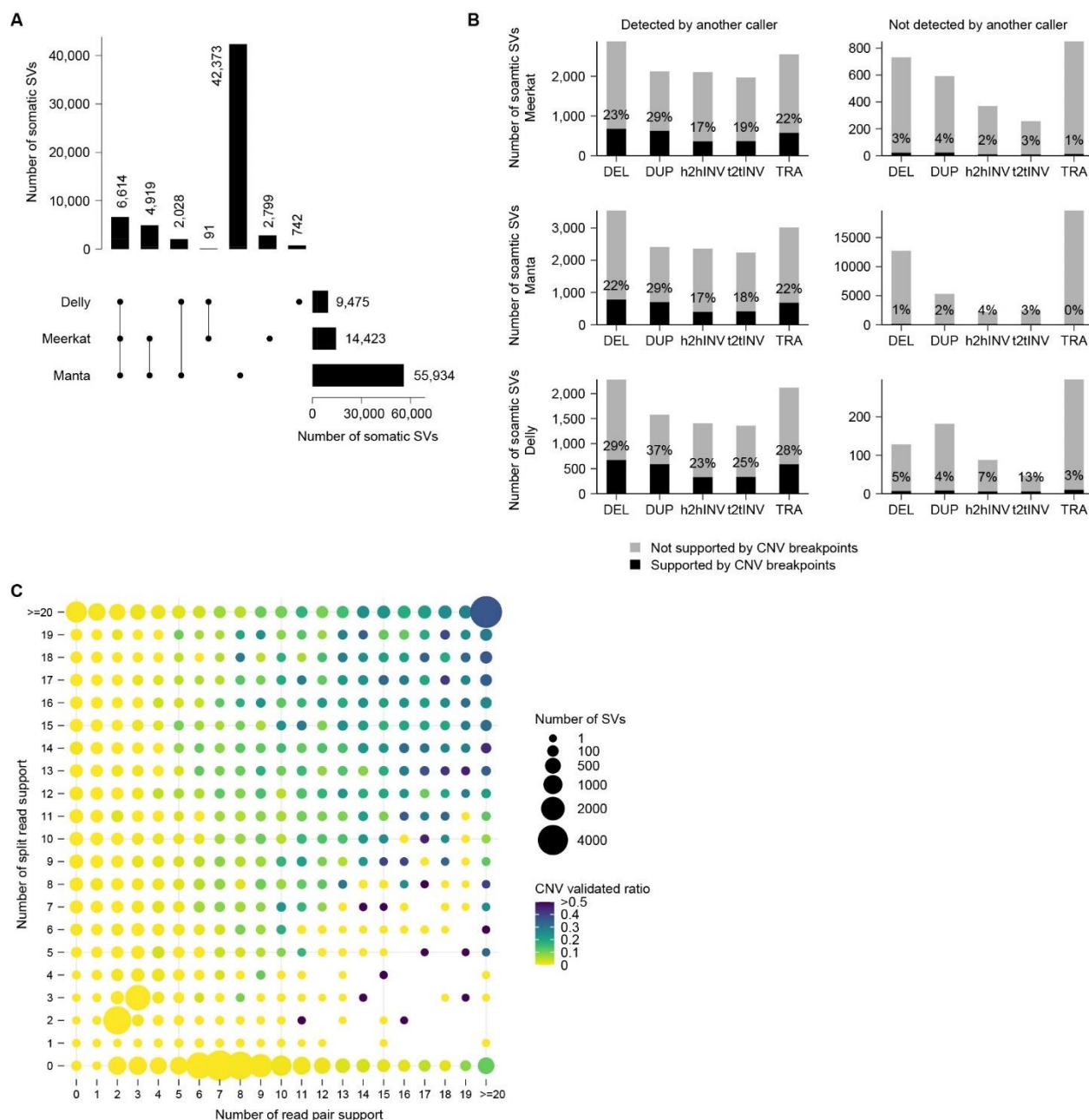


**Cell Reports, Volume 42**

**Supplemental information**

