Germline pathogenic variation impacts somatic alterations and patient outcomes in pediatric CNS tumors

Ryan J. Corbett[+1,2,3,14], Rebecca S. Kaufman[+3,4], Shelly W. McQuaid[5,6], Zalman Vaksman[3,4], Saksham Phul[1,2], Miguel A. Brown[1,2], Jennifer L. Mason[1,2,Δ], Sebastian M. Waszak[7], Bo Zhang[1,2], Chuwei Zhong[1,2], Emily Blauel[3], Heena Desai[8,9,] Ryan Hausler[8,9], Ammar S. Naqvi[1,2], Jessica M. Daggett[14], Alex Sickler[14], Evan C. Cresswell-Clay[3,4], Patricia J. Sullivan[14], Antonia Chroni[1,2,#], Zhuangzhuang Geng[1,2], Elizabeth M. Gonzalez[3], Yuankun Zhu[1,2], Allison P. Heath[1,2], Marilyn Li[10], Penn Medicine BioBank[11–], Regeneron Genetics Center[12–], Phillip B. Storm[1,2], Adam C. Resnick[1,2], Kara N. Maxwell[8,9], Kristina A. Cole[3,13], Angela J. Waanders[5,6], Miriam Bornhorst[5,6], Suzanne P. MacFarland[3,13], Jo Lynne Rokita*[,^,1,2,4,14,15], Sharon J. Diskin*[,^,3,4,9,13]

[1]Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[2]Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[3]Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[4]Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[5]Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA

[6]Division of Hematology, Oncology, Neuro-Oncology, and Stem Cell Transplantation, Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

[7]Laboratory of Computational Neuro-Oncology, Swiss Institute for Experimental Cancer Research, School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

[8]Division of Hematology/Oncology, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[9]Abramson Family Cancer Research Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[10]Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[11]Penn Medicine BioBank, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[12]Regeneron Genetics Center, Regeneron Pharmaceuticals Inc., Tarrytown, NY, USA

[13]Division of Oncology, Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[14]Center for Cancer and Immunology Research, Children's National Hospital, Washington, DC, USA

[15]Department of Pediatrics, The George Washington University School of Medicine and Health Sciences, Washington, DC, USA

[#]Present affiliation: Department of Developmental Neurobiology, St. Jude Children's Research Hospital, Memphis, TN, USA

[+]Equal contribution

[Δ]Deceased

[–]Full author list in supplement

*Senior author equal contribution

[^]Correspondence:

Sharon J. Diskin, PhD

Jo Lynne Rokita, Ph.D.

Abstract

The contribution of rare pathogenic/likely pathogenic (P/LP) germline variants to pediatric central nervous system (CNS) tumor development remains understudied. Here, we characterized the prevalence and clinical significance of germline P/LP variants in cancer predisposition genes across 830 CNS tumor patients from the Pediatric Brain Tumor Atlas (PBTA). We identified germline P/LP variants in 23.3% (193/830) of patients and the majority (137/193) lacked clinical reporting of genetic tumor syndromes. Among P/LP carriers, 34.6% had putative somatic second hits or loss of function tumor alterations. Finally, we linked pathogenic germline variation with novel somatic events and patient survival to highlight the impact of germline variation on tumorigenesis and patient outcomes.

Keywords

Pediatric brain tumors, pathogenic germline variants, loss of heterozygosity, aberrant splicing

Introduction

Central nervous system (CNS) tumors are the leading cause of childhood cancer death in the United States, with over 47,000 children and young adults diagnosed annually worldwide[1,2]. The overall prevalence of children with cancer with reported rare pathogenic or likely pathogenic (P/LP) germline variants in cancer predisposition genes (CPGs) ranges from 4-27%[3–14], although comprehensive studies in pediatric cohorts with CNS tumor malignancies are limited. There are over 20 characterized genetic cancer predisposition syndromes (CPS) observed in patients with CNS tumors (**Table S1**), and loss-of-function variants in over 30 CPGs have been linked to increased risk of distinct pediatric CNS tumor histologies and molecular subtypes[15]. Targeted germline sequencing is being increasingly implemented in clinical settings to inform clinical practice and treatment strategies[16–18]. However, the full spectrum and prevalence of pathogenic germline variants in CPGs in patients with CNS tumors remains unknown, limiting the clinical utility of targeted testing.

The integration of germline and somatic multi-omic modalities with clinical data in cancer patients has the potential to uncover novel mechanisms underlying tumor biology and reveal clinical outcomes driven by genetic predisposition[19–21]. For example, rare pathogenic germline variants in CPGs in patients with medulloblastoma have been linked to molecular subgroups, specific evolutionary trajectories, and/or biological pathway deficiencies, including chromothripsis (associated with *TP53* variants), homology-directed repair deficiency (*PALB2* and *BRCA2*), and proteome instability (*ELP1*)[22,23].

In this study, we investigate rare pathogenic germline variants in CPGs across 830 pediatric CNS tumor patients in the Pediatric Brain Tumor Atlas (PBTA), consisting of individuals enrolled on either the Children's Brain Tumor Network (CBTN) or a Pediatric Neuro-Oncology Consortium (PNOC) protocol[24]. We integrate multiple data modalities from matched tumors[25,26], including DNA and RNA sequencing, proteomics, DNA methylation, and clinical data. We observe higher prevalence of CPG germline P/LP variants than previously reported in PBTA patients, with enrichment in PBTA relative to two cancer-free control cohorts. Importantly, we

identify CPS-associated pathogenic germline variants in patients with undiagnosed tumor predisposition syndromes, underscoring the need for implementing germline testing at diagnosis. Finally, we report new associations between rare germline pathogenic CPG variants and somatic alterations in patients with high-grade glioma, *BRAF* fusion-positive low-grade gliomas, and medulloblastoma.

