature*. 2016;538(7626):523-527. doi:10.1038/nature19847
87. Li H. Toward better understanding of artifacts in variant calling from high-coverage samples.
1290 *Bioinformatics*. 2014;30(20):2843-2851. doi:10.1093/bioinformatics/btu356
88. Imielinski M, Guo G, Meyerson M. Insertions and Deletions Target Lineage-Defining Genes in Human
1292 Cancers. *Cell*. 2017;168(3):460-472.e14. doi:<https://doi.org/10.1016/j.cell.2016.12.025>
89. Smith DI, Zhu Y, McAvoy S, Kuhn R. Common fragile sites, extremely large genes, neural development and
1294 cancer. *Cancer Lett*. 2006;232(1):48-57. doi:<https://doi.org/10.1016/j.canlet.2005.06.049>
90. Hadi K, Yao X, Behr JM, et al. Distinct Classes of Complex Structural Variation Uncovered across
1296 Thousands of Cancer Genome Graphs. *Cell*. 2020;183(1):197-210.e32. doi:10.1016/j.cell.2020.08.006
91. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to
1298 multiple testing. *J R Stat Soc B*. 1995;57.
92. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*.
1300 2012;29(1):15-21. doi:10.1093/bioinformatics/bts635
93. DeLuca DS, Levin JZ, Sivachenko A, et al. RNA-SeQC: RNA-seq metrics for quality control and process
1302 optimization. *Bioinformatics*. 2012;28(11):1530-1532. doi:10.1093/bioinformatics/bts196
94. Aguet F, Brown AA, Castel SE, et al. Genetic effects on gene expression across human tissues. *Nature*.
1304 2017;550(7675):204-213. doi:10.1038/nature24277
95. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with
1306 DESeq2. *Genome Biol*. 2014;15(12):550. doi:10.1186/s13059-014-0550-8
96. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical
1308 Bayes methods. *Biostatistics*. 2006;8(1):118-127. doi:10.1093/biostatistics/kxj037
97. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and
1310 other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882-883.
doi:10.1093/bioinformatics/bts034
- 1312 98. Zhang Y, Liu T, Meyer CA, et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.
doi:10.1186/gb-2008-9-9-r137
- 1314 99. Rao SSP, Huntley MH, Durand NC, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals
Principles of Chromatin Looping. *Cell*. 2014;159(7):1665-1680. doi:10.1016/j.cell.2014.11.021
- 1316 100. Rajarajan P, Borrmann T, Liao W, et al. Neuron-specific signatures in the chromosomal connectome
associated with schizophrenia risk. *Science (80-)*. 2018;362(6420). doi:10.1126/science.aat4311
- 1318 101. Durand NC, Robinson JT, Shamim MS, et al. Juicebox Provides a Visualization System for Hi-C Contact
Maps with Unlimited Zoom. *Cell Syst*. 2016;3(1):99-101. doi:10.1016/j.cels.2015.07.012
- 1320 102. Cameron CJF, Dostie J, Blanchette M. HIFI: estimating DNA-DNA interaction frequency from Hi-C data at
restriction-fragment resolution. *Genome Biol*. 2020;21(1):11. doi:10.1186/s13059-019-1913-y
- 1322 103. Grasso CS, Tang Y, Truffaux N, et al. Functionally defined therapeutic targets in diffuse intrinsic pontine
glioma. *Nat Med*. 2015;21(6):555-559. doi:10.1038/nm.3855
- 1324 104. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and
item tracking. *Bioinformatics*. 2010;26(12):1572-1573. doi:10.1093/bioinformatics/btq170
- 1326 105. Ramos AH, Lichtenstein L, Gupta M, et al. Oncotator: Cancer Variant Annotation Tool. *Hum Mutat*.
2015;36(4):E2423-E2429. doi:10.1002/humu.22771
- 1328 106. Cerami E, Gao J, Dogrusoz U, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring
Multidimensional Cancer Genomics Data. *Cancer Discov*. 2012;2(5):401 LP - 404. doi:10.1158/2159-
1330 8290.CD-12-0095
107. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles
1332 Using the cBioPortal. *Sci Signal*. 2013;6(269):pl1 LP-pl1. doi:10.1126/scisignal.2004088
108. Amin SB, Anderson KJ, Boudreau CE, et al. Comparative Molecular Life History of Spontaneous Canine

- 1334 and Human Gliomas. *Cancer Cell*. 2020;37(2):243-257.e7. doi:<https://doi.org/10.1016/j.ccell.2020.01.004>
- 1336 109. Shinde J, Bayard Q, Imbeaud S, et al. Palimpsest: an R package for studying mutational and structural
variant signatures along clonal evolution in cancer. *Bioinformatics*. 2018;34(19):3380-3381.
doi:[10.1093/bioinformatics/bty388](https://doi.org/10.1093/bioinformatics/bty388)
- 1338 110. Caravagna G, Heide T, Williams MJ, et al. Subclonal reconstruction of tumors by using machine learning
and population genetics. *Nat Genet*. 2020;52(9):898-907. doi:[10.1038/s41588-020-0675-5](https://doi.org/10.1038/s41588-020-0675-5)
- 1340 111. Storey JD. False Discovery Rate BT - International Encyclopedia of Statistical Science. In: Lovric M, ed.
Berlin, Heidelberg: Springer Berlin Heidelberg; 2011:504-508. doi:[10.1007/978-3-642-04898-2_248](https://doi.org/10.1007/978-3-642-04898-2_248)

