Somatic structural variation signatures in pediatric brain tumors

Yang Yang[1], Lixing Yang[1,2,3,**]

1. Ben May Department for Cancer Research, University of Chicago, Chicago, IL, USA
2. Department of Human Genetics, University of Chicago, Chicago, IL, USA
3. University of Chicago Comprehensive Cancer Center, Chicago, IL, USA

Correspondence: lixingyang@uchicago.edu

Summary

Brain cancer is the leading cause of cancer-related death in children. Somatic structural variations (SVs), large scale alterations in DNA, remain poorly understood in pediatric brain tumors. Here, we detect a total of 13,199 high confidence somatic SVs in 744 whole-genome-sequenced pediatric brain tumors from Pediatric Brain Tumor Atlas. The somatic SV occurrences have tremendous diversity among the cohort and across different tumor types. We decompose mutational signatures of clustered complex SVs, non-clustered complex SVs, and simple SVs separately to infer the mutational mechanisms of SV formation. Our finding of many tumor types carrying unique sets of SV signatures suggests that distinct molecular mechanisms are active in different tumor types to shape genome instability. The patterns of somatic SV signatures in pediatric brain tumors are substantially different from those in adult cancers. The convergence of multiple signatures to alter several major cancer driver genes suggesting the functional importance of somatic SVs in disease progression.

Keywords

Genome instability, genomic rearrangement, chromothripsis, gene fusion

Introduction

Brain and central nervous system cancers are the most prevalent solid tumors in children under 19 and are the leading cause of cancer-related death among children (Ostrom et al., 2015). There are more than 100 types of pediatric brain tumors, which differ markedly from adult brain tumors (Louis et al., 2007). Although the 5-year survival rate of pediatric brain tumors is 75%, the survivors often suffer over their lifetimes from both the disease and the side effects of treatments. Therefore, there is an urgent need to better understand the disease mechanisms and to develop new therapeutic strategies in order to further increase survival and improve the quality of life for patients and their families.

Genetic alterations in cancer include single nucleotide variants (SNVs), copy number variants (CNVs), and structural variations (SVs). Pediatric brain tumors have few somatic SNVs, but carry more somatic SVs than other pediatric cancers (Gröbner et al., 2018). SVs are large scale structural changes of DNA such as deletions, tandem duplications, inversions, translocations, and some can be quite complex. For example, chromothripsis refers to a single catastrophic event resulting in numerous SVs through one cell cycle (Maciejowski et al., 2015; Stephens et al., 2011; Zhang et al., 2015). Understanding the mechanisms behind these alterations not only can improve our knowledge of disease biology, but also can reveal therapeutic opportunities. For instance, translocations at the immunoglobulin gene locus in B cell lymphoma are caused by aberrant V(D)J recombination (Alt et al., 2013; Hakim et al., 2012) and often result in activations of *MYC* and *BCL2* oncogenes (Bakhshi et al., 1985; Gostissa et al., 2009). Furthermore, breast and ovarian cancer patients carrying *BRCA1* and *BRCA2* mutations have deficiency in DNA double strand break repair and elevated level of somatic SVs in their tumors (Lee et al., 2016; Nik-Zainal et al., 2016). Patients with BRCA deficiency can be effectively treated by PARP inhibitors (Farmer et al., 2005; Fong et al., 2009). Mutational signatures have been widely used to study the molecular mechanisms of SNVs (Alexandrov et al., 2013, 2020; Degasperi et al., 2022), CNVs (Drews et al., 2022; Steele et al., 2022), SVs (Li et al., 2020) and complex SVs (Bao et al., 2022) in adult cancers. However, comprehensive studies of somatic SV signatures in pediatric brain tumors are still lacking. A recent study of SV signatures in pediatric high-grade glioma has revealed that histone genes, *TP53*, *CDKN2A*, and *RB1* aberrations are associated with complex SVs (Dubois et al., 2022). However, whether other types of pediatric brain tumors harbor similar SV signatures remains unclear.

Here, we decompose complex and simple SV signatures from 744 pediatric brain tumors. We find tremendous heterogeneity in SV occurrences and SV signatures across tumor types. We also report that somatic SVs are frequent cancer drivers in pediatric brain tumors.

Results and Discussion

High confidence somatic SVs in pediatric brain tumors

Pediatric Brain Tumor Atlas (PBTA) has collected more than 1,000 pediatric brain tumors across more than 30 types. There were 744 samples in PBTA with whole-genome sequencing data after removing non-tumorous lesions, non-brain cancer, and non-primary cancer samples (Table S1). We focused on tumor types with at least 10 samples including: 220 low-grade astrocytic tumors (LGATs), 97 medulloblastomas, 71 ependymomas, 70 high-grade gliomas (HGGs), 44 gangliogliomas, 38 craniopharyngiomas, 27 atypical teratoid rhabdoid tumors (ATRTs), 23 meningiomas, 23 dysembryoplastic neuroepithelial tumors (DNETs), 17 non-meningothelial mesenchymal tumors, 13 schwannomas, 13 germ cell tumors, 13 neurofibromas, and 12 choroid plexus papillomas. Tumor types with less than 10 samples were all classified as “Others”.

A previous study used Manta (Chen et al., 2016) to detect somatic SVs in the PBTA cohort to study the effects of SVs on gene expression (Zhang et al., 2021). However, the quality of variants called by a single algorithm is always not ideal (Campbell et al., 2020). In order to produce high confidence somatic SVs, we integrated three SV calling algorithms: Meerkat (Yang et al., 2013), Manta and Delly (Rausch et al., 2012a). Caller-specific SVs were discarded, and the somatic SVs detected by more than one algorithm were considered as high confidence. Meerkat, Manta and Delly detected 14,423, 55,934 and 9,475 somatic SVs in the 744 samples, respectively (Figure S1A). Since tumor DNA was not available, the SV quality could not be directly measured. Instead, we used CNV breakpoints detected by read depth approach to assess the quality of SVs, since a portion of somatic SVs change DNA copy numbers. We found SVs detected by only one algorithm were not well supported by CNVs (Figure S1B), which suggested that caller-specific SVs had poorer quality. SVs detected by Manta with fewer read pair and split read support were particularly of poor quality (Figure S1C). SVs detected by more than one algorithm were better supported by CNVs (Figure S1B), which suggested that they were of high quality. We also removed deletions that resided at exon-intron boundaries which were likely caused by cDNA contamination (Campbell et al., 2020). As a result, a total of 13,199 high confidence SVs were detected from 744 pediatric brain tumors with a median of 3 SVs per sample. In each type of pediatric brain tumor, the number of somatic SVs per sample varied by nearly three orders of magnitude (Figure 1A). There was also considerable heterogeneity across tumor types. HGGs were most abundant in somatic SVs followed by meningiomas and medulloblastomas, whereas choroid plexus papillomas had no SV at all (Figure 1A). Although most LGATs, ependymomas, gangliogliomas, and neurofibromas had very few SVs, a small fraction of them had very unstable genomes with more than 100 SVs (Figure 1A).

Complex SVs in pediatric brain tumors

A non-negative-matrix-factorization-(NMF)-based approach has been very effective in decomposing mutational signatures for somatic SNVs (Alexandrov et al., 2013, 2020; Degasperi et al., 2022) and CNVs (Drews et al., 2022; Steele et al., 2022). Several studies, including the recent SV signature study in pediatric high-grade glioma (Dubois et al., 2022), also used this

approach to extract SV signatures by combining both complex SVs and simple SVs (Li et al., 2020). Meaningful signatures can be reliably detected if DNA damage and repair mechanisms generate variants independently and repeatedly in cancer genomes. However, it is well-established that chromothripsis events occur as one-time events and the numbers of SVs vary extensively (Maciejowski et al., 2015; Shoshani et al., 2020; Zhang et al., 2015). Furthermore, multiple molecular mechanisms can lead to chromothripsis. For example, lagging chromosomes trapped in micronuclei during mitosis can shatter into many pieces and some fragments can be ligated together as chromothripsis events (Zhang et al., 2015). Dicentric chromosomes can form chromatin bridges during cell division, shatter into pieces, and also produce chromothripsis (Maciejowski et al., 2015). NMF-based SV signature decomposition cannot differentiate these mechanisms. To better deduce SV signatures in pediatric brain tumors, we studied clustered complex SVs, non-clustered complex SVs and simple SVs separately. Clustered complex SVs are those with breakpoints enriched in certain genomic regions including chromothripsis. Circular extrachromosomal DNA (ecDNA) with many SV breakpoints is also a clustered complex SV (Bao et al., 2022). We recently developed Starfish, a clustering-based approach, to infer clustered complex SV signatures based on their SV and CNV patterns (Bao et al., 2022). We reported six clustered complex SV signatures using nearly 2,500 adult tumors including micronuclei-induced chromothripsis, chromatin-bridge-induced chromothripsis, as well as ecDNA. There are three other signatures that cannot be linked to biological processes, namely “Large loss”, “Large gain” and “Hourglass”. Non-clustered complex SVs are complex SVs with scattered breakpoints including chromoplexy and cycle of templated insertions (Li et al., 2020). Chromoplexy events are likely to form through repair of multiple co-occurring DNA double strand breaks similar to reciprocal translocations (Baca et al., 2013; Berger et al., 2011), whereas templated insertions may reflect replication-based mechanisms (Li et al., 2020; Yang et al., 2013; Zhang et al., 2009). After detecting clustered and non-clustered complex SVs, the remainder of SVs were classified as simple SVs, which include deletions, tandem duplications, balanced/unbalanced/foldback inversions, and balanced/unbalanced translocations.

Among the 13,199 SVs in 744 pediatric brain tumors, 7,601 (57.6%) were clustered complex SVs that belonged to 146 individual complex events, 2,377 (18.0%) were non-clustered complex SVs that belonged to 346 events, and 3,221 (24.4%) were simple SVs (Table S2). Out of the 744 tumors, 108 (14.5%) and 150 (20.2%) carried clustered and non-clustered complex SVs respectively, whereas 552 (74.2%) did not have any complex SVs. The high numbers of SVs in tumors with very unstable genomes (>100 SVs) were mainly due to their abundance of complex SVs (Figure 1A and Figure S2). HGGs were most abundant in complex SVs, whereas DNETs did not carry any complex SVs (Figure 1A). We used Starfish (Bao et al., 2022) to classify clustered complex SV signatures (Table S3) and used junction pattern (Li et al., 2020) to determine non-clustered complex SVs. HGGs and medulloblastomas carried almost all complex SV signatures, whereas other tumor types only harbored small numbers of complex SV signatures (Figure 1B). The “ecDNA” (Figure 1C) signature was predominantly found in HGGs; the “Large loss” signature (Figure 1C), characterized by complex SVs with large amount of DNA loss, was mainly observed in HGGs and meningiomas; micronuclei-induced chromothripsis (“Micronuclei” signature) (Figure 1C) was enriched in mesenchymal tumors and ependymomas (Figure 1B). Chromatin-bridge-induced chromothripsis (“Chr bridge” signature)

(Figure 1C) only occurred in HGGs and neurofibromas (Figure 1B). Hourglass chromothripsis events ("Hourglass" signature), complex SVs with a small amount of DNA loss and highly concentrated SV breakpoints, were detected in a small number of samples in several tumor types (Figure 1B), such as HGGs, meningiomas, medulloblastomas, and ependymomas. Regarding to non-clustered complex SVs, chromoplexy was found in many different tumor types and occurred in as many as 18.6% (13/70) of HGGs and 17.6% (3/17) of mesenchymal tumors (Figure 1D). Cycle of templated insertions was abundant in HGGs (Figure 1D).