Results

Cohort characteristics and study workflow

For this study, we analyzed germline sequencing data from 830 pediatric CNS tumor patients with matched tumor sequencing, including 790 with whole genome sequencing (WGS) and 40 with whole exome sequencing (WXS) (**Figure 1A, Table S2**). Tumors were classified into 15 broad histology groups using the 2021 World Health Organization CNS tumor classifications[27], and molecular subtypes were assigned using annotations from the Open Pediatric Cancer Project[26] (**Figure 1B**). We further distinguished diffuse intrinsic pontine gliomas (DIPG) and diffuse midline gliomas (DMG) from other high-grade gliomas (HGG) due to their distinct anatomical, molecular, and clinical profiles. Demographic and genomic characteristics, including genetically-inferred sex and ancestry from germline data, and tumor-specific alterations, are summarized in **Figures 1C** and **S1**. Rare germline variants (allele frequency <0.1% in all non-bottleneck populations in gnomAD non-cancer population database[28]) were evaluated for pathogenicity, and we focused specifically on variants in 211 cancer predisposition genes (CPGs) defined by an expert panel of physicians and scientists (**Table S3**). Germline P/LP variants were integrated with matched tumor data, including RNA-Seq (N = 770, 92.8%), proteomics (N=192; 23.1%), and DNA methylation array data (N=730; 88.0%). Sixty-seven patients (8.1%) had documented CPS, with the most prevalent being neurofibromatosis type 1 (NF-1; N=25), followed by tuberous sclerosis (N=10), neurofibromatosis type 2 (NF-2; N=7), and Li-Fraumeni syndrome (N=7) (**Table S2**).

Rare P/LP germline variants in CPGs are prevalent in pediatric CNS tumor patients

Using AutoGVP[29], we identified 197 rare P/LP germline SNVs/InDels in 71 CPGs (**Figure 2A**, **Figure S2A**, **Table S4A-B**), with 80.2% supported by ClinVar evidence (N=158/197; **Figure 2B**). We also detected 18

germline CPG structural variants (SVs) classified as P/LP by AnnotSV[30] or ClassifyCNV[31] (**Figure 2C**, **Table S4C**). To validate these findings and further characterize their clinical relevance, we cross-referenced CPG P/LP variants with known CNS CPS (**Table S1**). We identified 104 variants in 100 patients, including 48 variants in 48 distinct patients (5.7% of cohort) without a clinically-reported CPS (**Figure 2D**, **Table S5**). These findings may reflect previously unrecognized germline contributions to tumor development and highlight the potential value of systematic germline analysis in uncovering clinically-meaningful variation beyond known diagnoses. Among patients with a clinically reported CPS in the Open Pediatric Cancer (OpenPedCan) Project[26], 86% (49/57) had a matching P/LP germline variant in our study (**Figure S2B**). In the remaining eight cases, we extended our search to include low allele frequency variants (0.08-0.20 VAF) and variants of uncertain significance (VUS), uncovering potentially deleterious variants in five patients (**Table S6**). These included four VUS: one *TP53* missense variant in an LFS patient (PT_PFP1ZVHD), and one *NF1* and two *NF2* splice donor variants in patients with neurofibromatosis. We also recovered a P/LP *NF2* variant with a VAF of 0.175, leading us to identify nine additional low-VAF P/LP variants (**Table S7**).

Overall, P/LP CPG variants were identified in 23.3% (193/830) of patients, including 30 with two variants and two with three variants: one patient with HGG and three *MSH6* P/LP variants and another with subependymal giant cell astrocytoma (SEGA, PT_JW6FBEFK) that harbored *TSC2*, *ASXL1*, and *APC* P/LP variants (**Table S4A**). The former was clinically diagnosed with constitutional mismatch repair deficiency syndrome (CMMRD; PT_3CHB9PK5), suggesting biallelic *MSH6* activation (**Figure S2C**). In the latter SEGA case, the *ASXL1* variant is unlikely to represent clonal hematopoiesis of indeterminate potential (CHIP): the patient was four years old, the variant had higher VAF in tumor vs. blood (0.60 vs. 0.41), and no other CHIP-associated variants were found. No homozygous P/LP variants were observed (**Figure S2D**); however, in addition to the CMMRD patient with compound heterozygous *MSH6* variants, a patient with medulloblastoma (MB, PT_2FK75B27) had two adjacent *PMS2* P/LP InDels on the same allele (**Figure S2E**), despite no reported CPS.

P/LP carriers were non-randomly distributed across histologies (p=5.0e-4, **Table 1),** and enriched in neurofibroma plexiform (NF; n=11/15, OR=9.5, 95% CI=2.7-41.0, p=4.6e-06) and HGG (n=26/76, OR=1.8, 95% CI=1.1-3.2, p=0.02) cohorts (**Figure 2E**, **Figure S3A**). Enrichment was also observed among tumor

molecular subtypes with known genetic predispositions (p<0.05, OR lower 95% CI>1.0; **Figure S3B, Table S8**). All SEGA patients (N=10) had P/LP variants, primarily in *TSC1* (n=3) or *TSC2* (n=6). The remaining SEGA patient had a *MUTYH* P/LP variant, not previously linked to SEGA. SHH-activated MB (SHH-MB) patients were enriched for P/LP carriers (n=12/19), including variants in *PTCH1* (n=3), *SUFU* (n=3), and *GPR161* (n=1). Among histone H3 wildtype HGGs, 64% (n=16/25) of patients were P/LP carriers, mainly in mismatch repair (MMR) genes (n=6) or *TP53* (n=4). Pineoblastoma (PB) patients were significantly enriched for P/LP variants (n=4/5; **Figure S3C**). None harbored *DICER1* variants known to be associated with PB[32], but one had a *DROSHA* and two had *ATM* P/LP variants. Demographic features were not associated with P/LP carrier status (**Table 1, Table S9**), except among mixed glioneuronal and neuronal tumors (GNTs), in which male carriers were overrepresented (n=10 male vs. 1 female, OR=9.8, 95% CI=1.3-445.5, p=0.02; **Table S9C**). In total, we identified 225 germline CPG P/LP variants (n=207 SNVs/InDels, n=18 SVs) in 23.3% of pediatric CNS tumor patients.

Germline pathogenic variation in CPGs and cancer pathways is enriched in pediatric CNS tumor patients

We compared the prevalence of P/LP germline variants in CPGs between pediatric CNS tumor patients and non-cancer cohorts from the Penn Medicine BioBank (PMBB; n=6,295) and gnomAD (n=74,023) (see **Methods**). CPG P/LP variants were significantly enriched relative to PMBB (OR=1.6, 95% CI=1.3-1.9, p=9.4e-07) and gnomAD (OR=1.7, 95% CI=1.4-2.0, p=7.3e-9) (**Figure 2F, Table S10**). Histology-specific analyses revealed significant CPG P/LP variant burden in patients with NF, HGG, and MB compared to controls (Bonferroni-adjusted p<0.05, OR lower 95% CI>1.0; **Figure S4**). Genes with the strongest burden included *NF1*, *TSC2*, *TSC1*, *TP53*, and *NF2* (**Figure S5A**), largely reflecting the composition of the cohort, which includes many patients with LGGs, NF, and HGG. We confirmed known associations, such as *PTCH1* with MB, *SMARCB1* with ATRT, and *SMARCE1* with meningioma (MNG, **Figure S5B**). Notably, we also identified a novel enrichment of *ATM* P/LP variants in patients with PB, though the small sample warrants cautious interpretation (n=2/5 patients).