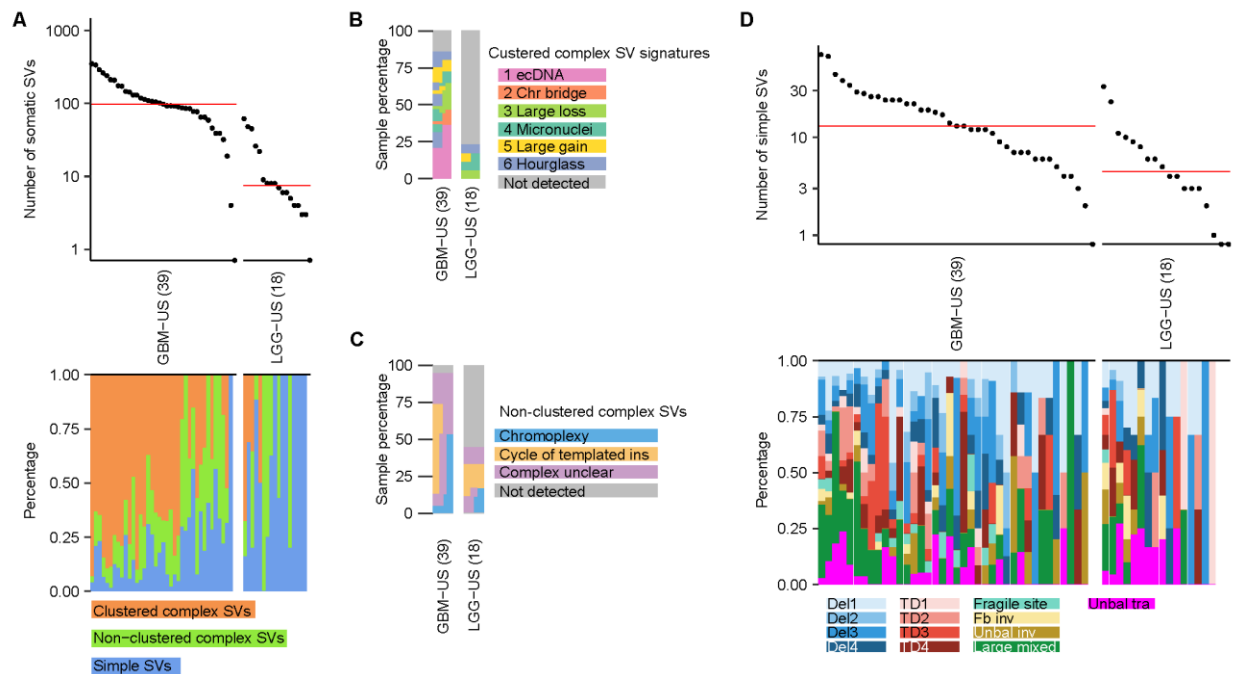
**Somatic structural variation signatures  
in pediatric brain tumors**