1342

Figures

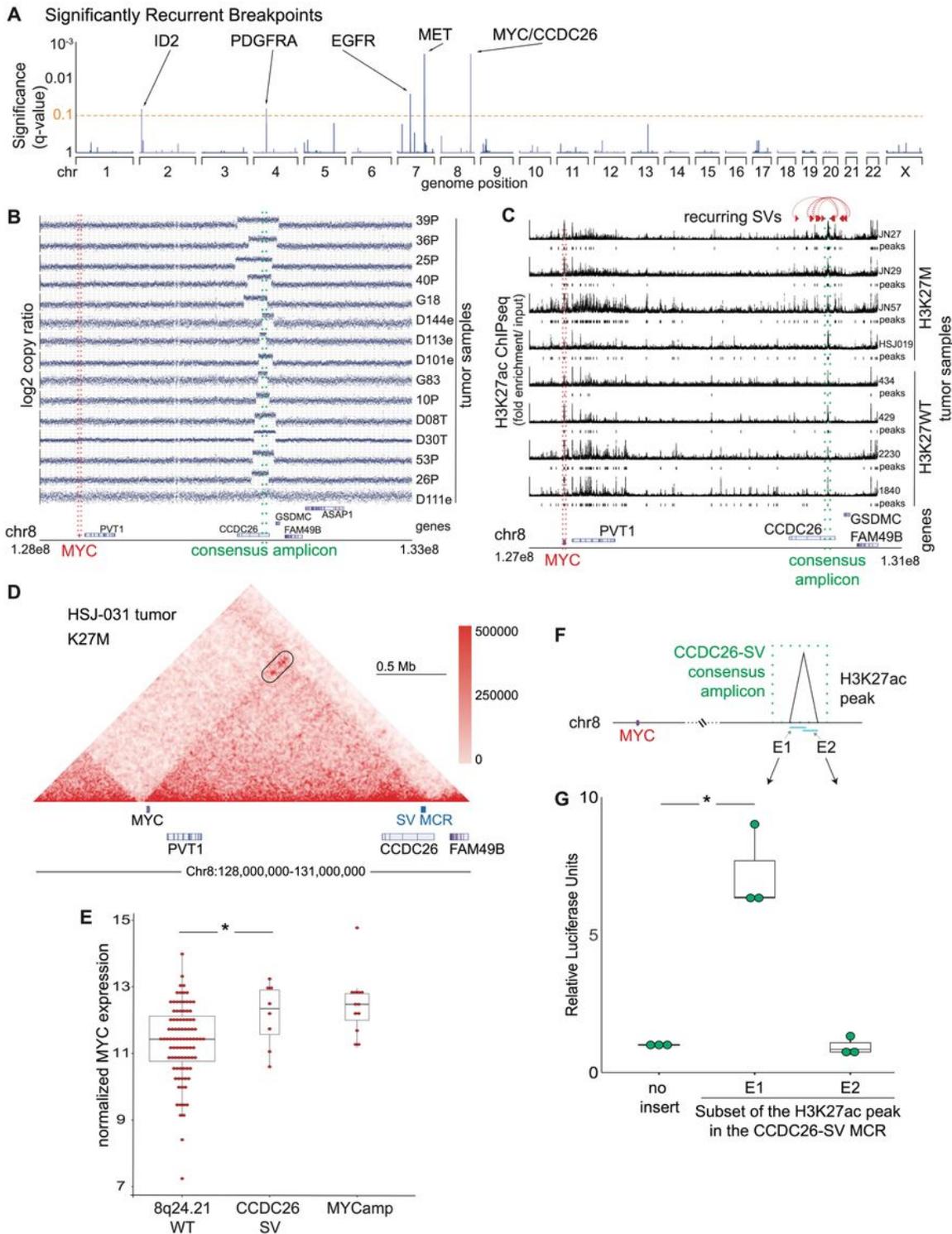


Figure 1

Significantly recurrent breakpoints within CCDC26. (A) Significance (multiple hypothesis corrected q-values, vertical axis) of recurrent breakpoints (genomic positions on the horizontal axis) across the 179 pHGG genomes. (B) Copy-number profiles across the MYC TAD for the 15 tumors with the recurrent