Expanding beyond CPGs, we assessed rare P/LP variants across all genes and found 13 enriched KEGG pathways in the PBTA, including five involving the previously implicated CPGs (*TP53*, *TSC1*/*2*, and *NF1*). Additional enriched pathways included cell cycle regulation (G1/S) and broader cancer signaling networks such as the *MTOR* pathway (**Figure S6, Table S10C**). We further assessed DNA repair pathways[33] and observed significant enrichment of P/LP variants across 1) all DNA repair genes, 2) MMR genes, and 3) other DNA repair genes in PBTA (OR=1.4-2.7, 95% CI=1.2-5.0, p<1.3e-03; **Figure 2G, Table S10D**). This is likely driven by patients with HGG with high P/LP burden in DNA repair and MMR genes (OR=2.9-8.7, 95% CI=1.7-25.0, p<2.0e-03; **Figure 2H**), consistent with previous studiess[25,34]. In summary, germline P/LP variants in CPGs are enriched in the PBTA CNS tumor cohort, with a novel potential association between *ATM* P/LP variants and PB.

Germline CPG P/LP carriers exhibit loss of function (LOF) somatic events

We assessed all matched tumors for evidence of somatic LOF, including SNVs/InDels, copy number (CN) alterations, and loss of heterozygosity (LOH; **Figure 3A, Table S11**). Of the 225 germline P/LP variants, 195/207 (94.2%) SNVs/InDels and 17/18 (94.4%) SVs were detected in tumors. The 12 undetected SNVs/InDels had significantly lower germline VAFs compared to those detected in tumors (median VAF 0.20 vs. 0.50, p=2.9e-07), suggesting potential genetic mosaicism. Among detected variants, 72 events in 70 patients had a putative somatic second hit. These included eight patients with somatic SNVs/InDels annotated as oncogenic or likely oncogenic (O/LO) by OncoKB[35] or otherwise predicted as deleterious (see **Methods**): *NF1* mutations in four distinct P/LP carriers, an *APC* nonsense mutation and *SUFU* splice site mutation in two MB patients (PT_QEP13FH4 and PT_WWZ2QQ14R, respectively), a *MSH6* missense mutation in the HGG CMMRD patient (PT_3CHB9PK5), and a *SMARCE1* splice site mutation in a MNG case (PT_NSXP8AWV; **Figure 3B, Figure S7**). Next, we identified 63 LOH events (28.0% of all variants), with 55 (87.3%) resulting in loss of the wildtype allele (**Figure 3C**). Sixteen of these were accompanied by CN loss or deep deletions, while the remaining 47 events occurred in CN-neutral tumors (**Figure S8A**). Recurrent LOH was observed in known predisposition contexts, including: *NF1* (LGG, GNT, NF), *NF2* (MNG, SWN), *PTCH1* and *SUFU* (MB), *MSH2* (HGG), *SMARCB1* (ATRT), *TP53* (CPT, HGG), and *TSC1* (LGG)[22,36–38].

To assess the potential functional impact of germline P/LP variants on gene expression, we surveyed transcriptional and proteomic profiles in matched tumors. We observed 18 cases of gene expression loss and six cases of gene expression gain (**Figure 3D**). Variants associated with gene expression loss were primarily stop-gained (n=7), frameshift (n=3), SV deletions (n=3), or splice variants (n=2) (**Table S11**). In contrast, only two variants associated with expression gain were frameshift, while the remainder were either SV duplications (n=1) or missense variants (n=3), the latter of which are unlikely to directly inhibit transcription. Importantly, carriers with a somatic second hit (i.e., LOF SNV/InDel or wildtype LOH) exhibited significantly reduced transcript abundance relative to P/LP carriers without a second hit (p=9.7e-04; **Figure 3E**). Recurrent transcript loss in P/LP carriers versus non-carriers was observed for *MSH2*, and *TP53* (HGG), *NF1* (LGG), *SUFU* (SHH-MB), and *TSC1* and *TSC2* (SEGA) (Wilcoxon p<0.05, **Figure 3F**). Proteomics analysis of matched tumors revealed protein loss in three P/LP carriers: *TSC2* (PT_66XN3MT1, SEGA) with transcript loss (z-score=-2.43), *SLC25A13* (PT_MPRBGGEJ, LGG), and *ELP1* (PT_2JDDX6TJ, H3K27M DMG) (**Table S11**). While *ELP1* is a known SHH-MB predisposition gene[23], its role in H3K27M DMG has not been reported. In total, we identified somatic genomic or expression LOF events associated with germline P/LP variation in 78/225 (34.6%) cases.

We also analyzed tumor DNA methylation data to investigate potential epigenetic changes driven by germline P/LP variation. We found 102 hypermethylated and 317 hypomethylated probes annotated to the same gene in matched tumors of P/LP carriers relative to non-carriers (**Figure S8B, Table S12**). P/LP-associated hyper- and hypo-methylation were enriched in gene promoters (OR=1.9, 95% CI=1.2-2.9, p=5.4e-03) and intronic regions (OR=1.5, 95% CI=1.1-1.8, p=3.6E-06), respectively. Promoter hypermethylation was associated with gene expression loss (**Figure S8C**), implicating epigenetic silencing as a secondary mechanism. For example, *RECQL4* promoter hypermethylation was observed in in two unrelated P/LP carriers (LGG and GNT; **Figure S8D-E**), both with marginal gene expression loss (z-scores=-0.89, -1.18) but no somatic mutations. Notably, 14 P/LP carriers exhibited somatic promoter hypermethylation in the absence of other somatic hits or transcript loss, supporting epigenetic inactivation as an additional tumor mechanism driven by germline predisposition.