**Yang Yang and Lixing Yang**

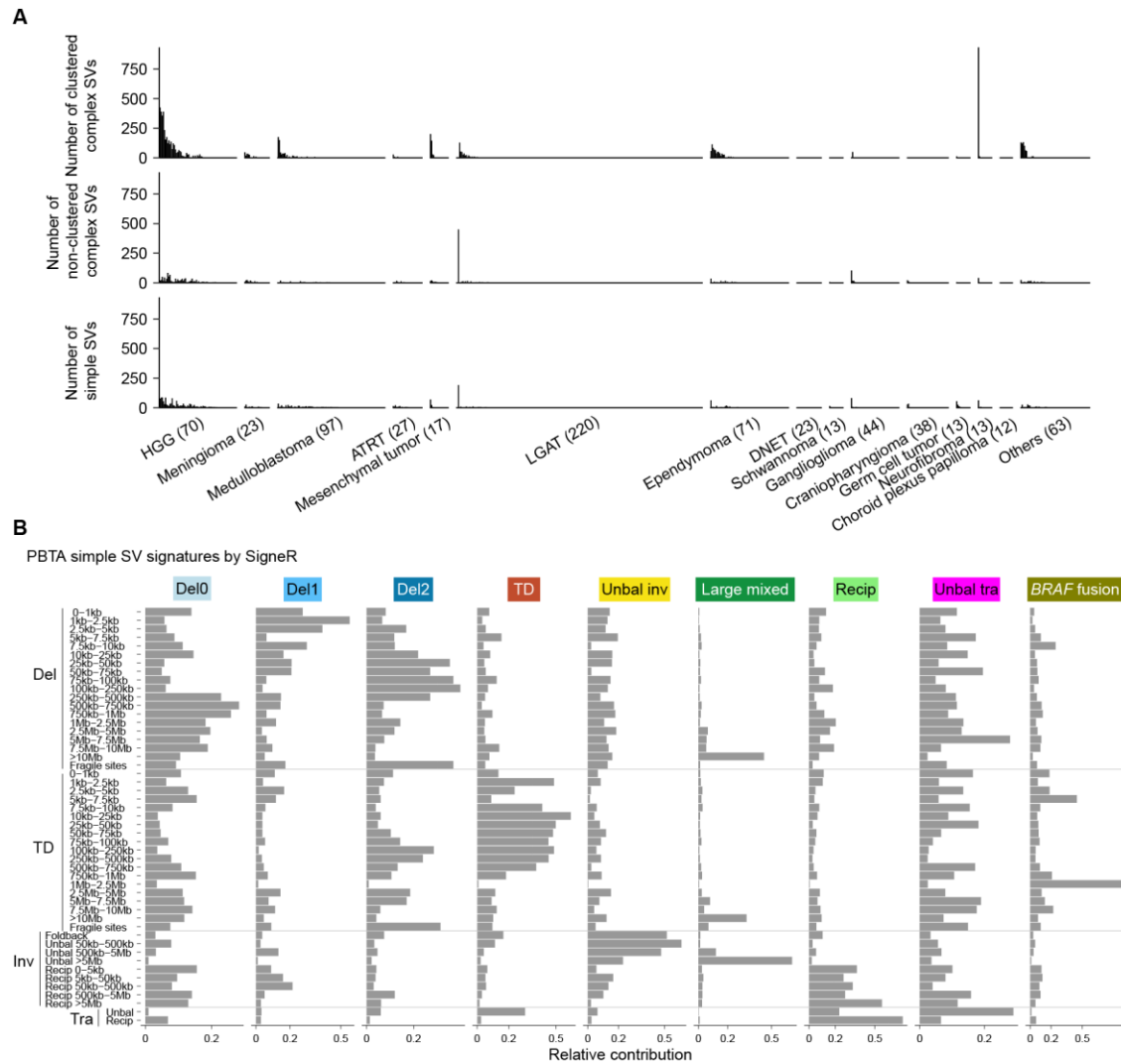


**Figure S1 Quality control of somatic SVs called by Manta, Meerkat, and Delly, related to Figure 1 and STAR Methods.** (A) UpSet plot of somatic SVs called by Meerkat, Manta, and Delly. The bars on the bottom right are the total SVs detected by the three tools respectively. The bars on the top show the number of SVs identified by one or more tools. The black dots under the bars indicate tools used. The numbers on the top and on the right side of the bars are numbers of SVs. (B) SVs called by different tools validated by CNVs. The left panels are high confidence SVs called by more than one tool, and the right panels are tool-specific low confidence SVs. The black and grey bars indicate SVs supported or not supported by CNV breakpoints, respectively. (C) CNV validation of Manta SVs. The x- and y-axes are the numbers of read pair and split-read