CCDC26 SV. (C) H3K27ac ChIP seq tracks within the TAD containing MYC (red lines) and CCDC26, for four H3K27M and four H3K27WT pHGGs tumors. Significantly enriched peaks (q -value < 0.01) are indicated below each H3K27ac ChIP-seq track. The CCDC26 amplicon boundaries for individual samples are indicated by the paired red arrows at the top. The consensus amplicon is indicated by the green dotted lines, and centers on an H3K27ac peak. (D) Hi-C heatmap across the MYC-CCDC26 locus from a midline glioma. Increasing interaction frequencies are indicated by brighter shades of red. The black oval outlines the frequent connections between MYC and the H3K27ac-enriched region within CCDC26. The minimal common region of amplification (SV MCR) is indicated at the bottom. (E) Normalized MYC expression in DMG samples with wild-type copy-number profiles at 8q24.21 ($n=92$), CCDC26-SVs ($n=8$) or amplifications of the MYC coding sequence ($n=12$). *denotes a $p<0.05$ as determined by Wilcoxon rank sum test. (F) Schematic illustrating the luciferase reporter used to validate the enhancer in CCDC26, showing the positions of the E1 and E2 sequences with respect to the enhancer within CCDC26. (G) Normalized luciferase activity in DIPG13 cells following transduction of the E1 and E2 enhancer reporters or empty vector controls. Values represent the average of four technical replicates in each of three independent experiments. * denotes $p=0.019$, T-test (unpaired, two-sided).

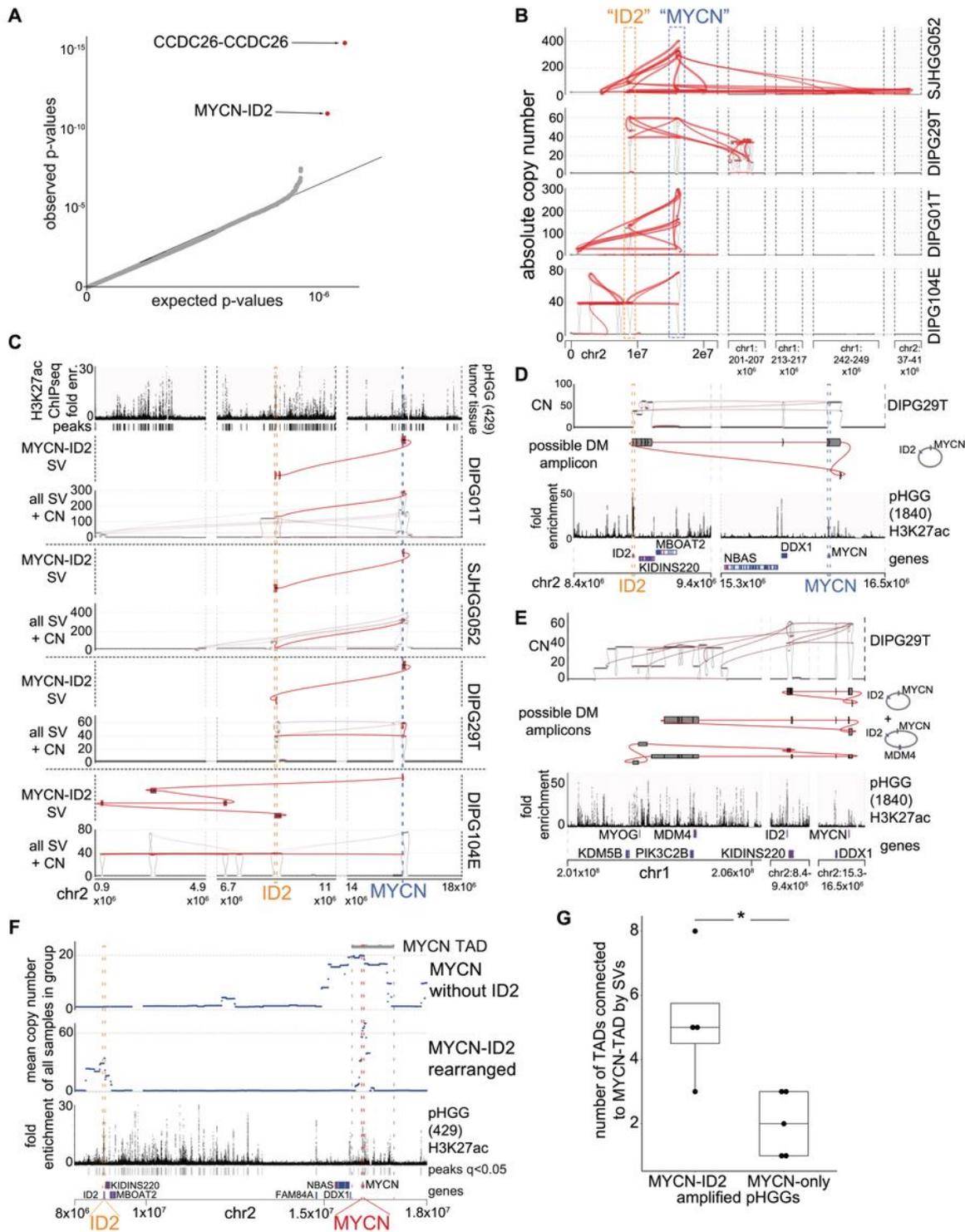


Figure 2

Significantly recurrent juxtapositions: MYCN-ID2. (A) Quantile-Quantile plot indicating the significance of juxtapositions between pairs of genomic loci. SVs that reached statistical significance are depicted in red. (B) SVs (red lines) involving MYCN and ID2 in samples with MYCN-ID2 rearrangements, and the number of copies at each connected locus (vertical axis). The dashed boxes indicate loci encompassing ID2 and MYCN. (C) SV maps as in panel B, with SVs juxtaposing ID2 and MYCN highlighted in red. The top track