P/LP variation is significantly associated with differential somatic alternative splicing and impacts gene expression and protein function

Due to the large number of germline splice variants in the PBTA cohort (37/217) and their documented impact on pediatric CNS tumor development[22,23,39], we investigated their influence on somatic alternative splicing (AS) and subsequent functional consequences. We queried splice events (see **Methods**) in matched tumors and observed that germline P/LP splice region variation was associated with significantly increased proximal intron retention or exon skipping relative to other P/LP variants (**Figure 4A, Table S13**). Splice region variants were associated with reduced transcript and protein abundance (**Figure S9A-B**), similar to levels observed for frameshift and stop gained variants thus suggesting similar functional effects. Frameshift and stop gained P/LP variants were associated with proximal intron retention and alternative splice site usage, respectively, each significantly negatively correlated with transcript abundance (Pearson's R=-0.64, p=0.013 and R=-0.38, p=0.036), implicating aberrant splicing as a mechanism influencing gene expression in P/LP carriers (**Figure 4B**).

We identified 32 significant P/LP variant-proximal AS events in matched tumors (|percent spliced in [PSI] z-score| $\geq$ 2) including skipped exon (SE; N=25), retained intron (RI; N=4), and alternative 3' splice site usage (A3SS; N=3) events (**Figure 4C**). Intriguingly, 20 (80%) of the 24 SE events lack annotated splice junctions are therefore considered novel. Several tumors showed AS-associated gene expression loss or putative disruption of Pfam protein domains[40,41] (**Figure 4D-I, Figure S9C**). In a germline *PTCH1* splice acceptor P/LP carrier with SHH-MB (PT_NCDHZ8H8) we observed novel exon 2 skipping, which affects a conserved sterol transporter domain (**Figure 4D**). A germline *LZTR1* stop gained P/LP carrier with schwannoma (PT_BWFTKJXT) exhibited increased intron 14 retention directly upstream of this variant that was associated with significant transcript loss relative to other schwannomas (R=-0.71, p=9.5e-3; **Figure 4E,G**), suggesting a mechanism of *LZTR1* inactivation. A germline *TSC2* polypyrimidine tract P/LP carrier with SEGA exhibited exon 11 skipping resulting in a novel transcript isoform and significant TSC2 protein loss compared to other LGGs (**Figure 4F,H**). Lastly, a patient with AT/RT and a germline *LZTR1* synonymous variant exhibited somatic exon 4 skipping, predicted to disrupt the Kelch substrate recognition domain (**Figure S9D**). This variant has previously been shown to disrupt an exonic *LZTR1* splice enhancer[42]. We observe

significant LOH in this patient's tumor, implicating a novel role for *LZTR1* splice variants in AT/RT tumorigenesis. These findings support a broad role of germline splice and non-splice variants in shaping the AS landscape and its downstream functional consequences in pediatric CNS tumors.

Mismatch repair gene pathogenic variation drives distinct genetic and molecular signatures in pediatric high-grade gliomas

Given the significant P/LP variant burden in DNA repair pathways, we further explored their functional impact in tumors. We identified 105 P/LP variants across 31 DNA repair CPGs in 97 patients (11.7% of cohort; **Figure S10A**). HGG tumors were enriched in patients with P/LP variants in MMR, all repair, and other repair pathways (FDR<0.05 and OR lower CI>1 for all comparisons), while mesenchymal and other tumors were nominally enriched among base and nucleotide excision repair (BER and NER) P/LP carriers, respectively (p<0.05 and OR lower CI>1 for all comparisons; **Figure 5A**). We assessed COSMIC mutational signature data from these tumors and confirmed that H3-wildtype HGGs from HR and MMR gene P/LP carriers exhibited significantly higher SBS3 and MMR-deficiency signature exposures, respectively, relative to other HGGs (p=0.029 and 0.0033, **Figure 5B-C**, **Figure S10B, Table S14**). MMR P/LP carriers also harbored significantly higher tumor mutation burden (TMB) compared to non-carriers (p=1.2e-04), and, consistent with previous findings in MMR-deficient HGG[43], 4/6 tumors were considered hypermutant (>10 mutations/Mb; **Figure 5D**). Mesenchymal tumors with BER P/LP variants exhibited increased SBS30 exposure weights relative to non-carriers (p=8.6e-4, **Figure 5E).** We observed significantly lower genome-wide mean methylation beta values in H3-wildtype HGGs from MMR P/LP carriers compared to non-carriers, with the CMMRD patient with 3 *MSH6* germline P/LP variants exhibiting the lowest global methylation (p=0.003, **Figure 5F**). Since the tumors of MMR P/LP carriers were hypomethylated and hypermutated, we asked whether TMB and global methylation were broadly correlated in HGGs. Indeed, global methylation was significantly negatively correlated with TMB exclusively in H3 wildtype HGG tumors (Pearson's R=-0.72, p=7.6e-06), but not in other histologies or HGG subtypes (**Figure S10C-D, Table S15**). Clustering of HGG tumors by beta values at the 20,000 most variable promoter CpGs revealed a distinct cluster of heterozygous MMR P/LP carriers, while the CMMRD patient sample clustered separately (**Figure 5G**). Finally, two MMR gene P/LP carriers exhibited somatic promoter

hypermethylation of the same gene, suggestive of epigenetic silencing associated with P/LP variation despite observed global hypomethylation (**Figure 5H-I**). These findings highlight distinct DNA methylation and mutational signatures associated with germline MMR P/LP variation in pediatric HGGs.

CPG P/LP carriers exhibit distinct survival outcomes

CPG P/LP variants have previously been linked to poorer overall survival (OS) in pediatric cancers including neuroblastoma and leukemia[22,44–46]. We therefore evaluated whether P/LP carriers in the PBTA exhibited distinct survival outcomes (**Figure 6A, Table S16**). We found that among all *KIAA1549*::*BRAF* fusion-positive LGGs, P/LP carriers had significantly worse event-free survival (EFS) versus non-carriers in univariate (p=0.031, **Figure 6B**) and multivariate cox proportional hazards models accounting for extent of tumor resection, predicted sex, and age at diagnosis (HR=2.67, p=0.03; **Figure 6C**). Most P/LP carriers with *KIAA1549::BRAF* fusion positive LGG (N=13/20) harbored variants in DNA repair genes, and this cohort exhibited the second greatest excess DNA repair P/LP variant burden relative to controls; this represent a novel finding, but was not significant after multiple test correction (OR=1.8, 95% CI=1.1-3.0, p=0.01; **Figure S11A**).