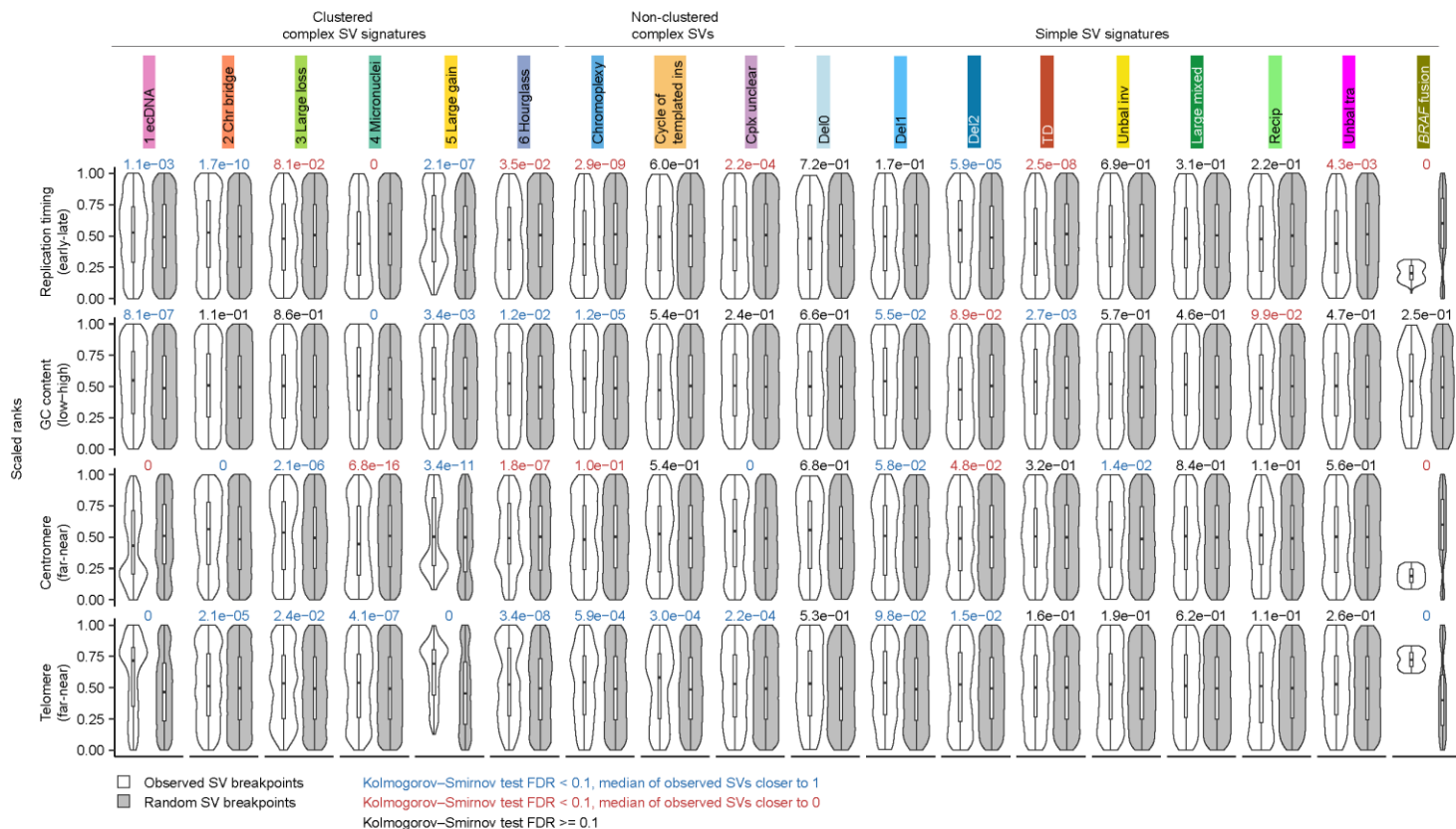
support for SVs detected by Manta, respectively. The size of each dot represents the number of SVs, and the color indicates the proportion of SVs validated by CNVs.