indicates H3K27ac marks in a pHGG without a known MYCN-ID2 rearrangement, showing a strong enhancer within ID2. In each case, the MYCN-ID2 juxtaposition reduces the somatic distance MYCN and ID2 (each indicated by a dashed line) from 7 Mbp to less than 700kbp. (D-E) Examples of possible reconstructions of circular extrachromosomal amplicons containing MYCN and ID2 from a single DMG. (D) The simplest possible circle connecting MYCN and ID2. The top track shows copy-number and SVs, the middle track indicates the reconstructed topology, and the bottom track shows H3K27ac binding at the indicated loci in a different pHGG. (E) The chr1 loci connected to the MYCN-ID complex in this DMG. Short-read reconstructions allow for several extrachromosomal amplicons incorporating MYCN-ID2 and MDM4. The difference in the copy number could be explained either by a mix of amplicons containing respectively MYCN-ID2 alone and MYCN-ID2-MDM4 or by more complex amplicons incorporating multiple copies of MYCN878 ID2 for each copy of MDM4. (E) MYCN amplicons in tumors without ID2 amplification (top track) incorporate a larger fraction of the MYCN TAD (60%) relative to MYCN amplicons in tumors with MYCN-ID2-SV (second track; 23%, $p=0.03$, T-test). As a result, the former tend to include more loci with H3K27ac enrichment (third track) from the MYCN-TAD; significantly enriched H3K27ac peaks ($q\text{-value} < 0.01$) are indicated with black bars between the H3K27ac fold enrichment and the gene track. (F) MYCN-ID2 amplified pHGGs contain significantly more SVs between the MYCN-TAD and other TADs than pHGGs amplifying only MYCN and its neighborhood. ($p = 0.03$, Wilcoxon rank sum test).