Among patients with MB, CPG P/LP carriers exhibited significantly better EFS and OS relative to non-carriers (p<0.01 for both EFS and OS in univariate and multivariate models; **Figure 6D, Figure S11B-C**). We evaluated additional molecular and clinical data in patients with MB, and observed that P/LP carriers exhibited significantly lower metastasis rates compared to non-carriers (OR=0.31, 95% CI=0.09-0.91, p=0.04; **Figure 6E**), and metastasis was associated with significantly worse OS (HR=7.88, 95% CI=2.1-29.1, p=4.2e-03; **Figure S11D**). We identified significantly lower TMB in WNT- and SHH-MB tumors of P/LP carriers compared to non-carriers, which has not been previously reported (**Figure 6F**). In the case of SHH-MB tumors, this may be explained in part by earlier ages at diagnosis in P/LP carriers (**Figure S11E**). We further delineated SHH-MB tumors and observed that infant SHH (β and γ) tumors were enriched in P/LP carriers (N=6, OR=1.7, 95% CI=1.4-1395.7, p=0.01; **Figure 6G**), while child-adult SHH (α and δ) tumors were enriched in non-carriers and exhibited worse OS and EFS **(Figure S11F)**. Similarly, we only identified P/LP carriers with subgroup 7 (N=3/19) and 8 (N=3/15) Group 4 MBs among this subtype, and these patients had significantly better EFS

and OS than non-carriers with subgroup 5 tumors, though we acknowledge the low numbers in these groups (**Figure 6H, Figure S11G**). In summary, P/LP carriers with MB Group 4 subgroups 7 or 8, infant-type SHH MB, and/or MB with decreased metastasis confer better OS.

Discussion

We identified germline CPG P/LP variants in 193/830 (23.3%) patients, a frequency that is within the range reported in previous germline studies (4-27%)[3–5,7,8,11–13,22]. While the majority of CPG P/LP SNVs/InDels called by AutoGVP utilized ClinVar evidence, we identified 42 additional P/LP variants through AutoGVP's modified InterVar classification, 18 P/LP SVs and one deep intronic *NF1* variant previously classified as P/LP[47], highlighting the importance of leveraging multiple approaches to characterize the full spectrum of disease-causing germline variation in pediatric cancers. We observed high concordance between clinically-reported CPSs and prevalence of associated CPG P/LP variants, and identified several VUS in patients lacking P/LP variants that warrant further study into their pathogenicity. The majority of identified germline P/LP variants were absent from clinical records, including a high proportion of variants in genes linked to CPSs. Several factors may contribute to this, such as the extended timeframe of sample collection (spanning over 13 years), the non-routine nature of germline testing, and a negative family history due to incomplete penetrance or *de novo* variants[48]. These findings underscore the importance of increasing routine germline testing in pediatric CNS tumor patients to better guide monitoring and treatment strategies even in the absence of family history. For example, we identified several patients with germline P/LP *TP53* or MMR gene variants with no clinically reported CPS; the former would benefit from lifetime monitoring for early detection of secondary cancers as is routine for patients with LFS, and the latter could receive immunotherapies given the prevalence of chemoresistant tumors[49]. While the majority of histology-specific P/LP CPGs have been previously described, we observed that 2 of 5 patients with PB harbored an *ATM* germline P/LP variant. To our knowledge, *ATM* has not been implicated in development of PB, although future studies are needed to assess their prevalence in a larger cohort.

We detected a putative somatic second hit (LOF SNV/InDel, LOH) in tumors of 27.5% (N=53/193) of P/LP carriers, suggesting that observed pathogenic germline variation in these patients is indeed contributing

to tumorigenesis. While representing roughly one-third of cases, and within the range of second hit frequencies reported in previous pediatric cancer studies (10-55%)[3–5,8,44], there may be additional somatic inactivating mutations in functionally-related genes and/or pathways, and/or haploinsufficiency alone may initiate tumorigenesis[50] in some cases. Furthermore, a small percentage of germline CPG P/LP variants (13/225) were not detected in matched tumors, and these exhibited low median VAF (0.2) relative to those variants found in tumors (VAF=0.5). Such observations may suggest that these low VAF variants represent genetic mosaicism and may not contribute to tumorigenesis; however, it is possible that these variants play critical roles in tumor initiation and be lost at later stages of tumor evolution, as has been reported in glioma and medulloblastoma initiating events[51,52]. Somatic SNVs/InDels in the same P/LP gene were rare, while CN-neutral LOH resulting in loss of wildtype allele was pervasive, particularly in *NF1*, *TP53*, *PTCH1*, *MSH2*, and *TSC2*.

We report transcriptomic, regulatory, and proteomic somatic alterations indicative of gene inactivation, many in cases in which a tumor DNA second hit was not observed. The majority of somatic differential gene expression in CPG P/LP carriers indicated loss of transcript in the same gene, though there were rare cases of gene expression gain including in a *CHEK2* P/LP carrier with a large germline duplication event. For the remaining cases, increased gene expression associated with P/LP germline variation may be due in part to compensatory upregulation as a result of loss of function at the protein level; however, we did not have proteomics data from any of these samples to confirm protein loss. We further identified splice region P/LP variants driving somatic aberrant splicing and disruption of protein domains with subsequent gene expression and protein loss. Beyond splice region variants, we report cases of frameshift or stop gained variant-associated AS, indicating that splicing may contribute to LOF driven by these variant classifications as has been reported previously[53]. Our work emphasizes the need for further investigation into the contributions of AS to pediatric CNS tumor development and progression. Lastly, we observed global functional consequences of germline MMR gene P/LP variation in patients with histone H3-wildtype HGG, including increased MMR-deficiency mutational signatures, TMB, and global hypomethylation, the latter of which has not been reported previously. Strikingly, we found increased TMB to be associated with global hypomethylation exclusively in H3-wildtype HGG, but not in other HGG subtypes nor in other tumor histologies.