In summary, different types of pediatric brain tumors often carry distinct complex SV signatures.

Simple SVs in pediatric brain tumors

Next, we used the NMF approach to decompose simple SV signatures; a total of nine signatures were extracted (Figure 2A and Table S2). These signatures included deletions smaller than 1kb ("Del0"), deletions between 1kb to 5kb ("Del1"), deletions larger than 5kb ("Del2"), tandem duplications ("TD"), unbalanced inversions ("Unbal inv"), large intra-chromosomal SVs ("Large mixed"), reciprocal inversions and reciprocal translocations ("Recip"), as well as unbalanced translocations ("Unbal tra"). Interestingly, large tandem duplications resulting in *KIAA1549-BRAF* fusions belonged to a stand-alone signature, namely "*BRAF* fusion". HGGs had more simple SVs than other tumor types, while a considerable number of samples from various tumor types, including LGATs and ependymomas, showed no evidence of simple SVs (Figure 2B). Most tumors with simple SVs carried multiple simple SV signatures (Figure 2B), and apparent enrichments could be observed. For example, DNETs predominantly carried the "TD" signature; schwannomas mainly harbored the "Recip" signature; the "Unbal inv" signature was mainly found in HGGs; the "TD" signature was enriched in medulloblastomas; and "Del2" signature was abundant in ATRTs; and the "*BRAF* fusion" signature was almost exclusively observed in LGATs (Figure 2B). Intriguingly, 102 out of 220 LGATs had *KIAA1549-BRAF* fusions, and among the fusion positive LGATs, 88 had only one SV present in their genomes. The SVs producing *KIAA1549-BRAF* fusions were exclusively tandem duplications between 1Mb and 2.5Mb in size (Figure 2A).

Taken together, our complex SV and simple SV signature analysis demonstrated that numerous mutational mechanisms are active in pediatric brain tumors to induce genome instability, with unique molecular mechanisms present in different tumor types.

Genomic features associated with SV signatures

Somatic SVs are not evenly distributed across the genome (Li et al., 2020). Many factors, such as replication timing, GC content, repeat content, and 3D genome organization, have been associated with SV breakpoint distribution (Akdemir et al., 2020; Li et al., 2020). Here, we surveyed 31 genomic features for their relationships with somatic SVs in pediatric brain tumors (Figure 3).

For clustered complex SV signatures, SV breakpoints of the "ecDNA", "Chr bridge" and "Large gain" signatures were significantly enriched in late replicated regions (Figure 3). In contrast, all

clustered complex SV signatures in adult cancers were enriched in early replicated regions (Bao et al., 2022). In adult cancers, all clustered complex SV signatures were enriched in GC-rich regions and near CpG islands (Bao et al., 2022), whereas only the "ecDNA" and "Micronuclei" signatures in pediatric brain tumors were enriched in GC-rich regions (Figure 3). SV breakpoints of the "Large loss" signature were significantly closer to centromeres than expected in pediatric brain tumors (Figure 3), in a pattern opposite to adult cancers (Bao et al., 2022). In adult cancers, the "ecDNA" and "Chr bridge" signatures were significantly farther away from telomeres, whereas other clustered complex SVs were significantly closer to telomeres (Bao et al., 2022). However, all clustered complex SV signatures were significantly closer to telomeres in pediatric brain tumors (Figure 3). All clustered complex SV signatures in adult cancers were significantly closer to many types of repetitive elements (simple repeats, short tandem repeats and transposable elements) (Bao et al., 2022), whereas in pediatric brain tumors, the repetitive elements had variable effects. For example, the "Micronuclei" signature in pediatric brain tumors were enriched near Alu elements but depleted around L1s (Figure 3). Therefore, the SV breakpoint distributions of clustered complex SVs in pediatric brain tumors were quite different from those of adult cancers.

Chromoplexy was proposed to form similar to reciprocal translocations through simultaneous ligation of multiple broken chromosomal ends (Baca et al., 2013; Berger et al., 2011). In adult cancers, chromoplexy breakpoints and reciprocal translocations shared similar patterns. For example, they were both enriched in late replicated regions. In sharp contrast, chromoplexy breakpoints in pediatric brain tumors were enriched in early replicated regions while reciprocal translocations did not display any bias in replication timing (Figure 3). In addition, chromoplexy breakpoints in pediatric brain tumors were also enriched in GC-rich regions and near telomeres, while reciprocal translocations were enriched in AT-rich regions and not enriched towards either centromeres or telomeres (Figure 3). These results suggested that chromoplexy in pediatric brain tumors may form through a different mechanism than reciprocal translocations. Cycle of templated insertion breakpoints in pediatric brain tumors had little association with most genomic features except for proximity to telomeres (Figure 3).

Simple SV breakpoint distributions of pediatric brain tumors were also quite distinct from those of adult cancers. In adult cancers, deletions were enriched in early replicated regions, and tandem duplications and unbalanced translocations were enriched in late replicated regions (Li et al., 2020). In contrast, in pediatric brain tumors, smaller deletions ("Del0" and "Del1") were not associated with replication timing, large deletions ("Del2") were enriched in late replicated regions, and tandem duplications ("TD") as well as unbalanced translocations ("Unbal tra") were enriched in early replicated regions (Figure 3). In adult cancers, all deletions regardless of their sizes were enriched in AT-rich regions (Li et al., 2020); whereas in pediatric brain tumors, small deletions ("Del1") were enriched in GC-rich regions and large deletions ("Del2") were enriched in AT-rich regions (Figure 3). In adult cancers, tandem duplications were significantly closer to Alu elements (Li et al., 2020); whereas in pediatric brain tumors, tandem duplications ("TD") were farther away from L1s, LTR transposons, and DNA transposons (Figure 3). In adult cancers, both small and large deletions were depleted from topologically associated domain (TAD) boundaries (Li et al., 2020). However, in pediatric brain tumors, only large deletions

("Del2"), but not small deletions ("Del0" and "Del1"), were depleted from TAD boundaries (Figure 3).

Taking the above results together, our results suggest that the forming mechanisms of somatic SVs, both complex and simple SVs, in pediatric brain tumors were likely to be different from those of adult cancers.

Next, we sought to identify mutations associated with genome instability. As expected, *TP53* mutations in HGGs were associated with "Chr bridge" and "Large loss", and cycle of templated insertions, as well as six simple SV signatures: "Unbal inv", "Unbal tra", "Large mixed", "TD", "Del1" and "Del2" (Figure S3). In addition, *H3F3A* mutations in HGGs were associated with cycle of templated insertions (Figure S3) suggesting that histones may play important roles in genome instability in pediatric brain tumors. *ATRX* mutations in HGGs were associated with "Large mixed" signatures (Figure S3). No other mutations were associated with any SV signatures in any other pediatric brain tumors. In adult cancers, small deletions are associated with *BRCA2* mutations, and small and large tandem duplications are associated with *BRCA1* and *CDK12* mutations (Li et al., 2020; Nik-Zainal et al., 2016). In the 741 non-hypermutated pediatric brain tumors, only one sample carried a *BRCA1* mutation, and another sample carried a *BRCA2* mutation. This again suggested that the mechanisms of formation of deletions and tandem duplications in pediatric brain tumors are likely to be different from those of adult cancers.

SV breakpoint sequences

We then investigated microhomology and insertion sequences at the SV breakpoints across SV signatures. SVs are the results of erroneous repair of DNA double strand breaks or replication errors. Various repair pathways are involved (Aguilera and García-Muse, 2013; Bunting and Nussenzweig, 2013), such as non-homologous end joining (NHEJ), alternative end joining (alt-EJ), and microhomology-mediated break-induced repair (MMBIR). NHEJ usually ligates blunt DNA ends or ends with short 1-4bp homology. Alt-EJ often uses slightly longer homology for repair. MMBIR is considered a replication based template switching mechanism and non-template insertions are frequently present at the breakpoints (Zhang et al., 2009). Several SV signatures in pediatric brain tumors, including "Large loss", "Hourglass", chromoplexy, and "Unbal tra", had a majority of SV breakpoints being blunt DNA ends (no homology nor insertion at the breakpoints) (Figure 4). This suggested that these SVs are likely to form through NHEJ. Some other signatures, such as "Chr bridge", "Micronuclei", "Del0", "Del1", "Unbal inv", and "Large mixed", had slightly longer homology at the breakpoints with 1bp microhomology being the most frequent and also consistent with NHEJ (Figure 4). In addition, "TD", "Recip", and "*BRAF* fusion" signatures had 2bp microhomology being the most frequent (Figure 4) suggesting alt-EJ might play more important roles in these SVs. The observation of extended microhomology in the "Recip" signature was consistent with breakpoints found in the Philadelphia chromosome, the most prevalent reciprocal translocation in leukemia, often involving 2-8 bp microhomology (McVey and Lee, 2008). Although previous studies proposed

that chromoplexy forms in a manner similar to reciprocal translocation (Baca et al., 2013; Berger et al., 2011), our observation of chromoplexy and reciprocal translocation breakpoints with different microhomology patterns suggested that they may form via different mechanisms. Furthermore, a fraction of SVs in "Del1" had 5 bp or longer microhomology (Figure 4) especially in ependymomas, suggesting alt-EJ being the dominant mechanism in ependymoma. Intriguingly, "ecDNA" and "Large gain" signatures had frequent inserted sequences that were more than 10 bp (Figure 4), suggesting possible involvement of replication-based mechanisms such as MMBIR.

SV hotspots and tumor drivers

Hotspots of SV breakpoints often represent genetic alterations under positive selection and genes driving diseases. After binning the reference genome into 1 Mb window and counting SV occurrences, we found different SV signatures having quite distinct hotspots. *MYCN* and *MYC* are two frequently amplified oncogenes in pediatric brain tumors (Dubois et al., 2022). Interestingly, *MYCN* was amplified exclusively by "ecDNA" in HGGs and exclusively by "Large gain" in medulloblastomas (Figure 5), since "ecDNA" and "Large gain" were the most abundant clustered complex SV signatures in HGGs and medulloblastomas, respectively. Both signatures converged on *MYCN* amplifications, which are the main oncogenic events in HGGs and medulloblastomas. *MYCN* was also amplified by non-clustered complex SV with unclear pattern ("Complex unclear") in various tumor types (Figure 5). Tandem duplications can also amplify DNA. However, *MYCN* was not amplified by the "TD" signature in any samples, whereas *MYC* was amplified by the "TD" signature in a few HGGs and medulloblastomas (Figure 5). In addition, *FGFR1* was frequently amplified by the "TD" signature in DNETs (Figure 5). DNETs had very few somatic SVs and did not carry any clustered or non-clustered complex SVs. "TD" was the dominant SV signature in DNETs. These findings suggested that "TD" and *FGFR1* amplifications were the major oncogenic events in DNETs. Furthermore, chromoplexy and "Del2" frequently disrupted *CDKN2A* in various tumor types (Figure 5). The "*BRAF* fusion" signature produced *KIAA1549-BRAF* fusions mainly in LGATs (Figure 5). The "Recip" signature often led to *EWSR1* fusions in various tumor types (Figure 5).