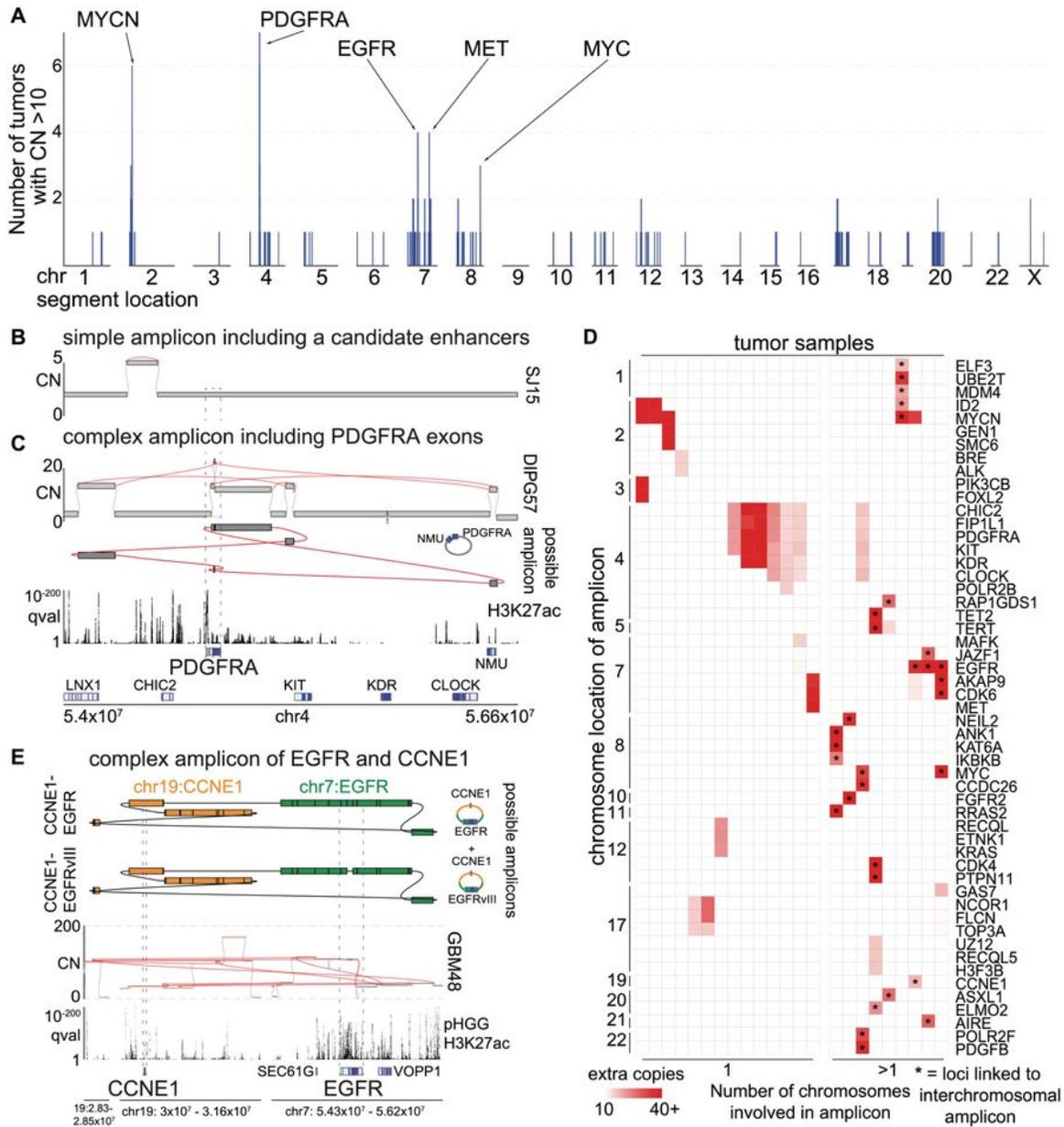


Figure 3

High-level amplicons. (A) For each genomic locus (horizontal axis), the number of tumors containing a high-level (CN>10) amplicon is indicated (vertical axis). SRBs are highlighted at the top. (B-C) Simple and complex-SVs exhibit distinct mechanisms to active PDGFRA. The top and bottom tracks indicate copy-numbers and the significance of H3K27ac enrichment (as calculated from eight pHGGs), respectively. SVs are highlighted in red. Selected gene loci are indicated on the bottom. (B) A simple amplicon of a

region with known 19,52 PDGFRA enhancers. (C) A complex high-level PDGFRA amplicon, displayed as in panel (B) with the addition of a track (second from top) indicating the topology of the amplicon. The complex-SV cluster around PDGFRA connects several segments on chr4, which are amplified to ten absolute copies. The SV calls support the reconstruction of an extrachromosomal amplicon incorporating PDGFRA exons and these segments. (D) Cancer genes involved in high-level amplicons (>10 copies) within the cohort. 9/23 tumors (grouped on right) contain high-level amplicons encompassing 898 loci from two or more chromosomes. These linked loci are marked by *. The color of each cell represents the number of extra copies due to the amplicon. (E) Example of a tumor with an extrachromosomal amplicon including two oncogenes from different chromosomes. This tumor shows a cluster of SVs connecting the EGFR and CCNE1 loci. The regions of both oncogenes are amplified to different CNs but in both cases reach several dozen absolute copies. (top) The complexity of the SVs allows for the reconstruction of several possible extrachromosomal amplicons. The CN differences in the bulk profile (middle) could be explained by either a mix of different circles or by more complex circles incorporating some segments repeatedly. The SV calls also reveal that a small fraction of the EGFR amplicons in this patient already show the EGFRvIII variant. The bottom two tracks show the genes of interest at the location and q-value H3K27ac track calculated from eight pHGG tumors.

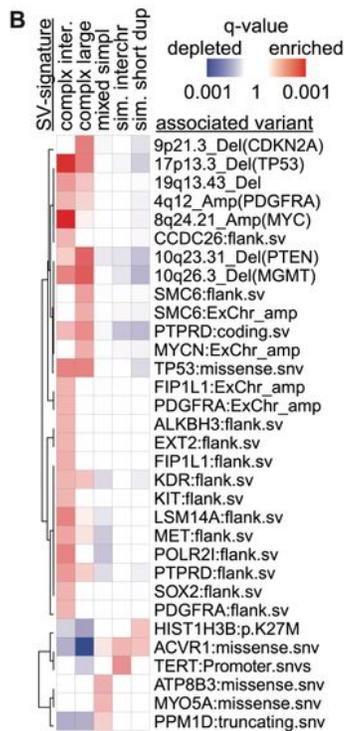
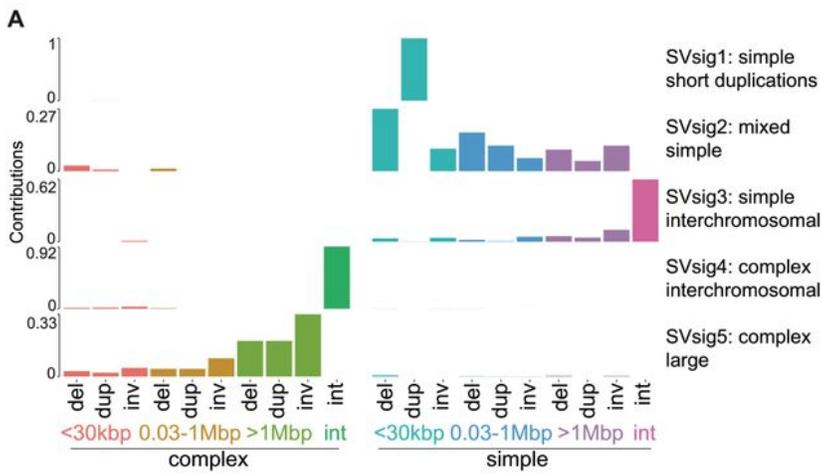


Figure 4

SV signatures in pHGG. (A) The horizontal axis indicates the size and type of SVs. Del stands for deletion, dup for duplication, inv for inversion, and int for interchromosomal rearrangement. The vertical axis indicates the fraction of SVs within each signature that are contributed by each SV type. (B) The statistical significance of positive (enriched) and negative (depleted) associations between each SV

signature and of all recurrently altered somatic genetic alterations that are documented in the Cancer Gene Census53.