Germline CPG pathogenic variation was associated with distinct clinical outcomes in a subset of patients. CPG P/LP carriers with *KIAA1549::BRAF* fusion-positive LGGs exhibited worse EFS compared to non-carriers. Interestingly, this cohort disproportionately harbored DNA repair gene variants, which were also nominally enriched among *KIAA1549::BRAF*-fusion positive LGGs relative to control populations. Future research should seek to assess the contribution of DNA repair gene germline pathogenic variation to *KIAA1549::BRAF* LGG tumorigenesis. Conversely, CPG P/LP carriers with MB exhibited better survival relative to non-carriers, which may be explained by lower metastasis rates and TMB, earlier ages at diagnosis, and clinically favorable methylation subgroups associated with P/LP carriers in this MB cohort. While inactivation of known MB predisposition genes has previously been associated with increased metastasis risk[54,55], this has not been explored in the context of germline variation to our knowledge. Future studies with a larger cohort are required to assess unique germline variant-associated tumor characteristics that may contribute to clinical outcomes in MB. In summary, our study has identified new functional links between germline variation, tumor molecular features, and clinical outcomes in pediatric CNS tumor patients, highlighting the importance of routine germline testing at diagnosis. Continued understanding of germline susceptibility and its influence on tumorigenesis will aid in identification of patients that may benefit from genetic counseling, surveillance, altered treatment regimens, and ultimately clinical outcomes.

Methods

Pediatric CNS tumor/normal samples and germline SNV/indel calling

The pediatric brain tumor cohort used in this study is composed of patients from the Children's Brain Tumor Network (CBTN, https://cbtn.org/, n=775) and the Pediatric Neuro-Oncology Consortium (PNOC, https://pnoc.us/, n=55). The CBTN is a collaborative international and multi-institutional research and biorepository initiative dedicated to the study of childhood brain tumors[24]. PNOC is an international clinical trials consortium dedicated to bringing new therapies to children and young adults with brain tumors. Together, these consortia have contributed samples to the Childhood Cancer Data Initiative (dbGaP phs002517.v4.p2). Germline variant calls from whole genome sequencing (WGS) data for paired tumor (~60X) and normal

peripheral blood (~30X)[25,26] and from whole exome sequencing (WXS) data for paired tumor (~470X) and normal peripheral blood (~308X) were obtained through an approved dbGaP data access request. Matched tumor WGS (~60X), WXS (~470X), RNA-Seq, methylation, and/or proteomics somatic outputs as well as clinical and demographic information (e.g. survival, metastasis, race, ethnicity, and clinically-reported predisposition syndrome) from the OpenPedCan[26] v15 data release[56] were utilized in this study. The list of genes associated with central nervous system (CNS) tumor predisposition syndromes was compiled using two key sources: Chapter 14 of the 2021 World Health Organization (WHO) Classification of Tumors of the Central Nervous System[57] and the recent publication describing *DROSHA*-associated tumor predisposition[58]. These references collectively informed the gene set used to identify P/LP germline variants relevant to CNS tumor syndromes in our cohort. We also performed pathology and electronic health record review (CHOP IRB 09-007316) for all patients with P/LP variants in CNS tumor predisposition associated CPGs to identify whether a tumor syndrome was clinically reported.

Germline SNV/indel calling

Briefly, paired-end WGS and WXS reads were aligned to the version 38 patch release 12 of the Homo sapiens reference genome using BWA-MEM[59]. The Broad Institute's Best Practices[60] were used to process Binary Alignment/Map files (BAMs) in preparation for variant discovery. Duplicate alignments were marked using SAMBLASTER[61], and merged and sorted BAMs using Sambamba[62]. The BaseRecalibrator submodule of the Broad's Genome Analysis Tool Kit (GATK)[63] the GATKHaplotypeCaller[64] submodule to generate a genomic variant call format (GVCF) file. This file was used as the basis for germline calling. Germline haplotype calling was performed using the GATK Joint Genotyping Workflow on individual gVCFs from the germline sample alignment workflow. The GATK generic hard filter suggestion was applied to the VCF (SNPs only), with an additional requirement of 10 reads minimum depth per variant. This Kids First workflow can be found at https://github.com/kids-first/kf-germline-workflow.

Cancer predisposition gene list curation

Cancer predisposition genes (CPGs) were selected based on review of the Children's Hospital of Philadelphia's (CHOP) Division of Genomic Diagnostics' hereditary cancer panel (www.testmenu.com/chop/Tests/786446), public databases including the Online Mendelian Inheritance in Man (www.omim.org) and the National Institute of Health's Genetics Home Reference (www.medlineplus.gov/genetics), and medical literature. We added *GJB2* due to its observed enrichment in neuroblastoma compared to tumor-free control cohorts[65] (**Table S3**).

Germline variant annotation and assessment of variant pathogenicity

Resulting variants were annotated using the Kids First Data Resource Center (KFDRC) Germline SNV Annotation Workflow and added the following annotations: ENSEMBL 105, ClinVar (2022May07), and InterVar (https://github.com/kids-first/kf-germline-workflow/blob/master/docs/GERMLINE_SNV_ANNOT_README.md). Germline variants in 211 cancer predisposition genes were further analyzed.

Pathogenicity was assessed for filtered variants *in silico* using AutoGVP to evaluate ClinVar (2022May07) and a modified execution of InterVar[29]. We applied a pathogenicity pre-processing workflow on VEP-annotated VCFs that encompasses variant annotation with ANNOVAR (database pulled on 2024May06), InterVar v2.2.1 variant classification, and AutoPVS1 v2.0.0 variant scoring. Variants with read depth coverage $\geq$ 10, variant allele fraction $\geq$ 0.20, and observed in <0.1% across non-bottleneck gnomAD v3.1.1 non-cancer populations (African/African American, Admixed American, non-Finnish European, South Asian, East Asian) were retained for variant classification. Pathogenicity assessment was performed using AutoGVP as previously described[29]. Briefly, we pulled classifications for rare variants annotated in the ClinVar database with $\geq$2 stars or with 1 star and no conflicting submissions. All variants with conflicting classifications in ClinVar were resolved by first filtering for submissions that have been associated with MedGen disease concept IDs, and taking the single submission or consensus submission if available. Remaining variants classifications were resolved by taking the classification from the submission at the last date evaluated. For remaining variants without ClinVar annotation, we adjusted PP5 based on this modified InterVar assessment and corrected PVS1 using AutoPVS1. Variants were assigned as pathogenic (P), likely pathogenic (LP), benign (B), likely benign (LB), or unknown significance (VUS) by first considering ClinVar results and then the modified InterVar output.