Interestingly, multiple SV signatures harbored hotspots on chromosome 11, and the hotspots were only found in ependymomas (Figure 5). *C11orf95-RELA* fusions are the major oncogenic events in 70% of supratentorial ependymomas (Parker et al., 2014). Recently, the World Health Organization (WHO) recommended the use of the *ZFTA* (*C110rf95*) fusion to classify supratentorial ependymoma instead of the *RELA* fusion because *ZFTA* can fuse to other partners as well (Louis et al., 2021). There were 71 ependymomas in our cohort and 23 of them (32%) were supratentorial (Figure 6A and Table S4). Among them, 13 (57%) carried *ZFTA* fusions (Figure 6A). There were 2 other *ZFTA* fusion positive ependymomas classified as "Others" (Figure 6A). Twelve out of 15 *ZFTA* fusions were driven by complex SVs and three were driven by tandem duplications ("TD") (Figure 6A). Among the 12 fusions resulting from complex events, 5 were micronuclei-induced chromothripsis ("Micronuclei"), 3 were hourglass chromothripsis ("Hourglass"), and 4 were non-clustered complex SVs (Figure 6A and 6B). Some

of the complex SVs involved the entire chromosome 11 (BS_K6A9Z04J), whereas others only affected a small region in chromosome 11 (BS_NWYBD9CA) (Figure 6B). These results showed that there are diverse mechanisms generating genome instability in ependymomas and they have shared oncogenic consequence of forming *ZFTA* fusions. The most prevalent complex SVs in ependymomas were micronuclei-induced chromothripsis events caused by erroneous chromosomal segregation. It is possible that the frequent complex SVs in ependymomas are due to frequent chromosomal segregation errors. However, aneuploidy is rare in ependymomas (Mack and Taylor, 2017). This suggested that chromosomal segregation errors are likely not frequent events in ependymomas, and complex SVs in ependymomas are under positive selection. Furthermore, since tandem duplications are sufficient to produce gene fusions, such as *ZFTA* fusions and *BRAF* fusions, and the fact that most somatic SVs in ependymomas involving chromosome 11 were complex SVs suggested that other genes altered by SVs may contribute to ependymoma tumorigenesis as well. "Unbal inv" signature also had a hotspot in a similar region on chromosome 11 (Figure 5). Three ependymomas had unbalanced inversions in gene *MARK2*, which is next to *ZFTA*, and no fusions were formed. Whether this "Unbal inv" hotspot in ependymoma reflects oncogenic events remains unclear.

In a significant fraction of pediatric brain tumors, the disease-causing SVs were the sole SVs of the corresponding signatures. These signatures did not produce additional passenger SVs in those samples. For example, 88 LGATs were driven by *BRAF* fusions and there was only one SV within the "*BRAF* fusion" signature in those samples (Figure 2B). Similarly, three ependymomas carried *ZFTA* fusions caused by the "TD" signature, and there was only one SV within the "TD" signature in two of these three samples (Figure 2B and Figure 5). In addition, nine DNETs had *FGFR1* amplifications resulting from the "TD" signature, and there was only one SV within the "TD" signature in all nine samples (Figure 2B and Figure 5). These results indicated that the molecular mechanisms leading to these disease-driving SVs are not highly active in tumor initiating cells. Cells independently acquire SVs through these mechanisms at very low rates, and cells that have acquired SVs that alter the major disease-driving genes eventually outcompete other cells and become tumors.

Impact of SVs on patient survival

Next, we sought to evaluate whether somatic SVs affect patient survival in pediatric brain tumors. Chromothripsis has been associated with worse patient survival in several previous studies (Cortés-Ciriano et al., 2020; Molenaar et al., 2012; Notta et al., 2016; Rausch et al., 2012b). However, we did not observe any complex SV signatures associated with patient survival in HGGs, LGATs, medulloblastomas or ependymomas (Figure S4). HGG patients with "Del2" and "Unbal tra" signatures in their tumors had significantly worse survival, and those with "Del1" and "Large mixed" signatures had marginally worse survival (Figure 7). HGGs were most abundant in simple SVs, which suggested that the SV forming mechanisms are relatively more active in tumor initiating cells of HGGs than those of other tumor types. No simple SV signatures were associated with patient survival in other tumor types. It is possible that simple

346 SV forming mechanisms are not very active in tumor types other than HGGs; therefore simple
347 SV signatures are associated with patient survival only in HGGs.

348

Acknowledgement

We thank the Center for Research Informatics at the University of Chicago for providing the computing infrastructure. We thank Dr. Ji Won Lee for assistance in tumor pathology. The work was supported by the National Institutes of Health grant R03CA246228 (L.Y.).

Author contributions

L.Y. conceptualized the study. Y.Y. developed the analysis workflow and performed the analysis. Y.Y. and L.Y. interpreted the results. L.Y. supervised the study. L.Y. wrote the paper. All authors have read and approved the final manuscript.

Declaration of interests

The authors have no competing interests to declare.

Figure Legends

Figure 1 Somatic SVs and complex SVs in 744 pediatric brain tumors. (A) Frequencies of somatic SVs and percentages of different types of SVs. In the upper panel, each dot represents one pediatric brain tumor sample. Samples are grouped by tumor type and tumor types are sorted by median SV frequency (red lines) except for the "Others" category. The numbers in parentheses are sample sizes for the corresponding tumor types. The bottom panel shows the percentages of clustered complex SVs, non-clustered complex SVs, and simple SVs in the corresponding samples on the top panel. HGG: high-grade glioma. ATRT: atypical teratoid/rhabdoid tumor. LGAT: low-grade astrocytic tumor. DNET: dysembryoplastic neuroepithelial tumor. (B) and (D) Percentages of clustered complex SV signatures and non-clustered complex SVs. Each vertical block represents one tumor type and each horizontal bar represent one sample. Samples are colored based on their SV signatures. Samples carry multiple signatures have multiple colors arranged horizontally. The height of each sample may differ across tumor types depending on sample sizes of the tumor types. (C) Examples of clustered complex SVs. Colored arcs represent SVs of different types. The red bars below the SVs indicate regions of clustered complex SVs. Copy number profiles are displayed by black bars above the chromosome models. The red bars within the grey chromosome models indicate the locations of centromeres. Tumor types and sample IDs are shown next to the names of clustered complex SV signatures. DEL: deletion, DUP: tandem duplication, h2hINV: head to head inversion, t2tINV: tail to tail inversion, TRA: translocation.

Figure 2 Simple SV signatures and their distributions. (A) Nine simple SV signatures of 744 pediatric brain tumors. The four major SV categories and 49 subcategories of simple SVs are shown on the y axis. The names of the nine simple SV signatures are displayed on the top. The relative contributions of SV subcategories to the corresponding signatures are shown on the x axis. (B) Frequencies of simple SVs and percentages of simple SV signatures. In the upper panel, each dot represents one sample. Samples are grouped by tumor types. Red bars indicate median frequencies. The numbers in parentheses are sample sizes for the corresponding tumor types. The bottom panel shows the percentages of simple SV signatures in the corresponding samples on the top panel.

Figure 3 Associations of SV signatures with 31 genomic features. SV signatures and genomic features are listed in the x and y axes respectively. Each dot represents the association between one SV signature and one genomic feature. The bigger the dots are, the more significant the associations are. The colors of the dots indicate the directions of median shift. The directions of biases are shown in the parentheses of genomic features. For instance, the "ecDNA" signature has red color (left on the color key) dot for centromere, indicating that the observed SV breakpoints of this signature are farther away (left in the parentheses) from centromeres than randomized breakpoints.

Figure 4 SV breakpoint homology. The distributions (x axis) of homology and insertion at SV breakpoints are shown for all SV signatures (y axis). The putative DNA repair mechanisms are inferred from the sizes of homology and insertion and annotated next to the signatures. The bars indicate number of somatic SVs and are colored by tumor type. MMBIR: microhomology-

mediated break-induced repair. NHEJ: non-homologous end joining. alt-EJ: alternative end joining.

Figure 5 SV breakpoint hotspots. SV breakpoint frequencies are shown for the entire reference genome (x axis) across different SV signatures (y axis). Chromosome models are shown as grey bars with red lines indicating locations of centromeres at the bottom. Hotspots containing known oncogenes, tumor suppressors, and fragile sites are annotated.

Figure 6 *ZFTA* fusions in ependymomas. (A) The prevalence of *ZFTA* fusions in 71 ependymomas. Samples are colored by signatures of the SVs resulting in *ZFTA* fusions and ependymoma subtypes. (B) Six examples of *ZFTA* fusions resulting from different SV signatures. SV signatures and sample IDs are shown on the top. Somatic SVs, regions of complex SVs and copy number profiles are displayed in the same scheme as Figure 1C. The *ZFTA* and *RELA* regions are zoomed in and *ZFTA* gene and *RELA* gene are further zoomed in respectively. Gene structures are shown at the bottom of six examples. Within gene structures, the SV breakpoints that lead to *ZFTA-RELA* fusions are shown as red vertical lines. The directions of gene transcription are indicated by arrows.

Figure 7 SV signatures associated with patient survival. Kaplan-Meier survival curves for HGG patients, stratified by the presence or absence of four simple SV signatures ("Del2", "Unbal tra", "Del1", and "Large mixed") are shown. P values are calculated by log-rank test.

STAR Methods

Sample and data collection

The raw normal and tumor whole-genome sequencing data for 744 pediatric brain tumor patients were downloaded from CAVATICA (https://cavatica.sbgenomics.com/). Sample characteristics, clinical data, somatic SNV, and somatic CNV data were retrieved from OpenPBTA GitHub (https://github.com/AlexsLemonade/OpenPBTA-analysis, https://www.biorxiv.org/content/10.1101/2022.09.13.507832v1). All patients were de-identified.

Gene annotation was obtained from ENSEMBL (GRCh38.p13) (https://useast.ensembl.org/index.html). Non-B DNA structures including A-phased repeats, direct repeats, G-quadruplex forming repeats, inverted repeats, mirror repeats, short tandem repeats and Z-DNA motifs were downloaded from non-B DB (Cer et al., 2013) (https://nonb-abcc.ncifcrf.gov/apps/nBMST/default/); Alu, L1, L2, LTR, MIR, simple repeat, transposon, and low complexity repetitive elements, as well as the coordinates of centromeres, telomeres, and CpG islands, were obtained from UCSC (https://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/); consensus estimate of the topologically associated domains (TADs) were downloaded from TAD Map (Singh and Berger, 2021) (https://cb.csail.mit.edu/cb/tadmap/); ChIP-seq data of epigenetic markers H3K4me1, H3K9me3, H3K36me3, H3K4me3, H3K27me3, H3K9ac, H3K4me2, H3K79me2, H3K27ac, and H4K20me1 from human astrocyte were downloaded from ENCODE (Zhang et al., 2020) (https://www.encodeproject.org/). The Wavelet-smoothed signal of replication timing data for the cell lines BG02ES and SK-N-SH were downloaded from UCSC (https://genome.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEncodeUwRepliSeq). The fragile site regions were obtained from a previous study (Li et al., 2020), and the coordinates were lifted over to hg38. All coordinates in this study were based on the hg38 genome assembly unless otherwise noted.