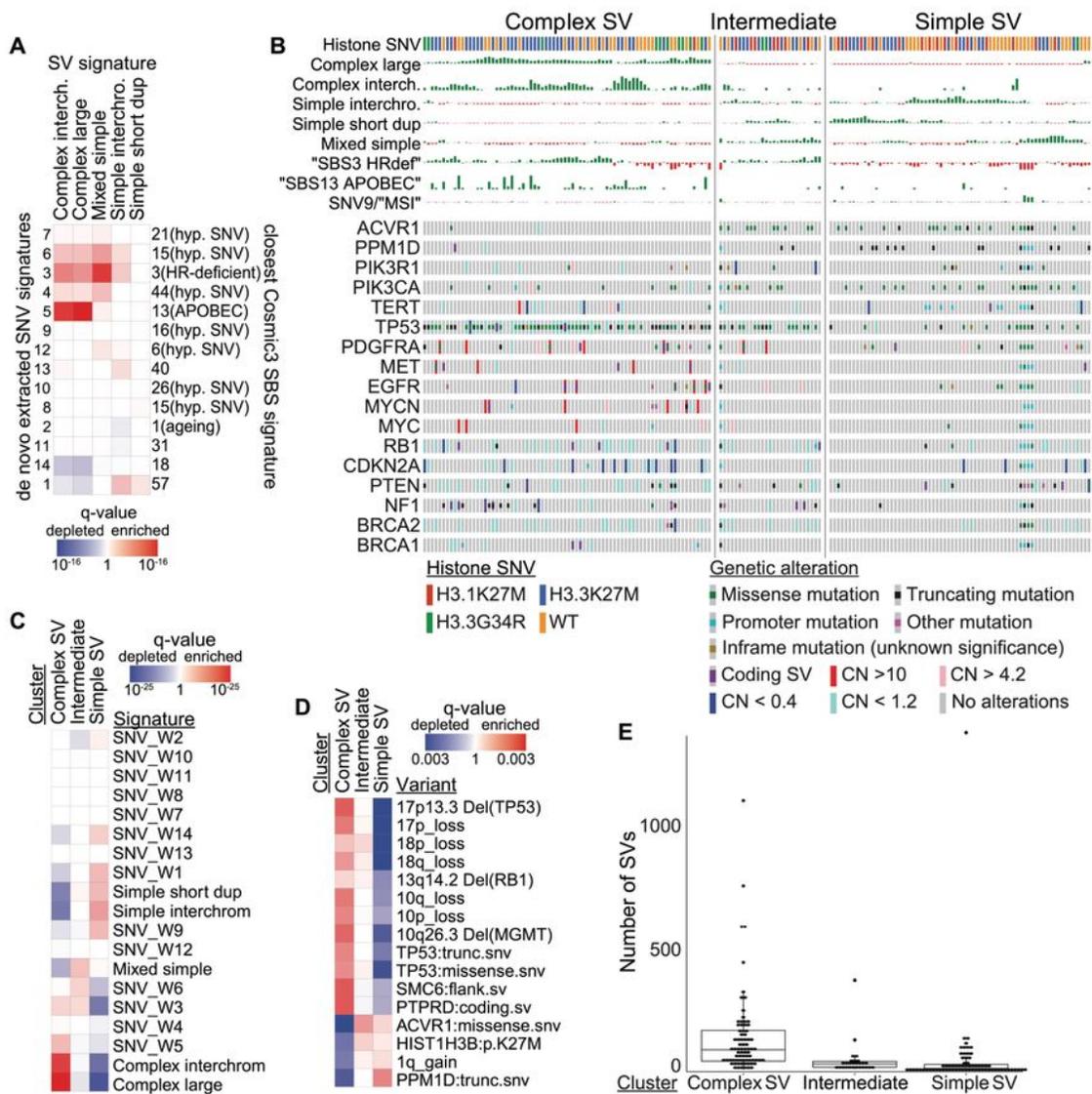


Figure 5

Integrated SV and SNV signature analyses reveal three pHGG subgroups with distinct variant generating processes. (A) Correlations between SV and SNV signatures. Signature labels from this analysis are indicated on the left; the nearest COSMICv3 signatures are indicated on the right, with their proposed

mechanisms in parentheses. Complex-SV signatures show a close correlation with APOBEC and homologous recombination deficiency SNV signatures (SBS3). q-values are based on Spearman rank correlations. (B) Consensus clustering of the normalized SNV and SV signature activities in each tumor sample (columns). Rows indicate signature activities (top) and potentially oncogenic variants (bottom). (C) Enrichment analysis for signature activities in each cluster from panel B. FDR q-values are based on Wilcoxon tests. (D) Significance of signature cluster associations for all variants with correlations reaching $q < 0.1$; q-values are based on Fisher's exact tests. Tumors in the complex-SV clusters are enriched for copy-number changes in cancer genes and SNVs in TP53, whereas simple-SV pHGGs tend to exhibit SNVs in different cancer genes. (E) Number of SVs per tumor in each cluster. All differences are significant to $q < 0.003$ by Wilcoxon tests.