This approach was applied to pediatric CNS cases and non-cancer control samples (PMBB and gnomAD 3.1.1). For variants with conflicting submissions in ClinVar that were resolved as non-P/LP, we reviewed cases for which there was a P or LP submission that provided functional evidence (i.e., experimental evidence showing this variant is associated with loss of function), and upgraded the classification accordingly when such evidence was sufficient.

All cases of proximal adjacent InDels in a single patient and InDels $\geq$ 20 bp were manually reviewed in Integrated Genomics Viewer (IGV) and removed if variants were not present. We also queried samples for intronic NF1 variants identified in Koczkowska et al. 2023[47] that were shown to associate with alternative splicing events and classified these as LP.

Cancer-free control cohorts

We utilized cancer-free control cohort data from the Penn Medicine BioBank (PMBB) and Genome Aggregation Database (gnomAD v3.1) to calculate relative enrichment of P/LP variants in the PBTA cohort. The PMBB is a precision medicine cohort that began in 2004, enrolling patients from the University of Pennsylvania Health System who consented to biospecimen collection and linkage of their samples to electronic health record (EHR) data. Whole-exome sequencing of participant DNA extracted from stored buffy coats was performed by the Regeneron Genetics Center (Tarrytown, NY). For this study, we leveraged data from 6,295 cancer- and tumor-free individuals classified based on the absence of relevant International Classification of Diseases (ICD)-9/10 codes in their EHRs, as previously described[65].

Germline structural variant calling and pathogenicity assessment

We applied the KFDRC Germline Structural Variant Caller Workflow to generate SV pathogenicity calls from normal sequencing data (https://github.com/kids-first/kf-germline-workflow/blob/master/docs/GERMLINE_SV_README.md). Briefly, Manta v1.6.0[66] was run on all aligned paired-end sequencing reads to call SVs, and AnnotSV v3.1.1[30] was run on Manta output to annotate SV pathogenicity. CNVnator[67] was run on aligned reads to call CNVs (https://github.com/kids-first/kf-germline-workflow/blob/master/docs/GERMLINE_CNV_README.md), and ClassifyCNV[31] was run on CNVnator output

to annotate CNV pathogenicity. We filtered out any P/LP SVs or CNVs present in >1% of the PBTA cohort, and any that overlapped with SV or CNV annotated in gnomAD at an allele frequency > 0.01% in a non-bottleneck population.

Cohort demographics and clinical characteristics

Genetic sex prediction was performed as previously described[25]. Briefly, we calculated the fraction of total normalized X and Y chromosome reads that were attributed to the Y chromosome. Values < 0.2 were annotated as female, values > 0.4 were predicted male, and samples with values between 0.2–0.4, inclusive, were marked as unknown. Genetic ancestry prediction was performed on all patients with normal WGS data (n=790) as previously described[68]. Differences in demographic and clinical characteristics between subjects with and without germline P/LP variants in CPGs were assessed across the cohort and within cancer histologies. Fisher's exact tests were used to compare categorical data variables (sex, race, ethnicity, predicted genetic ancestry) and Mann Whitney-U tests were used to compare numerical data variables (age at diagnosis, overall and event-free survival, tumor mutation burden). A two-sided Fisher's exact test or Mann Whitney-U test p-value <0.05 was considered significant.

CPG P/LP Enrichment

The gnomAD v3.1.1 and PMBB cancer-free control cohorts were run through the same pathogenicity preprocessing and assessment pipelines. The PMBB cohort consisted of tumor-free patients without a family history of cancer, as defined previously[44]. Variants identified in the PMBB cohort were subjected to the same manual curation used for the PBTA cohort. Enrichment of P/LP germline variants across all CPGs and within individual CPGs in the PBTA cohort vs. control cohorts was performed using Fisher's exact tests. The same statistical framework was applied to assess enrichment of P/LP germline variants among genes in KEGG pathways and DNA repair pathways as defined in Knijnenburg et al 2018[33], regardless of whether the gene was also defined as a CPG. Bonferroni adjustment was performed for gene and pathway analyses to correct for multiple tests. Enrichment of CPGs and pathway P/LP variants was also run within tumor histology cohorts using the same methodology.

Somatic second hit analyses

Somatic SNVs, CNVs, and SVs were obtained from OpenPedCan release v15[25,26]. Somatic mutations in CPGs were annotated with oncoKB to identify those classified as oncogenic or likely oncogenic (O/LO)[35]. We defined somatic SNV/indel second hits as those defined as those annotated as O/LO or otherwise classified as "probably_damaging" or "possibly_damaging" by PolyPhen or "deleterious" by SIFT. Copy number variants (CNVs) were identified in tumor samples as described in OpenPedCan[25] based on WXS CNVkit[69] calls or WGS consensus calls among Control-FREEC[70], CNVkit[69], GATK[63] and/or MantaSV[66]. We assessed LOH in matched tumors using two methods: 1) comparison of VAFs in tumor vs. germline, and 2) estimating tumor allele copy number using CNVkit[69], the latter of which accounts for low tumor purity which limited detection of LOH by VAF in some cases. We created the publicly-available tool AlleleCouNT (https://github.com/d3b-center/AlleleCouNT)[71] to obtain germline and tumor allele counts of each CPG P/LP variant in P/LP carriers, and calculated odds ratios of variant frequencies from 2x2 contingency tables using Fisher's exact tests. Loss of heterozygosity (LOH) affecting the wildtype allele was defined by one of the following criteria: 1) tumor VAF > 0.75, odds ratio > 1 and Fisher's exact $p < 0.05$, or 2) CNVkit copy number of wild type allele=0. LOH affecting the variant allele was defined by the following criteria: 1) tumor VAF < 0.25, odds ratio < 1 and Fisher's exact $p < 0.05$, or 2) CNVkit copy number of variant allele=0. Variant LOH score was calculated as tumor VAF - germline VAF. We calculated gene-level LOH score using the following equation across all rare variants showing evidence of heterozygosity in the germline:

$$gene_LOH_score = \sum |variant_AF_tumor - 0.5|$$

Genes were defined as exhibiting significant LOH if gene_LOH_score $\geq$ 0.25.