Tumor classifications

The histological classifications of tumor samples were determined based on diagnosis, pathological examination, and histological examination according to the 2021 WHO classifications of pediatric brain tumors (Louis et al., 2021). For tumor types with at least 10 samples, we attempted to subclassify them. Gliomas were subclassified into high-grade gliomas (HGGs), low-grade astrocytic tumors (LGATs), ependymomas, dysembryoplastic neuroepithelial tumors (DNETs), and gangliogliomas. Embryonal tumors were subclassified into medulloblastomas, and atypical teratoid rhabdoid tumors (ATRTs). Cranial and paraspinal nerve tumors included schwannomas and neurofibromas. Germ cell tumors included teratomas, germinomas, and other germ cell tumors. Mesenchymal non-meningothelial tumors included hemangioblastomas, Ewing sarcomas, rhabdomyosarcomas, myofibroblastomas, and other sarcomas. Meningiomas, craniopharyngiomas, and choroid plexus tumors were independently classified. Tumor types with less than 10 samples were all classified into “Others”. Ependymomas were further stratified into distinct subtypes based on the primary sites.

Somatic SV calling and filtering

Manta (https://github.com/Illumina/manta), Meerkat (https://github.com/guru-yang/Meerkat), and Delly (https://github.com/dellytools/delly) were used for somatic SV detection. SVs called by Manta were obtained from OpenPBTA. Meerkat was run as suggested (Yang et al., 2013). Delly was run with default settings. For SVs detected by Delly, at least four supporting read pairs and split read combined were required for SVs less than 500 bp. For all three SV detection algorithms, only SVs located in canonical chromosomes (chr1-22, X, Y) were retained. SVs identified by different algorithms were considered identical if their two breakpoints were on the same chromosomes, with the same orientations and within 10 bp. SVs identified by two or more algorithms were considered high-confidence SVs and used in the subsequent analysis. Deletions with both breakpoints within 3 bp of exon-intron boundaries of the same genes were excluded from further analysis.

Somatic CNVs were used to assess the quality of somatic SVs. For each SV, if the distances of both SV breakpoints were less than 1 kb to the nearest CNV breakpoints, the SV was considered validated.

Complex SVs and their signatures

We used Starfish (Bao et al., 2022) (https://github.com/yanglab-computationalgenomics/Starfish) to detect clustered complex SVs and classified them into six signatures. In cases where reported gender and germline estimated sex were inconsistent, gender identity was recorded as unknown for signature detection. After removing clustered complex SVs, we used ClusterSV (Li et al., 2020) (https://github.com/cancerit/ClusterSV) to identify non-clustered complex SVs. Non-clustered complex SVs include chromoplexy, cycle of templated insertions, and complex unclear.

Simple SV signatures

After removing clustered and non-clustered complex SVs, the remainder were simple SVs including four major categories: deletions, tandem duplications, inversions, and translocations. Deletions and tandem duplications with breakpoints falling within fragile site regions were classified as fragile site deletions and fragile site tandem duplications, respectively. The remaining deletions and tandem duplications were classified into 18 subcategories based on their sizes. Inversions and translocations were further subclassified into reciprocal inversions, fold-back inversions, unbalanced inversions, reciprocal translocations, and unbalanced translocations. Unbalanced inversions and reciprocal inversions were classified into 3 and 5 subcategories based on their sizes, respectively. Fold-back inversions, unbalanced translocations, and reciprocal translocations were three independent subcategories. As a result, all simple SVs were classified into 49 subcategories and SigProfilerExtractor (Islam et al., 2022) (https://github.com/AlexandrovLab/SigProfilerExtractor) with default parameters was used to extract simple SV signatures. According to the final signatures we chose, deletions smaller than 1kb were assigned as “Del0”; deletions ranging in size from 1kb-5kb were classified as “Del1”; fragile site deletions, fragile site tandem duplications, and deletions sized in 5kb-10Mb were assigned as “Del2”; tandem duplications between 1Mb-2.5Mb with breakpoints located within the *BRAF* region were classified as *BRAF* fusion signature; other tandem duplications were

classified as “TD”; foldback inversions and unbalanced inversions sized between 50kb-5Mb were categorized as “Unbal inv”; deletions and tandem duplications larger than 10Mb, as well as reciprocal unbalanced inversions larger than 5Mb, were classified as “Large mixed”; reciprocal inversions and reciprocal translocations were categorized as “Recip”; and unbalanced translocations were classified as “Unbal tra”.

Genomic feature tests

For each observed somatic SV, we generated four random SVs on the same chromosome, same size, and same type. All observed and randomized breakpoints were annotated with genomic features. Bedtools was used to compute the GC content within a ±50 bp interval of each SV breakpoint. The distances in kilobases (kb) from the breakpoints to the nearest Non-B DNA structures, repetitive elements, and CpG islands were logarithmically transformed, with the distances set to 0 if breakpoints were within any of the aforementioned elements. The distances in megabases (Mb) from the breakpoints to centromeres and telomeres and the distances (kb) to the closest TAD boundaries were also transformed to log scale. The SV breakpoints were annotated by signal -log10(p-values) for different epigenetic modifications. The replication timing data for cell lines BG02ES and SK-N-SH were quantile normalized. The SV breakpoints were lifted over to hg19 since the replication timing data were based on hg19. The replication timing values were then annotated for each SV breakpoint. Breakpoints of observed SVs and randomized SVs were tested as described in the previous study (Li et al., 2020). Briefly, scores of SV breakpoints for all genomic features were rescaled from 0 to 1. The distributions of scores between observed breakpoints and randomized breakpoints were compared using two-sided Kolmogorov-Smirnov test. False discovery rates (FDRs) were computed using the Benjamini-Hochberg procedure and 0.1 FDR cutoff was used to determine significant associations. Homology and insertion size at the SV breakpoints were provided by Meerkat and Manta.

Hotspot analysis

The reference genome was divided into 1Mb non-overlapping bins. The number of samples with SV breakpoints in each bin was counted for each SV signature. A sample with multiple SV breakpoints of the same SV signatures falling within the same bin was only counted once.

Mutation test

Only protein-altering somatic SNVs and indels were considered, including missense mutations, splice site mutations, frameshift indels, nonsense mutations, translation start site mutations, and nonstop mutations. Three HGG samples (BS_20TBZG09, BS_02YBZSBY, and BS_VW4XN9Y7) with hypermutation were excluded. The tests were performed within tumor types. Protein-coding genes with mutation frequencies >= 5% in each tumor type were analyzed. Samples were classified into two categories based on the presence and absence of the SV signatures. Fisher’s exact test was used to calculate p values. FDRs were computed using the Benjamini-Hochberg procedure. FDR < 0.1 was considered as significant.

Survival analysis

Since patient survival differs dramatically across tumor types, survival analysis was only performed within tumor types, but not across tumor types. For clustered complex SV signatures, samples with only one signature were assigned to the corresponding signatures; samples with more than one clustered complex SV signatures were classified into “Mixed”; and samples without any clustered complex SV were assigned into “None”. For simple SV signatures, samples were classified based on the presence and absence of the signatures. Log-rank test was used to calculate p values.

References

Aguilera, A., and García-Muse, T. (2013). Causes of genome instability. Annu. Rev. Genet. *47*, 1–32.

Akdemir, K.C., Le, V.T., Chandran, S., Li, Y., Verhaak, R.G., Beroukhim, R., Campbell, P.J., Chin, L., Dixon, J.R., Futreal, P.A., et al. (2020). Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. Nat. Genet. *52*, 294–305.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A. V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. Nature *578*, 94–101.

Alt, F.W., Zhang, Y., Meng, F.L., Guo, C., and Schwer, B. (2013). Mechanisms of programmed DNA lesions and genomic instability in the immune system. Cell *152*, 417–429.

Baca, S.C., Prandi, D., Lawrence, M.S., Mosquera, J.M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T.Y., and Ghandi, M. (2013). Punctuated evolution of prostate cancer genomes. Cell *153*, 666–677.

Bakhshi, A., Jensen, J.P., Goldman, P., Wright, J.J., McBride, O.W., Epstein, A.L., and Korsmeyer, S.J. (1985). Cloning the chromosomal breakpoint of t(14;18) human lymphomas: clustering around Jh on chromosome 14 and near a transcriptional unit on 18. Cell *41*, 899–906.

Bao, L., Zhong, X., Yang, Y., and Yang, L. (2022). Starfish infers signatures of complex genomic rearrangements across human cancers. Nat. Cancer 2022 310 *3*, 1247–1259.

Berger, M.F., Lawrence, M.S., Demichelis, F., Drier, Y., Cibulskis, K., Sivachenko, A.Y., Sboner, A., Esgueva, R., Pflueger, D., Sougnez, C., et al. (2011). The genomic complexity of primary human prostate cancer. Nature *470*, 214–220.

Bunting, S.F., and Nussenzweig, A. (2013). End-joining, translocations and cancer. Nat. Rev. Cancer *13*, 443–454.

Campbell, P.J., Getz, G., Korbel, J.O., Stuart, J.M., Jennings, J.L., Stein, L.D., Perry, M.D., Nahal-Bose, H.K., Ouellette, B.F.F., Li, C.H., et al. (2020). Pan-cancer analysis of whole genomes. Nature *578*, 82–93.

Cer, R.Z., Donohue, D.E., Mudunuri, U.S., Temiz, N.A., Loss, M.A., Starner, N.J., Halusa, G.N., Volfovsky, N., Yi, M., Luke, B.T., et al. (2013). Non-B DB v2.0: a database of predicted non-B DNA-forming motifs and its associated tools. Nucleic Acids Res. *41*.

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., and Saunders, C.T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics *32*, 1220–1222.

Cortés-Ciriano, I., Lee, J.J.-K., Xi, R., Jain, D., Jung, Y.L., Yang, L., Gordenin, D., Klimczak, L.J., Zhang, C.-Z., Pellman, D.S., et al. (2020). Comprehensive analysis of chromothripsis in

2,658 human cancers using whole-genome sequencing. Nat. Genet. *52*, 331–341.

Degasperi, A., Zou, X., Amarante, T.D., Martinez-Martinez, A., Koh, G.C.C., Dias, J.M.L., Heskin, L., Chmelova, L., Rinaldi, G., Wang, V.Y.W., et al. (2022). Substitution mutational signatures in whole-genome–sequenced cancers in the UK population. Science *376*.

Drews, R.M., Hernando, B., Tarabichi, M., Haase, K., Lesluyes, T., Smith, P.S., Morrill Gavarró, L., Couturier, D.L., Liu, L., Schneider, M., et al. (2022). A pan-cancer compendium of chromosomal instability. Nature *606*, 976–983.

Dubois, F.P.B., Shapira, O., Greenwald, N.F., Zack, T., Wala, J., Tsai, J.W., Crane, A., Baguette, A., Hadjadj, D., Harutyunyan, A.S., et al. (2022). Structural variants shape driver combinations and outcomes in pediatric high-grade glioma. Nat. Cancer 2022 38 *3*, 994–1011.

Farmer, H., McCabe, H., Lord, C.J., Tutt, A.H.J., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., et al. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature *434*, 917–921.

Fong, P.C., Boss, D.S., Yap, T.A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M.J., et al. (2009). Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from BRCA Mutation Carriers. N. Engl. J. Med. *361*, 123–134.