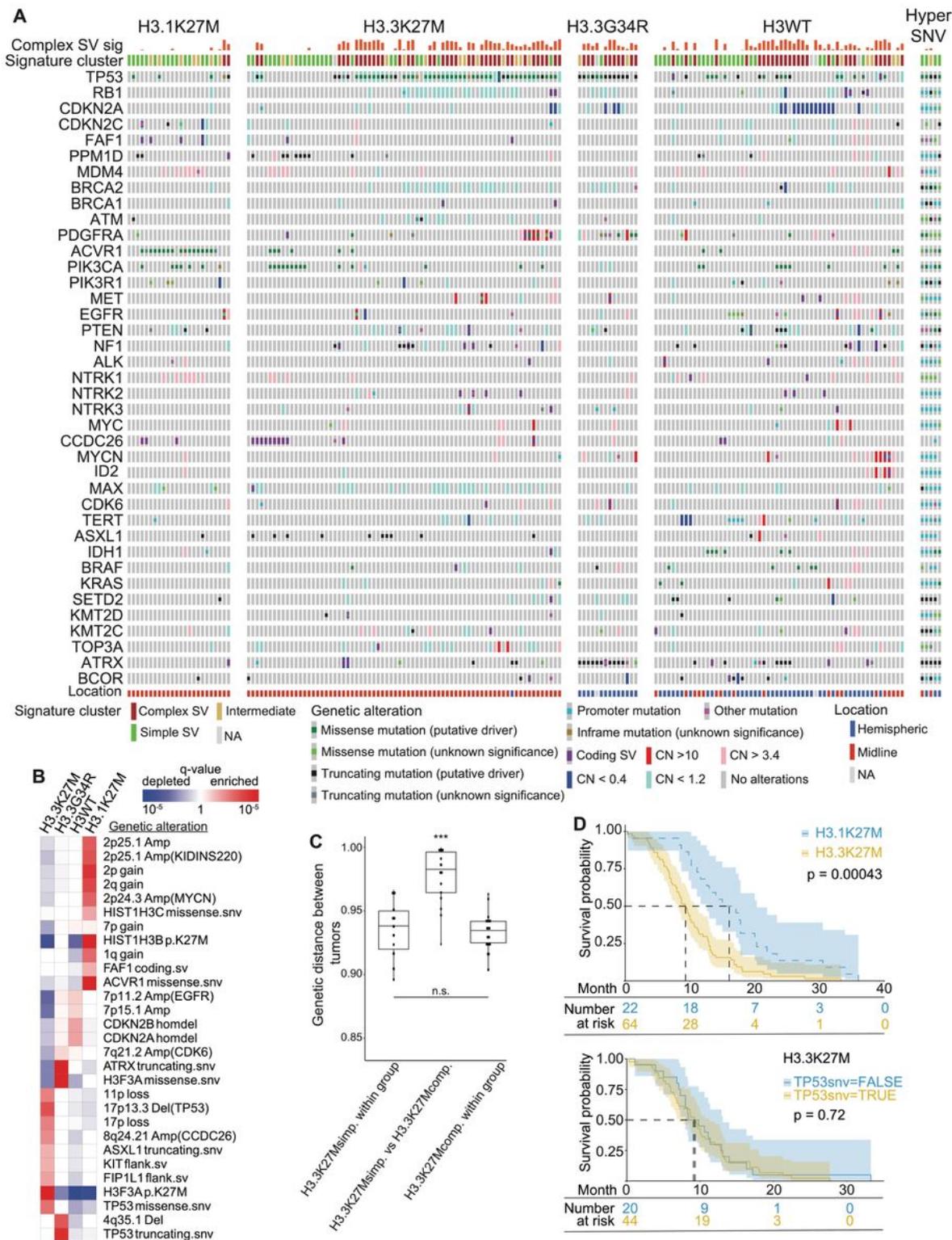


Figure 6 Overview of somatic variants and associated features within histone mutation-defined subgroups. (A) Co-mut plot of the 176/179 (98.3%) tumors with somatic variants in at least one well-known oncogene. Columns represent tumors, ordered within histone mutation-defined subgroups by hierarchical clustering of all potential driver variants. The top two rows show signature metadata. (B) Significance of histone subgroup associations for all variants with correlations reaching $q < 0.05$ (based on Fisher's exact tests)

Figure 7

Context of the significantly recurrent SVs. (SRSVs) (A) Co-mut plot for TAD-amp and Onc-amp SRSVs. TAD-amp SRSVs generate focal amplifications in the TAD of an oncogene without amplifying the protein coding sequence; Onc-amp SRSVs generate high-level (CN >3.4) amplifications or fusions of the coding sequence. The top three rows show associated metadata. The next seven rows indicate the genes affected by the SRSVs. The bottom seven rows show genes in DNA damage response pathways. Significant associations with the two groups are illustrated with pie charts below the plot, based on Fisher's exact tests. (B) TAD-amp pHGGs show significantly lower combined Complex-SV signature activity than Onc-amp pHGGs ($p = 4 \times 10^{-7}$, Wilcoxon). (C-D) Timing analysis of somatic variant acquisition in TAD-amp (C) and Onc-amp (D) pHGGs based on a Bradley-Terry model. The horizontal axis shows the log odds of the variant being an early event. The distributions indicate the results of 100 random subsamples of the data. Only variants in the SRSV-affected genes and pathways (growth factor and MYC signaling) and in DNA damage response genes altered in more than two samples are shown.