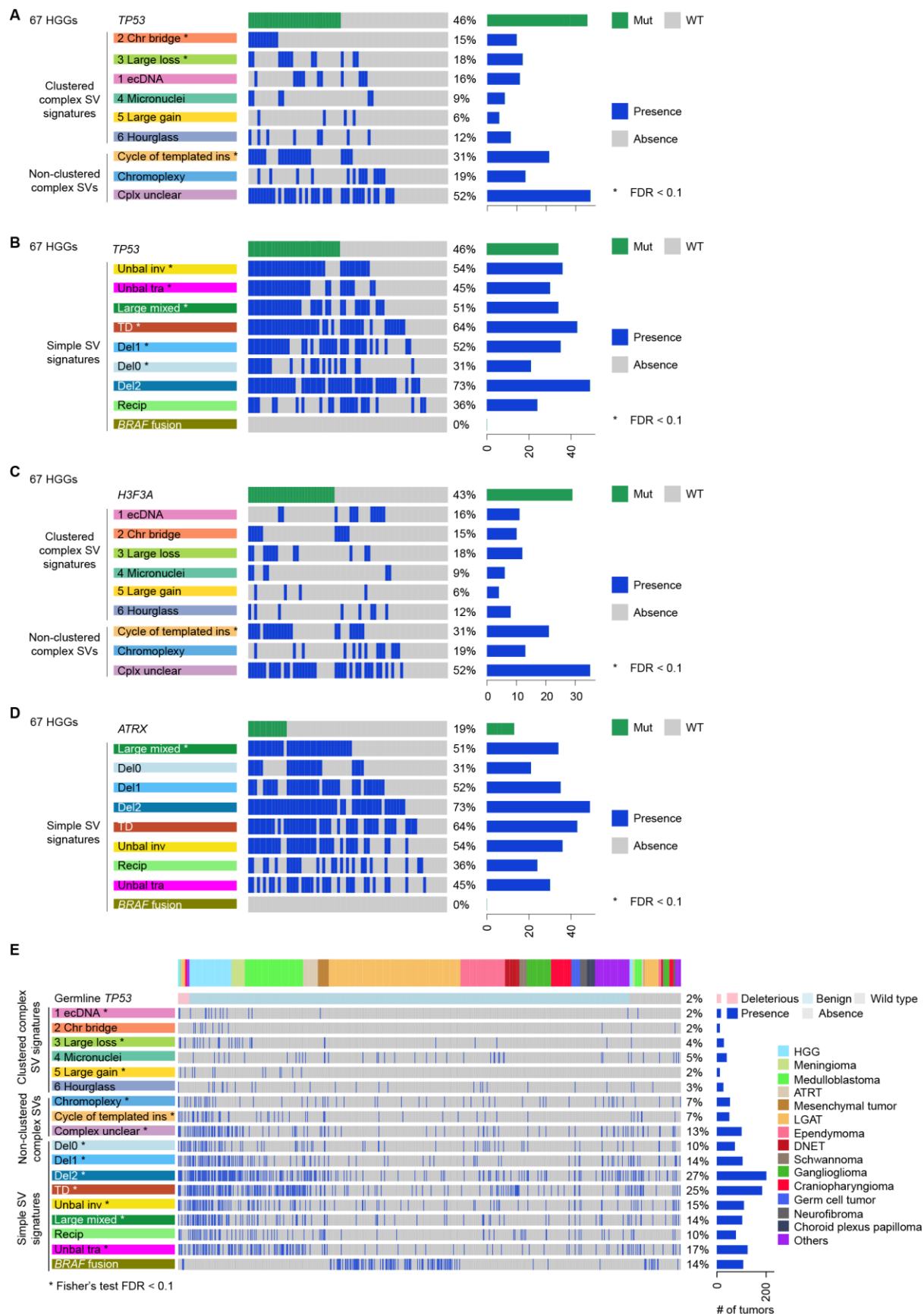
**Figure S2 Somatic SVs in PCAWG adult brain tumors, related to Figures 1 and 2.** (A) Frequencies of somatic SVs and percentages of different types of SVs in adult brain tumors. In the upper panel, each dot represents one adult brain tumor sample. Samples are grouped by tumor type and tumor types are sorted by median SV frequency (red lines). 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. GBM: glioblastoma multiforme. LGG: low-grade gliomas. (B) and (C) Percentages of clustered complex SV signatures and non-clustered complex SVs. Each vertical block represents one tumor type and each horizontal bar represents 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. (D) Frequencies of somatic simple SVs and percentages of different adult simple SV signatures in adult brain tumors. Samples are grouped by tumor type and tumor types are sorted by median SV frequency (red lines). The numbers in parentheses are sample sizes for the corresponding tumor types. The bottom panel shows the percentages of different adult simple SV signatures in the corresponding samples on the top panel.



**Figure S3 Frequencies of complex SVs, simple SVs, and simple SV signatures decomposed by signeR, related to Figures 1 and 2.** (A) The three tracks present the number of clustered complex SVs, non-clustered complex SVs, and simple SVs, respectively. The tumor type ordering is consistent with Figure 1, and samples are sorted by the total number of SVs. The numbers in parentheses represent the sample sizes for the corresponding tumor types. (B) 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.