Gostissa, M., Yan, C.T., Bianco, J.M., Cogné, M., Pinaud, E., and Alt, F.W. (2009). Long-range oncogenic activation of Igh-c-myc translocations by the Igh 3′ regulatory region. Nature *462*, 803–807.

Gröbner, S.N., Worst, B.C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V.A., Johann, P.D., Balasubramanian, G.P., Segura-Wang, M., Brabetz, S., et al. (2018). The landscape of genomic alterations across childhood cancers. Nature *555*, 321–327.

Hakim, O., Resch, W., Yamane, A., Klein, I., Kieffer-Kwon, K.R., Jankovic, M., Oliveira, T., Bothmer, A., Voss, T.C., Ansarah-Sobrinho, C., et al. (2012). DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. Nature *484*, 69–74.

Islam, S.M.A., Díaz-Gay, M., Wu, Y., Barnes, M., Vangara, R., Bergstrom, E.N., He, Y., Vella, M., Wang, J., Teague, J.W., et al. (2022). Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. Cell Genomics *0*, 100179.

Lee, J.-K., Choi, Y.-L., Kwon, M., and Park, P.J. (2016). Mechanisms and Consequences of Cancer Genome Instability: Lessons from Genome Sequencing Studies. Annu. Rev. Pathol. Mech. Dis. *11*, 283–312.

Li, Y., Roberts, N.D., Wala, J.A., Shapira, O., Schumacher, S.E., Kumar, K., Khurana, E., Waszak, S., Korbel, J.O., Haber, J.E., et al. (2020). Patterns of somatic structural variation in human cancer genomes. Nature *578*, 112–121.

Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., Burger, P.C., Jouvet, A., Scheithauer, B.W., and Kleihues, P. (2007). The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathol. *114*, 97–109.

Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H.K., Pfister, S.M., Reifenberger, G., et al. (2021). The 2021 WHO Classification of Tumors

of the Central Nervous System: a summary. Neuro. Oncol. *23*, 1231–1251.

Maciejowski, J., Li, Y., Bosco, N., Campbell, P.J., and de Lange, T. (2015). Chromothripsis and Kataegis Induced by Telomere Crisis. Cell *163*, 1641–1654.

Mack, S.C., and Taylor, M.D. (2017). Put away your microscopes: The ependymoma molecular era has begun. Curr. Opin. Oncol. *29*, 443.

McVey, M., and Lee, S.E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. *24*, 529–538.

Molenaar, J.J., Koster, J., Zwijnenburg, D.A., van Sluis, P., Valentijn, L.J., van der Ploeg, I., Hamdi, M., van Nes, J., Westerman, B.A., van Arkel, J., et al. (2012). Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. Nature *483*, 589–593.

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature *534*, 47–54.

Notta, F., Chan-Seng-Yue, M., Lemire, M., Li, Y., Wilson, G.W., Connor, A.A., Denroche, R.E., Liang, S.-B., Brown, A.M.K., Kim, J.C., et al. (2016). A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. Nature *538*, 378–382.

Ostrom, Q.T., de Blank, P.M., Kruchko, C., Petersen, C.M., Liao, P., Finlay, J.L., Stearns, D.S., Wolff, J.E., Wolinsky, Y., Letterio, J.J., et al. (2015). Alex's Lemonade Stand Foundation Infant and Childhood Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2007–2011. Neuro. Oncol. *16*, x1–x36.

Parker, M., Mohankumar, K.M., Punchihewa, C., Weinlich, R., Dalton, J.D., Li, Y., Lee, R., Tatevossian, R.G., Phoenix, T.N., Thiruvenkatam, R., et al. (2014). C11orf95–RELA fusions drive oncogenic NF-κB signalling in ependymoma. Nature *506*, 451–455.

Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012a). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics *28*, i333–i339.

Rausch, T., Jones, D.T.W., Zapatka, M., Stütz, A.M., Zichner, T., Weischenfeldt, J., Jäger, N., Remke, M., Shih, D., Northcott, P.A., et al. (2012b). Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. Cell *148*, 59–71.

Shoshani, O., Brunner, S.F., Yaeger, R., Ly, P., Nechemia-Arbely, Y., Kim, D.H., Fang, R., Castillon, G.A., Yu, M., Li, J.S.Z., et al. (2020). Chromothripsis drives the evolution of gene amplification in cancer. Nature *591*, 137–141.

Singh, R., and Berger, B. (2021). Deciphering the species-level structure of topologically associating domains. BioRxiv 2021.10.28.466333.

Steele, C.D., Abbasi, A., Islam, S.M.A., Bowes, A.L., Khandekar, A., Haase, K., Hames-Fathi, S., Ajayi, D., Verfaillie, A., Dhami, P., et al. (2022). Signatures of copy number alterations in human cancer. Nature *606*, 984–991.

660 Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D.,
661 Lau, K.W., Beare, D., Stebbings, L.A., et al. (2011). Massive genomic rearrangement acquired in
662 a single catastrophic event during cancer development. Cell *144*, 27–40.

663 Yang, L., Luquette, L.J., Gehlenborg, N., Xi, R., Haseley, P.S., Hsieh, C.-H., Zhang, C., Ren, X.,
664 Protopopov, A., Chin, L., et al. (2013). Diverse mechanisms of somatic structural variations in
665 human cancer genomes. Cell *153*, 919–929.

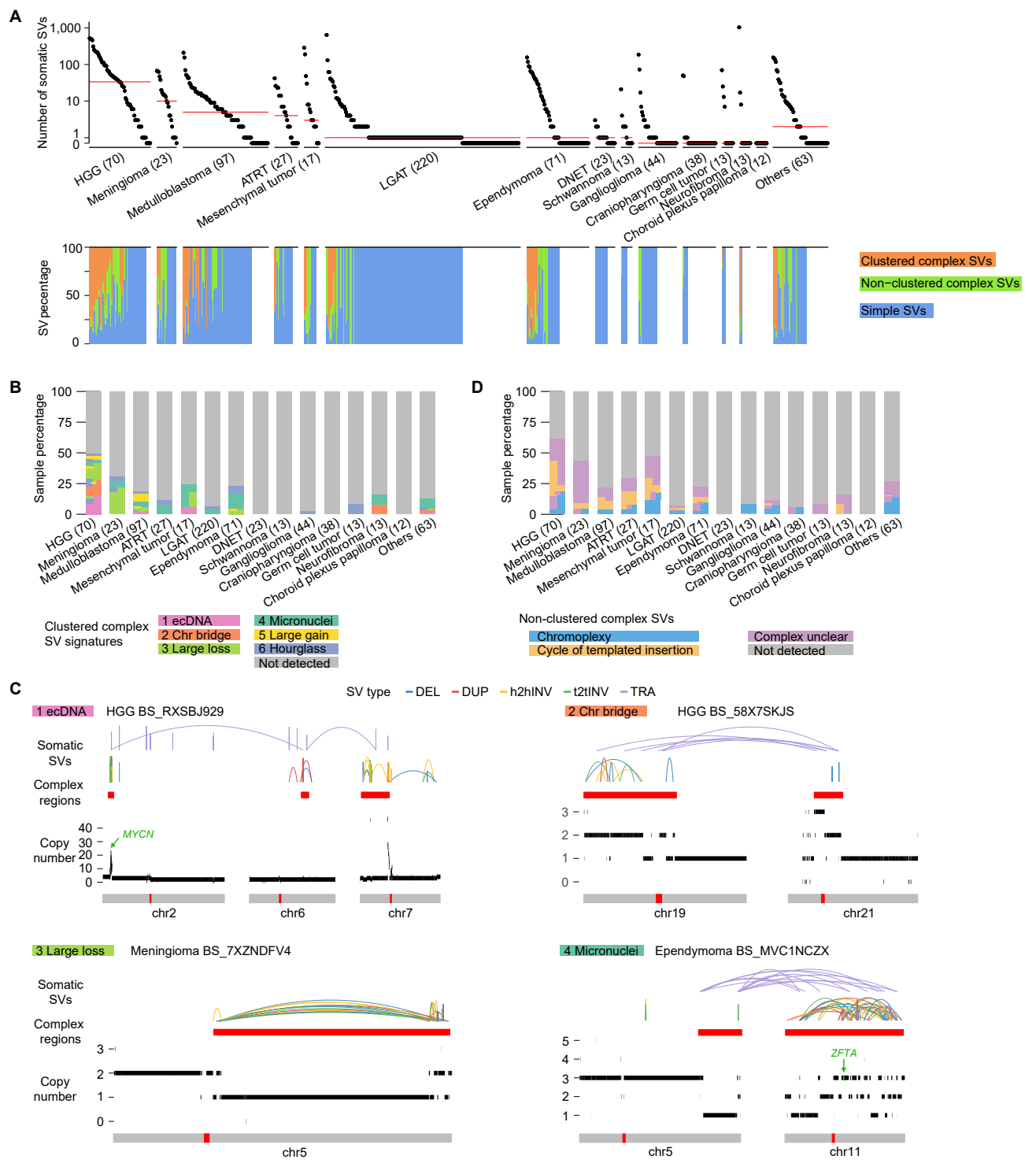
666 Zhang, C.-Z., Spektor, A., Cornils, H., Francis, J.M., Jackson, E.K., Liu, S., Meyerson, M., and
667 Pellman, D. (2015). Chromothripsis from DNA damage in micronuclei. Nature *522*, 179–184.