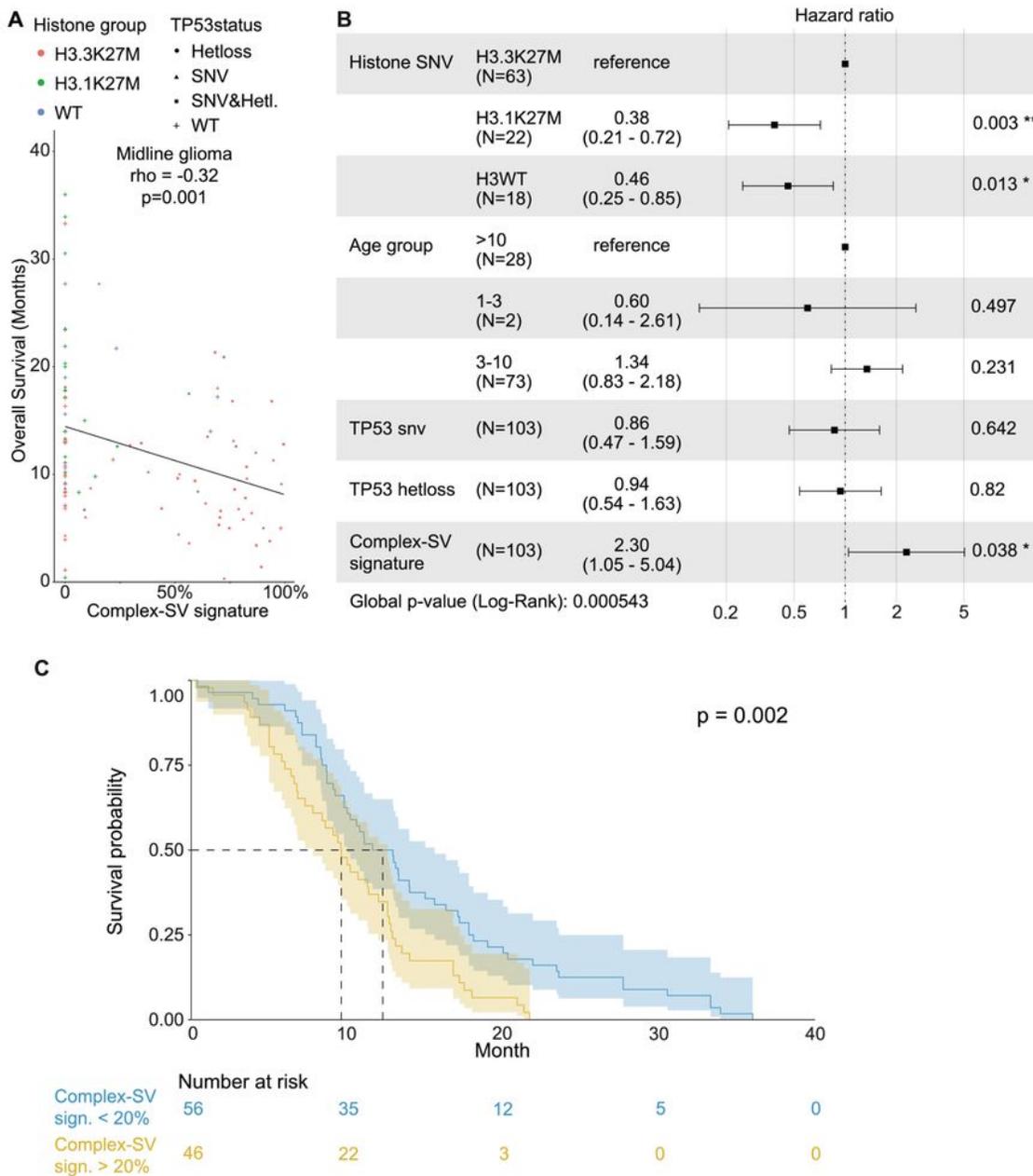


Figure 8

Association between the complex-SV signature and overall survival in midline gliomas. (A) Associations between the fraction of SV signature activity attributed to Complex-SV signatures and overall survival for all midline gliomas (n= 103; rho = -0.32 and p=0.001, Spearman correlation). (B) Cox proportional hazards analysis incorporating histone group, age, TP53 status and the combined Complex-SV signature. (C) Kaplan-Maier plot of a univariate analysis of the association between the Complex-SV signature and

overall survival, in which tumors were classified as Complex-SV if Complex-SV signatures contributed more than 20% of all SV signature activity. The number of patients in each group is indicated below the plot. The p-value represents a log-rank test.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supp.table1metaDatabySample.txt](#)
- [supp.table2SVmafs.txt](#)
- [supp.table3RSBs.txt](#)
- [supp.table4SVsigVarCorr.xlsx](#)
- [supp.table5VariantCombinationMatrix.txt](#)
- [supp.table6LuciferaseReporterPrimers.xlsx](#)
- [20210403pHGGsuppfigures.pdf](#)
- [20210403bExt.DataFigures.pdf](#)