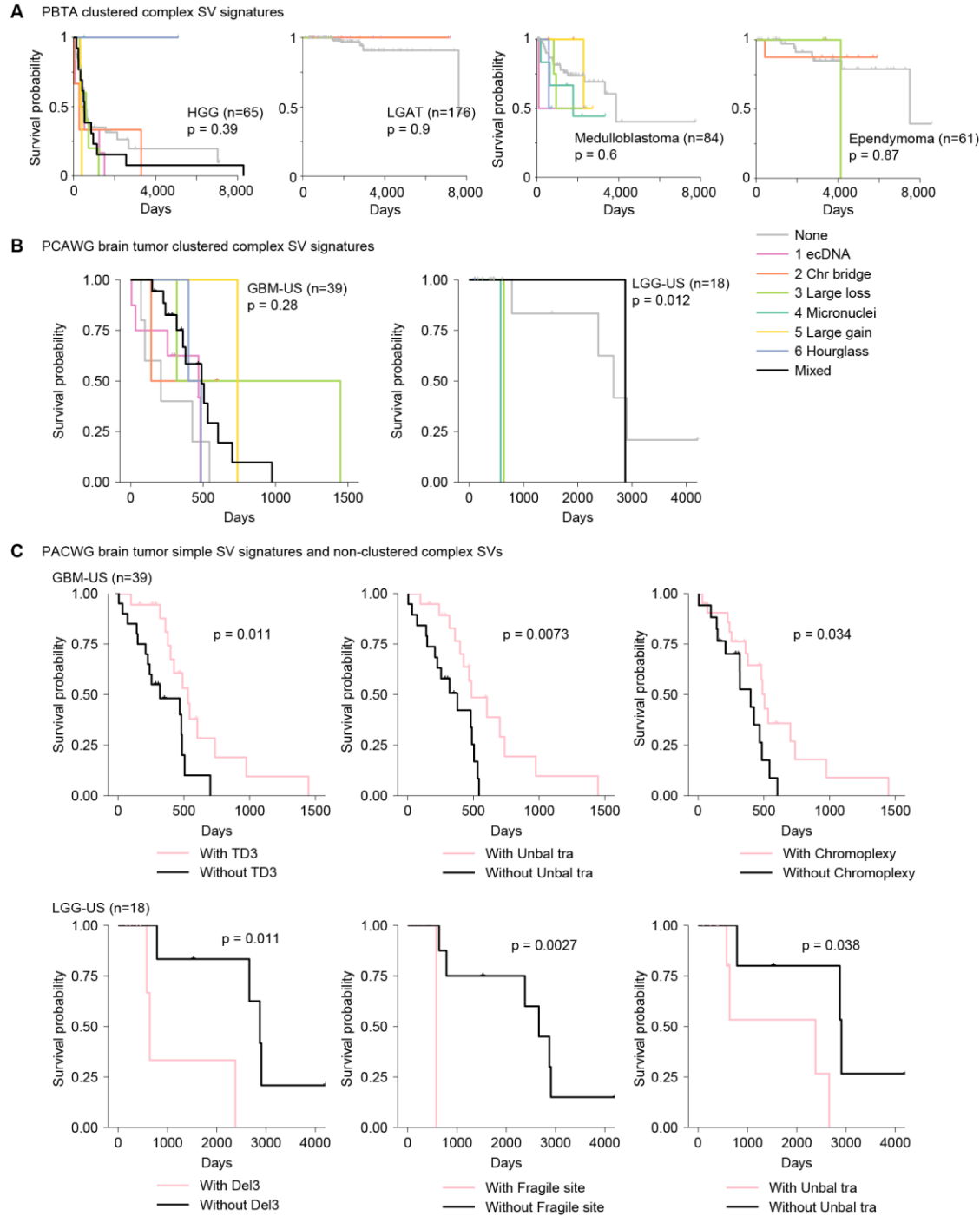


**Figure S4 Comparisons of four genomic features between observed and randomly generated SVs, related to Figure 3.** SV signatures and genomic features are listed in the x- and y-axes respectively. The violin and box plots display the distribution of scaled ranked genomic features. Colored numbers on top of each panel indicate the FDRs of the Kolmogorov-Smirnov test, as consistent with **Figure 3**.