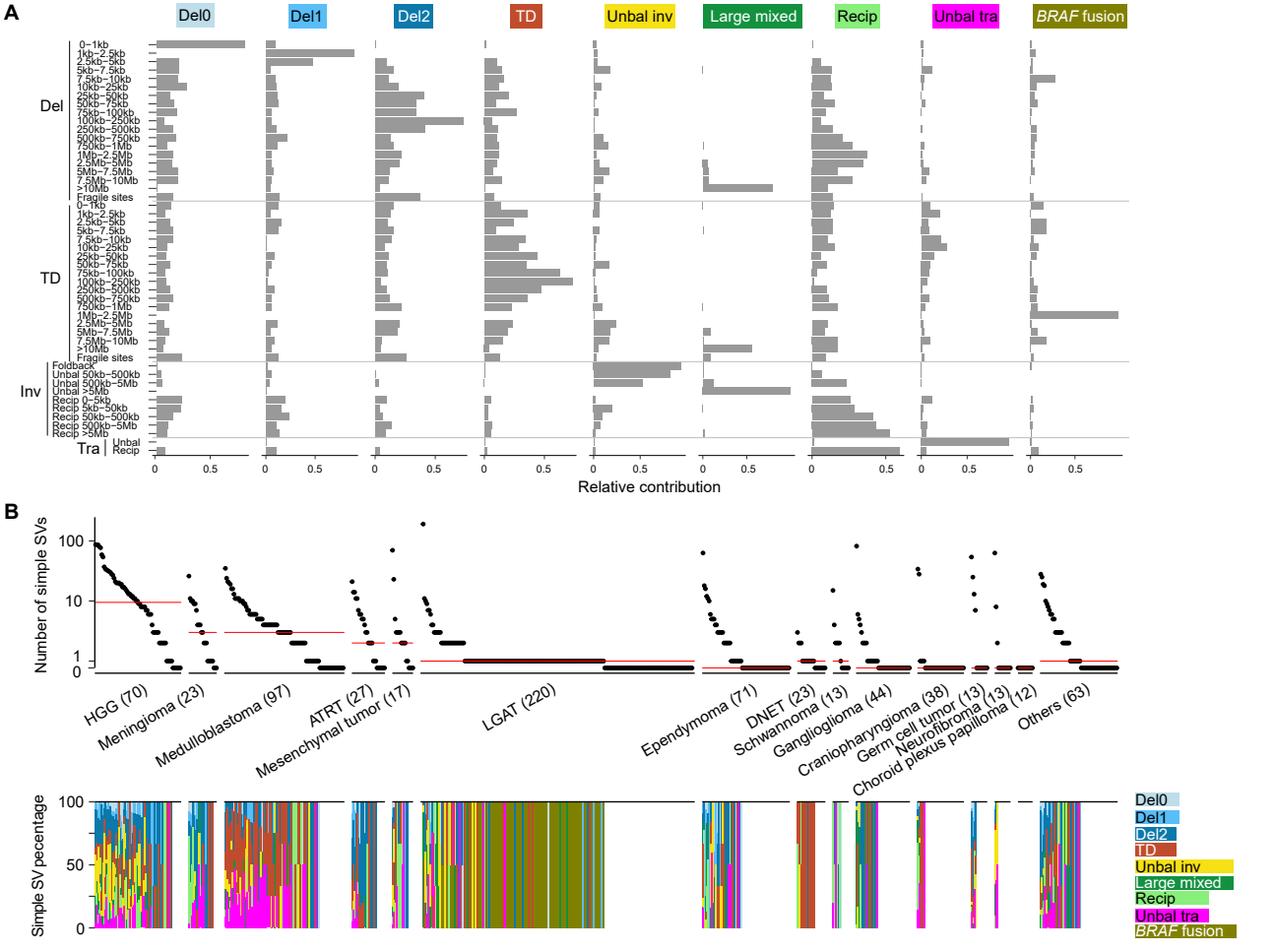
668 Zhang, F., Khajavi, M., Connolly, A.M., Towne, C.F., Batish, S.D., and Lupski, J.R. (2009). The
669 DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex
670 rearrangements in humans. Nat. Genet. *41*, 849–853.

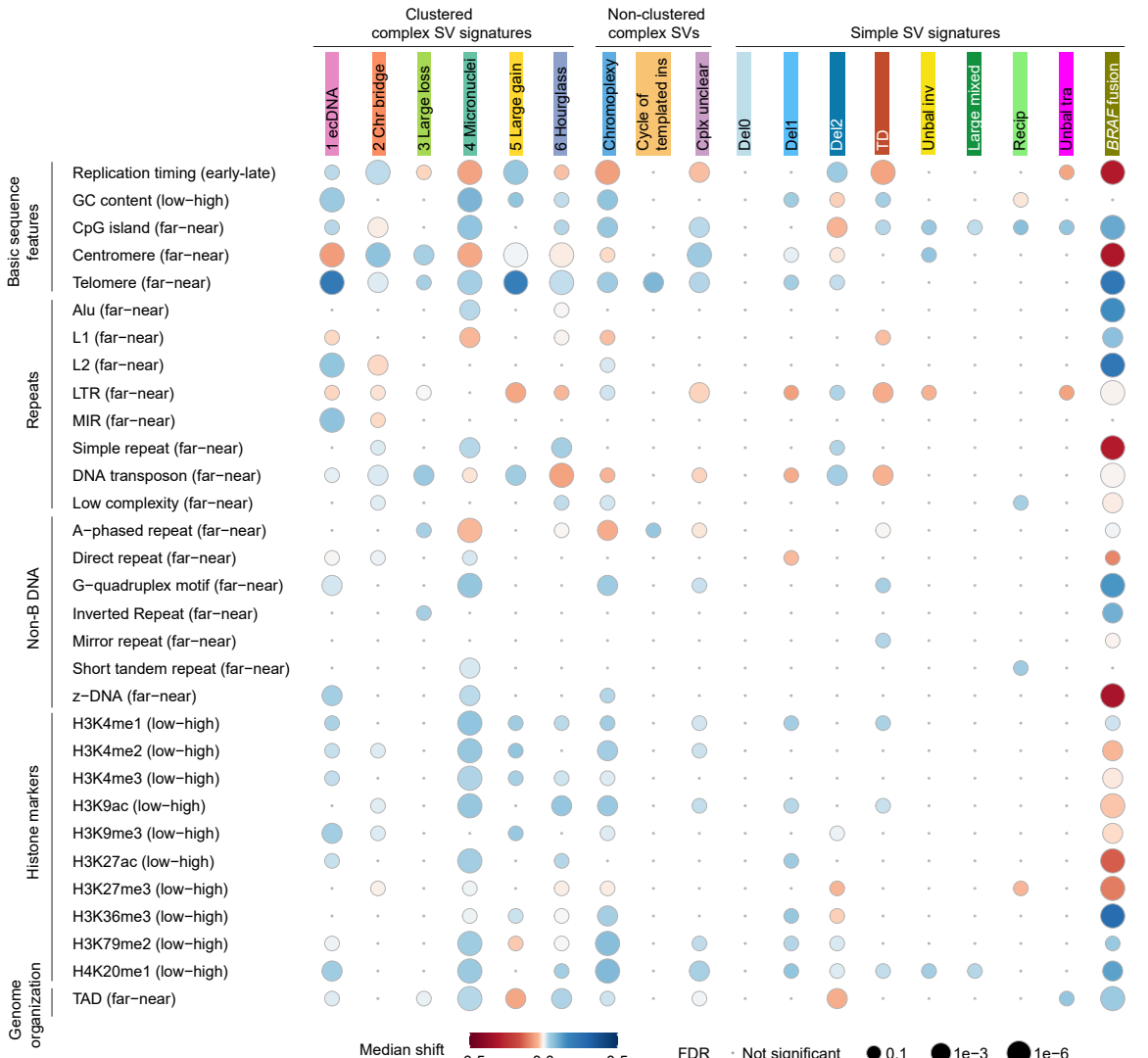
671 Zhang, J., Lee, D., Dhiman, V., Jiang, P., Xu, J., McGillivray, P., Yang, H., Liu, J., Meyerson,
672 W., Clarke, D., et al. (2020). An integrative ENCODE resource for cancer genomics. Nat.
673 Commun. *11*.

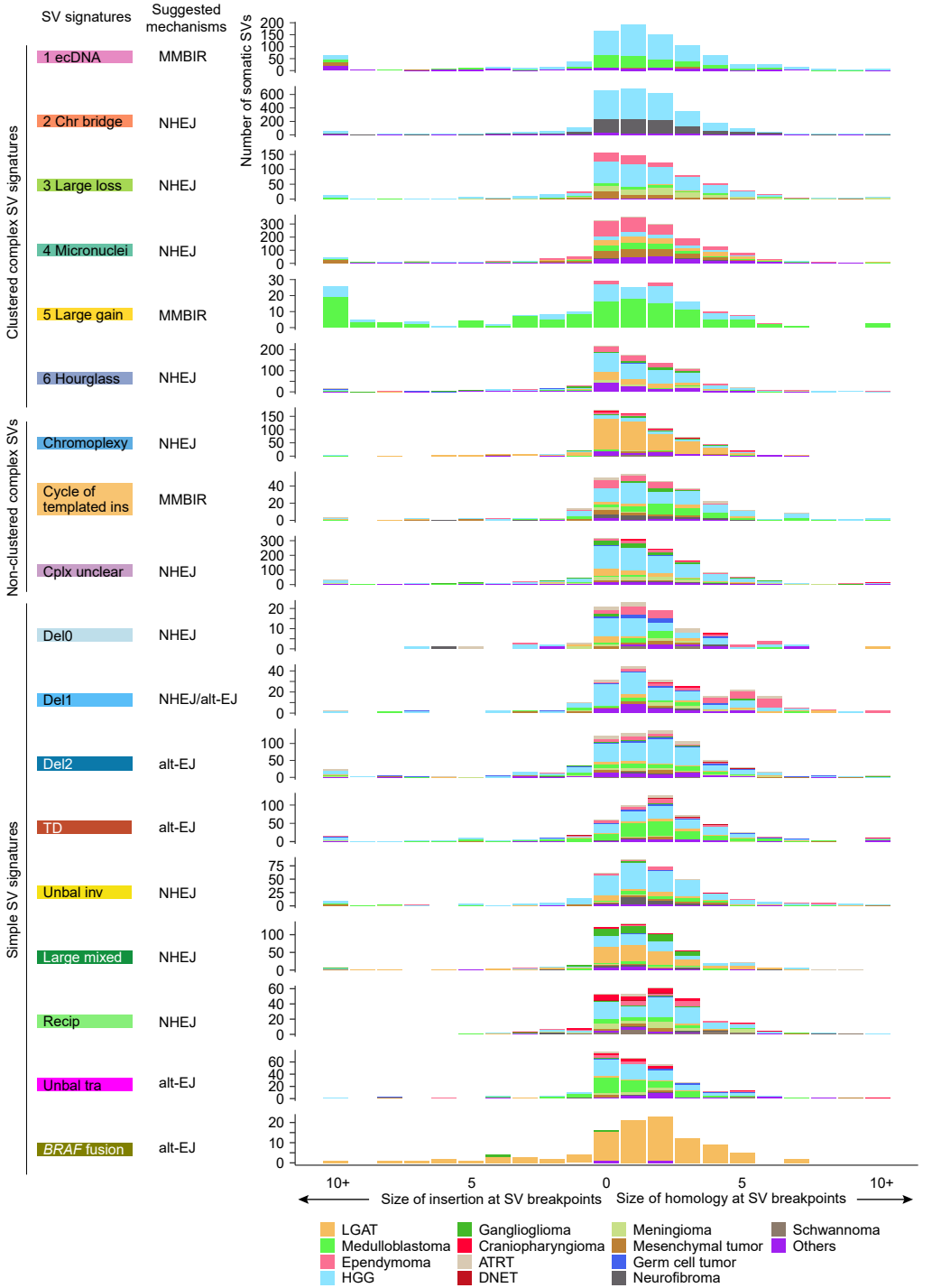
674 Zhang, Y., Chen, F., Donehower, L.A., Scheurer, M.E., and Creighton, C.J. (2021). A pediatric
675 brain tumor atlas of genes deregulated by somatic genomic rearrangement. Nat. Commun. *12*,
676 937.

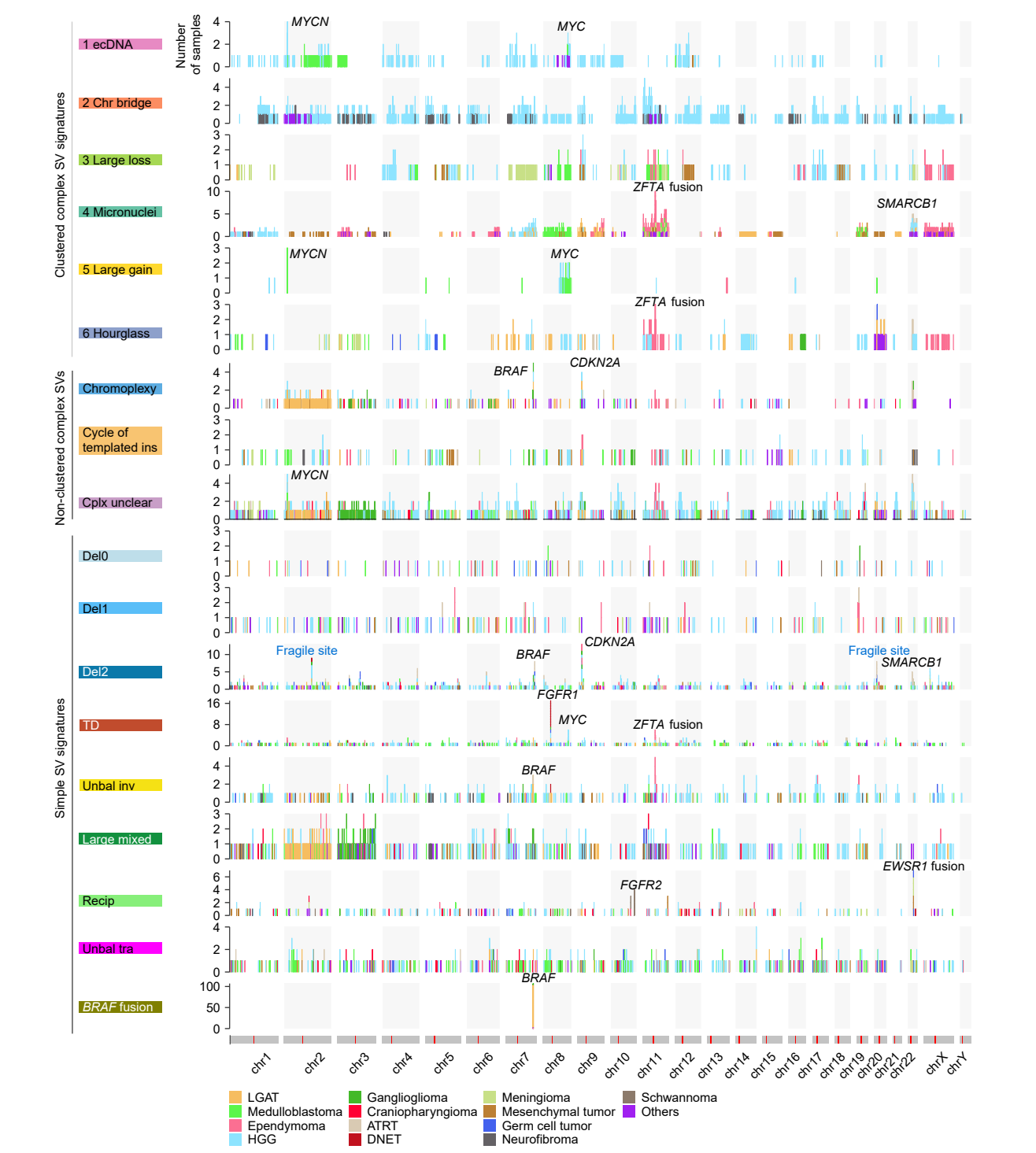
677

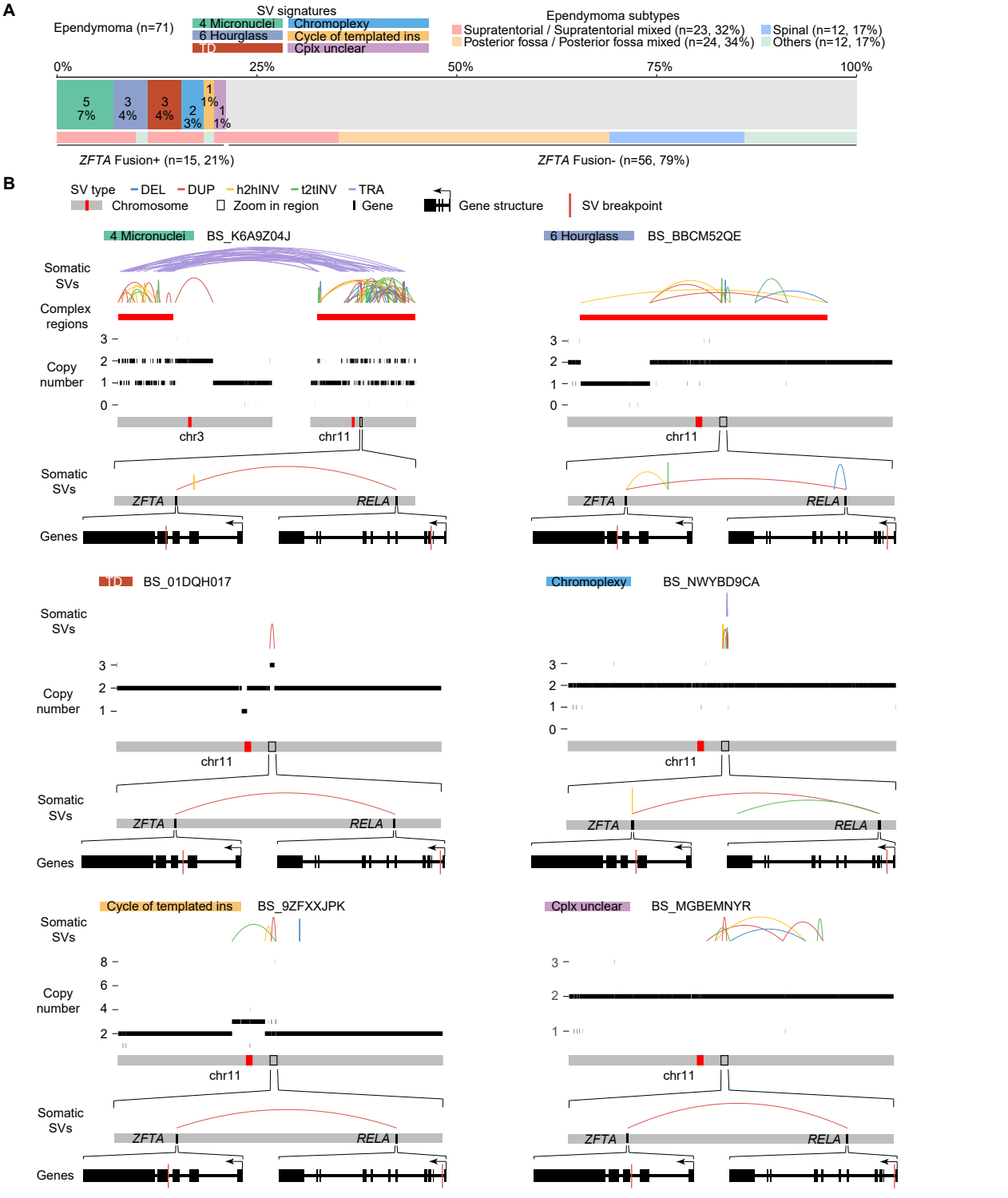

A
Number of somatic SVs
1,000
100
10
1
0
HGG (70)
Meningioma (23)
Medulloblastoma (97)
ATRT (27)
Mesenchymal tumor (17)
LGAT (220)
Ependymoma (71)
DNET (23)
Schwannoma (13)
Ganglioglioma (44)
Craniopharyngioma (38)
Germ cell tumor (13)
Neurofibroma (13)
Choroid plexus papilloma (12)
Others (63)
SV pecentage
100
50
0
Clustered complex SVs
Non−clustered complex SVs
Simple SVs
B
Sample percentage
100
75
50
25
0
Clustered complex SV signatures
1 ecDNA
2 Chr bridge
3 Large loss
4 Micronuclei
5 Large gain
6 Hourglass
Not detected
D
Sample percentage
Non-clustered complex SVs
Chromoplexy
Cycle of templated insertion
Complex unclear
Not detected
C
SV type
DEL
DUP
h2hINV
t2tINV
TRA
1 ecDNA
HGG BS_RXSBJ929
Somatic SVs
Complex regions
Copy number
MYCN
chr2
chr6
chr7
2 Chr bridge
HGG BS_58X7SKJS
chr19
chr21
3 Large loss
Meningioma BS_7XZNDFV4
chr5
4 Micronuclei
Ependymoma BS_MVC1NCZX
ZFTA
chr5
chr11


A
Del0
Del1
Del2
TD
Unbal inv
Large mixed
Recip
Unbal tra
BRAF fusion
Del
TD
Inv
Tra
Unbal
Recip
0–1kb
1kb–2.5kb
2.5kb–5kb
5kb–7.5kb
7.5kb–10kb
10kb–25kb
25kb–50kb
50kb–75kb
75kb–100kb
100kb–250kb
250kb–500kb
500kb–750kb
750kb–1Mb
1Mb–2.5Mb
2.5Mb–5Mb
5Mb–7.5Mb
7.5Mb–10Mb
>10Mb
Fragile sites
Foldback
Unbal 50kb–500kb
Unbal 500kb–5Mb
Unbal >5Mb
Recip 0–5kb
Recip 5kb–50kb
Recip 50kb–500kb
Recip 500kb–5Mb
Recip >5Mb
0
0.5
Relative contribution
B
Number of simple SVs
100
10
1
0
HGG (70)
Meningioma (23)
Medulloblastoma (97)
ATRT (27)
Mesenchymal tumor (17)
LGAT (220)
Ependymoma (71)
DNET (23)
Schwannoma (13)
Ganglioglioma (44)
Craniopharyngioma (38)
Germ cell tumor (13)
Neurofibroma (13)
Choroid plexus papilloma (12)
Others (63)
Simple SV pecentage
100
50
0
Del0
Del1
Del2
TD
Unbal inv
Large mixed
Recip
Unbal tra
BRAF fusion