**Figure S5 Associations between SV signatures and mutations, related to Figure 3.** (A) The associations between *TP53* mutations and complex SV signatures in 67 non-hypermuted HGGs. (B) The associations between *TP53* mutations and simple SV signatures in 67 non-hypermuted HGGs. (C) The associations between *H3F3A* mutations and complex SV signatures in 67 non-hypermuted HGGs. (D) The associations between *ATR*X mutations and simple SV signatures in 67 non-hypermuted HGGs. (E) The associations between SV signatures and germline *TP53* variants in 744 pediatric brain tumors. Tumor types are indicated by different colors on the top track. Protein-altering mutations are represented by green bars, while the presence of signatures is indicated by blue bars. The percentages reflect the proportions of samples carrying the corresponding mutations or signatures. The horizontal bars represent the frequencies of the corresponding mutations or signatures. The associations were assessed using Fisher's exact test, and the FDRs were adjusted using the Benjamini-Yekutieli method. Asterisks (\*) denote statistically significant positive correlations (FDR < 0.1) between the tested genes (A-D) or deleterious germline *TP53* variants (E) and the corresponding signatures.

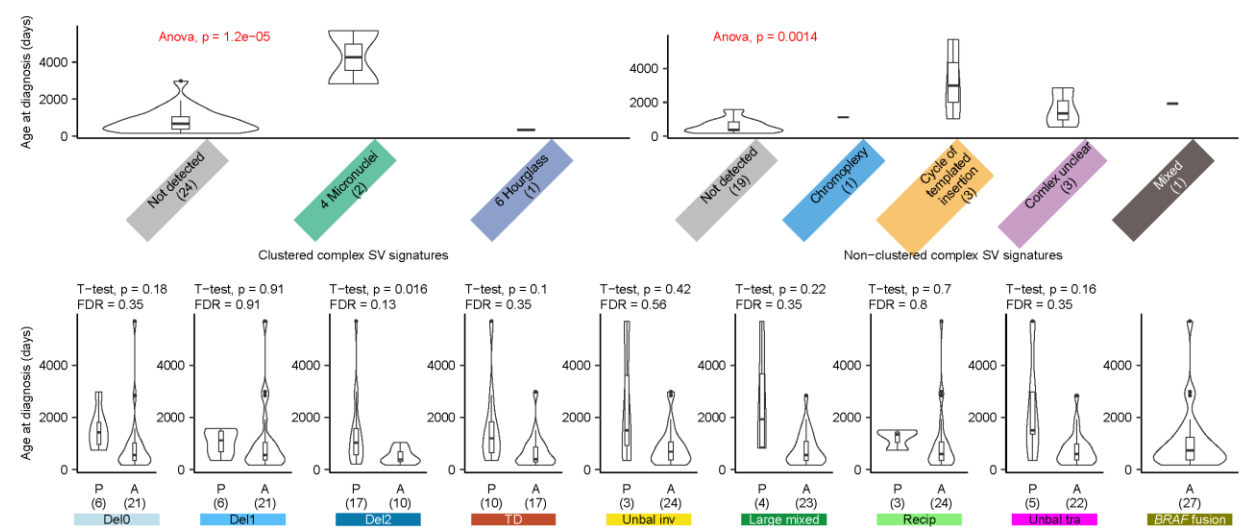




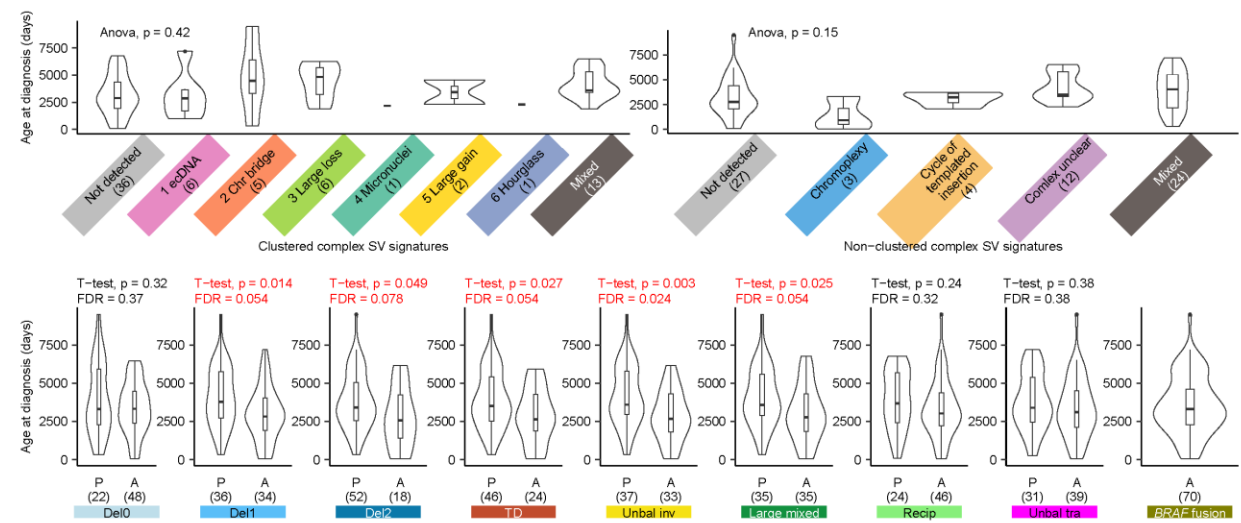
**Figure S6 Kaplan–Meier survival curves for PBTA and PCAWG brain tumors, related to Figure 7.** (A) Associations between clustered complex SV signatures and survival for PBTA four tumor types. The samples are colored by the clustered complex SV signatures. (B) Associations between clustered complex SV signatures and survival for PCAWG adult brain tumors. (C) Associations between simple SV signatures and survival for PCAWG adult brain tumors. Samples are stratified by the presence or absence of simple SV signatures. Samples are

colored by the clustered complex SV signatures (B and C). Log-rank test was used to calculate p-values. The sample sizes are provided in parentheses.

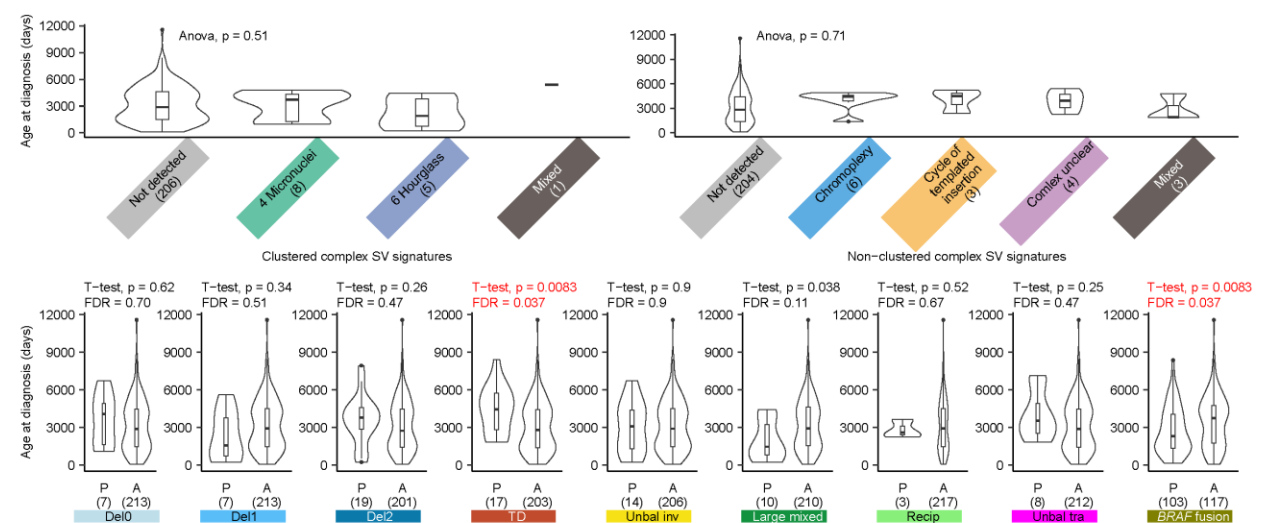
# A



# B



# C



P: Presence of specific signature  
A: Absence of specific signature

**Figure S7 Associations between SV signatures and age of diagnosis, related to Figure 7** (A) Associations between SV signatures and age of diagnosis in ATRT. (B) Associations between SV signatures and age of diagnosis in HGG. (C) Associations between SV signatures and age of diagnosis in LGAT. Samples are stratified based on clustered complex SV signatures (upper left of each panel), non-clustered complex SVs (upper right of each panel), and simple SV signatures (bottom of each panel). The differences between groups were tested using ANOVA and Student's t test. The FDRs were corrected using the Benjamini-Yekutieli method. The sample sizes of each group are indicated in parentheses.