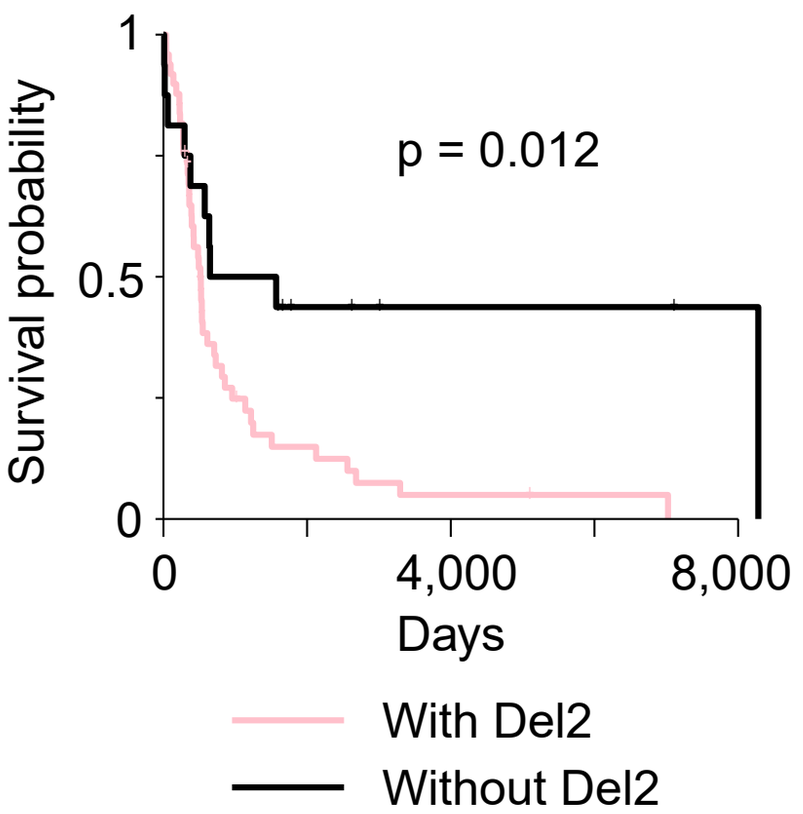

Clustered complex SV signatures
Non-clustered complex SVs
Simple SV signatures
1 ecDNA
2 Chr bridge
3 Large loss
4 Micronuclei
5 Large gain
6 Hourglass
Chromoplexy
Cycle of templated ins
Cplx unclear
Del0
Del1
Del2
TD
Unbal inv
Large mixed
Recip
Unbal tra
BRAF fusion
Basic sequence features
Replication timing (early-late)
GC content (low–high)
CpG island (far–near)
Centromere (far–near)
Telomere (far–near)
Repeats
Alu (far–near)
L1 (far–near)
L2 (far–near)
LTR (far–near)
MIR (far–near)
Simple repeat (far–near)
DNA transposon (far–near)
Low complexity (far–near)
Non-B DNA
A–phased repeat (far–near)
Direct repeat (far–near)
G–quadruplex motif (far–near)
Inverted Repeat (far–near)
Mirror repeat (far–near)
Short tandem repeat (far–near)
z–DNA (far–near)
Histone markers
H3K4me1 (low–high)
H3K4me2 (low–high)
H3K4me3 (low–high)
H3K9ac (low–high)
H3K9me3 (low–high)
H3K27ac (low–high)
H3K27me3 (low–high)
H3K36me3 (low–high)
H3K79me2 (low–high)
H4K20me1 (low–high)
Genome organization
TAD (far–near)
Median shift
−0.5
0.0
0.5
FDR
Not significant
0.1
1e−3
1e−6

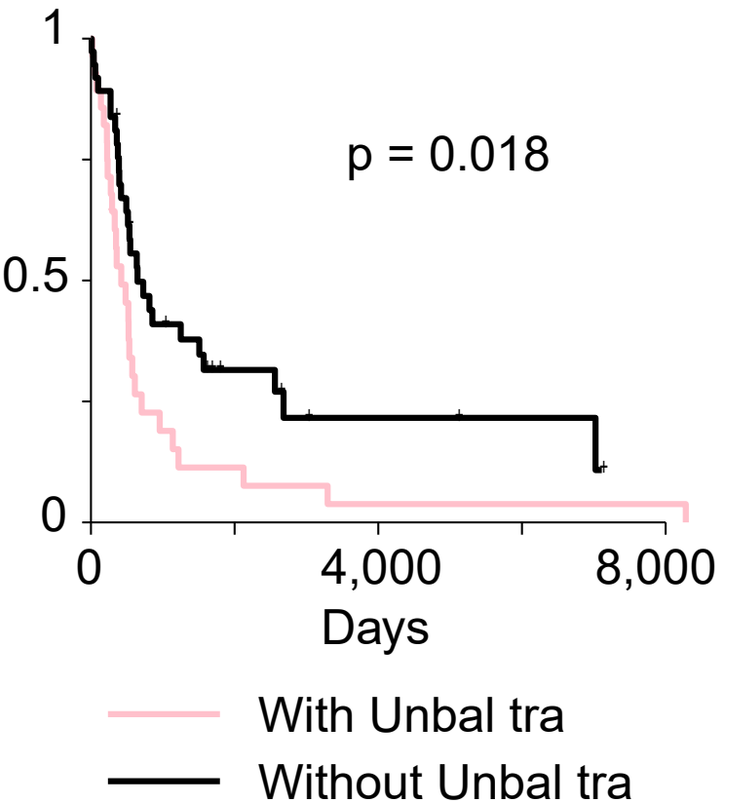

SV signatures
Suggested mechanisms
Clustered complex SV signatures
Non-clustered complex SVs
Simple SV signatures
1 ecDNA
MMBIR
2 Chr bridge
NHEJ
3 Large loss
NHEJ
4 Micronuclei
NHEJ
5 Large gain
MMBIR
6 Hourglass
NHEJ
Chromoplexy
NHEJ
Cycle of templated ins
MMBIR
Cplx unclear
NHEJ
Del0
NHEJ
Del1
NHEJ/alt-EJ
Del2
alt-EJ
TD
alt-EJ
Unbal inv
NHEJ
Large mixed
NHEJ
Recip
NHEJ
Unbal tra
alt-EJ
BRAF fusion
alt-EJ
Number of somatic SVs
10+
5
0
5
10+
Size of insertion at SV breakpoints
Size of homology at SV breakpoints
LGAT
Medulloblastoma
Ependymoma
HGG
Ganglioglioma
Craniopharyngioma
ATRT
DNET
Meningioma
Mesenchymal tumor
Germ cell tumor
Neurofibroma
Schwannoma
Others

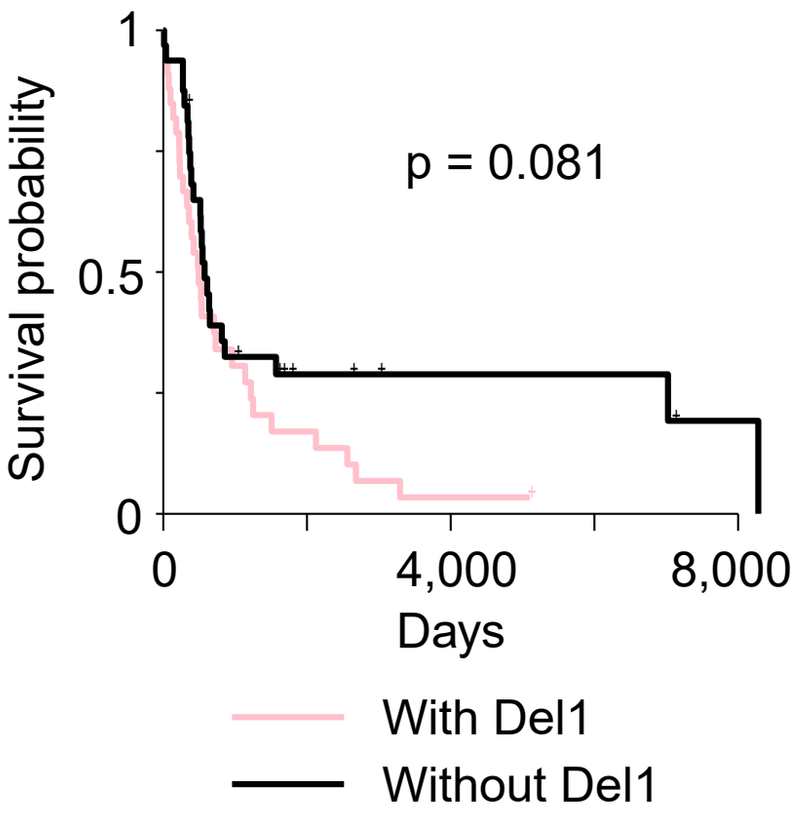

Clustered complex SV signatures
Non-clustered complex SVs
Simple SV signatures
1 ecDNA
2 Chr bridge
3 Large loss
4 Micronuclei
5 Large gain
6 Hourglass
Chromoplexy
Cycle of templated ins
Cplx unclear
Del0
Del1
Del2
TD
Unbal inv
Large mixed
Recip
Unbal tra
BRAF fusion
Number of samples
MYCN
MYC
ZFTA fusion
SMARCB1
BRAF
CDKN2A
Fragile site
FGFR1
FGFR2
EWSR1 fusion
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
chrY
LGAT
Medulloblastoma
Ependymoma
HGG
Ganglioglioma
Craniopharyngioma
ATRT
DNET
Meningioma
Mesenchymal tumor
Germ cell tumor
Neurofibroma
Schwannoma
Others

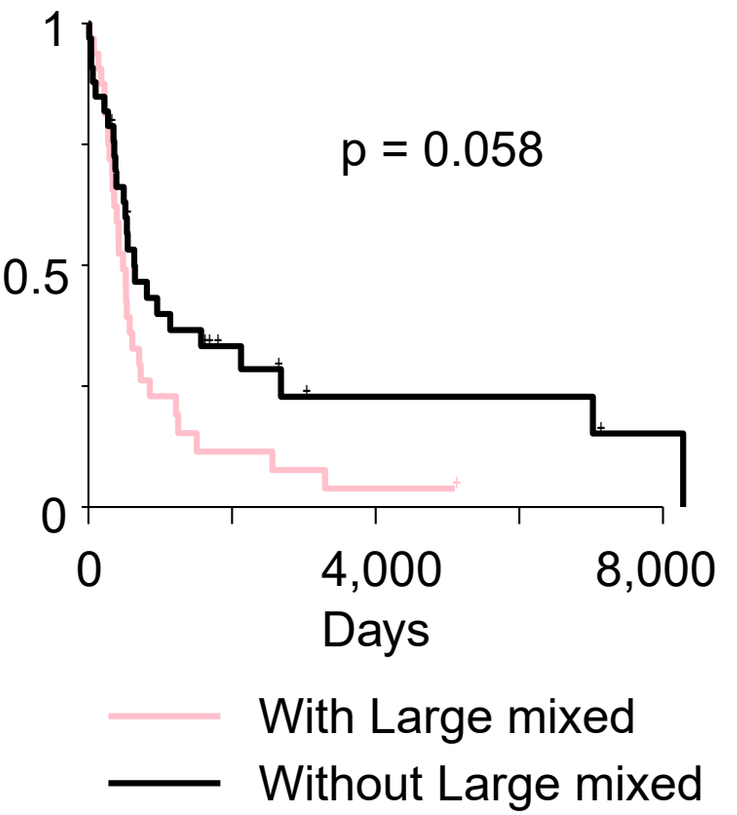

A
Ependymoma (n=71)
SV signatures
4 Micronuclei
6 Hourglass
TD
Chromoplexy
Cycle of templated ins
Cplx unclear
Ependymoma subtypes
Supratentorial / Supratentorial mixed (n=23, 32%)
Posterior fossa / Posterior fossa mixed (n=24, 34%)
Spinal (n=12, 17%)
Others (n=12, 17%)
0%
25%
50%
75%
100%
5 7%
3 4%
3 4%
2 3%
1 1%
1 1%
ZFTA Fusion+ (n=15, 21%)
ZFTA Fusion- (n=56, 79%)
B
SV type
DEL
DUP
h2hINV
t2tINV
TRA
Chromosome
Zoom in region
Gene
Gene structure
SV breakpoint
4 Micronuclei
BS_K6A9Z04J
6 Hourglass
BS_BBCM52QE
TD
BS_01DQH017
Chromoplexy
BS_NWYBD9CA
Cycle of templated ins
BS_9ZFXXJPK
Cplx unclear
BS_MGBEMNYR
Somatic SVs
Complex regions
Copy number
chr3
chr11
ZFTA
RELA
Genes

HGG (n=65)


Survival probability
1
0.5
0
p = 0.012
0
4,000
8,000
Days
With Del2
Without Del2


1
0.5
0
p = 0.018
0
4,000
8,000
Days
With Unbal tra
Without Unbal tra


Survival probability
1
0.5
0
p = 0.081
0
4,000
8,000
Days
With Del1
Without Del1


1
0.5
0
p = 0.058
0
4,000
8,000
Days
With Large mixed
Without Large mixed