Gene Expression

RNA-seq data were obtained from the OpenPedCan release v15[26,72]. Z-scores were calculated using the formula $z=(x - \mu)=\sigma$ where x is the sample TPM for a given gene, μ is the mean gene TPM across samples, and σ is the TPM standard deviation. We considered a gene as differentially expressed relative to other samples of the same tumor histology when |TPM z-score| $\geq$ 2.

Alternative Splicing

We queried replicate Multivariate Analysis of Transcript splicing (rMATs) files from OpenPedCan release v15[26,72] for the most proximal splice event of each case type (skipped exon [SE], retained intron [RI], and alternative 3' and 5' splice site [A3SS and A5SS, respectively] to each CPG P/LP variant, and calculated percent spliced in (PSI) value z-scores for these events in each P/LP carrier using the same method we applied to TPM values. Since rMATs only reports splice events in a sample that are either supported by 1) the provided gene annotation file or 2) sample RNA-seq reads, we classified each splice event as known (i.e., annotated in a transcript) or novel (i.e., not found in any annotated transcript). In the case of novel SE and RI events that were not reported in a given sample, we set PSI equal to 1 for SE events (indicating no evidence for exon skipping) and 0 for RI events (indicating no evidence for intron retention) to obtain the most accurate assessment of differential alternative splicing in P/LP carriers versus non-carriers. In cases where multiple exon skipping events were reported for the same exon but with different flanking exons, we retained the event with the greatest PSI difference in the P/LP carrier versus non-carriers. Significant germline P/LP variant-associated differential splicing events were defined as events with |PSI z-score| $\geq$ 2 and |PSI sample - mean PSI| > 0.05, where the P/LP variant was <250bp from the splice junction.

DNA Methylation

Matched tumor Illumina 450K or EPIC 850K Infinium Human Methylation Bead Array data from the PBTA-CBTN Open Access CAVATICA project (https://cavatica.sbgenomics.com/u/cavatica/pbta-cbttc/files-premium?path=CBTN-Methylation) were processed to obtain beta values as previously described in the OpenPedCan project[26]. Briefly, IDATs were processed using the minfi Bioconductor package, funnorm normalization, and probes containing common SNPs in dbSNP (minor allele frequency > 1%) were removed. We calculated detection p-values using minfi and filtered out probes with signal that was indistinguishable from background (p<0.05). Separately as a part of OpenPedCan, the IDATs were run through the DKFZ brain classifier version 12.6[73] and/or the NIH classifier v2 to obtain tumor classifications. To determine if P/LP carriers exhibited somatic differential methylation of affected CPG, we calculated beta value z-scores of each

probe associated with that CPG in the same manner as described for gene expression. Significant differential methylation was defined as |sample beta value z-score| $\geq$ 2 and |sample beta value - mean beta value| > 0.05. We also calculated mean beta values across all measured probes for each patient as an estimate of relative global methylation across individuals.

Proteomics

Proteomics data from the Clinical Proteomic Tumor Analysis Consortium (CPTAC PBTA, pediatric, Proteomic Data Commons PDC000180) and Project HOPE (adolescent and young adult HGG) were obtained from OpenPedCan release v15[26,74]. Sample-level protein abundance z-scores in P/LP carriers were calculated relative to other samples of the same histology were calculated as previously described. Significant differential protein abundance was defined as |sample protein abundance z-score| $\geq$ 2.

Tumor mutational signatures analysis

We utilized mutational signatures generated from the v15 release of OpenPedCan[26]. Mutational signature analysis was performed using the deconstructSigs[75] R package function `whichSignatures()`, input consensus SNV file, and COSMIC v3.3 mutational signatures database as described in the Open Pediatric Cancer Project[26]. Overall MMR deficiency exposure was calculated as the sum of the following SBS signatures: SBS6, SBS14, SBS15, SBS20, SBS21, SBS26, and SBS44. We performed Mann Whitney-U tests to assess differences in mutational signature exposure weights between samples with and without germline P/LP variants in DNA repair genes. Exposure weight differences were plotted using ggplot2 `geom_bar()` function[76]. Differences in mutational signatures exposure weights in >2 groups were assessed using one-way analysis of variance (ANOVA) and post-hoc pairwise comparisons, and were plotted using ggplot2 `geom_violin()` function[76].

Tumor Mutation Burden

We utilized tumor mutation burden (TMB) calculations from the v15 release of the Open Pediatric Cancer Project[26]. Briefly, coding-only TMB was computed using high- or moderate-impact nonsynonymous

variants from consensus SNV/MNV calls, i.e. those detected by at least two of four variant callers or hotspot mutations from one variant caller. For WGS and WXS samples, only variants within coding regions defined by GENCODE v39 were counted, and TMB was calculated by dividing these counts by the size of the genome or exome coding regions surveyed by the variant callers.

Survival analyses

We performed Kaplan-Meier analyses of overall and event-free survival (OS and EFS, respectively) to compare outcomes of patients with and without germline P/LP variants in CPGs. Patient events that were included in EFS calculations were as follows: tumor progression, recurrence, and second malignancy, and death due to disease. We generated Cox proportional-hazards regression models to identify variables that were predictive of outcome. Variables considered included presence of germline P/LP variant in a CPG, molecular subtype, extent of tumor resection, reported gender, and age at diagnosis. The cohort analyzed included all those for whom data were available for all variables in the model (n=654 for OS, n=652 for EFS). Survival analyses were run using the survival R package[77,78].

Inclusion and Ethics Statement

Human data analyzed for this manuscript was obtained through an approved data use agreement and utilized with patient consent for general research use.

Conflicts of Interest

AJW is on the Scientific Advisory Boards for Alexion and DayOne. AJW has served as a consultant for Ipsen. MB is a consultant for Alexion and Springworks.

Data availability

All pediatric brain tumor raw data are available upon request from the database of Genotypes and Phenotypes (dbGaP), accession number phs002517.v4.p2. Proteomics raw data from the Clinical Proteomic Tumor Analysis Consortium (CPTAC PBTA, pediatric) are openly available from the Proteomic Data Commons

PDC000180. Methylation array data are openly available from the PBTA-CBTN Open Access CAVATICA project (https://cavatica.sbgenomics.com/u/cavatica/pbta-cbttc/files-premium?path=CBTN-Methylation). All processed somatic data used in this study were derived from the OpenPedCan project[26] v15 data release[79], which can be found through https://github.com/d3b-center/OpenPedCan-analysis or CAVATICA https://cavatica.sbgenomics.com/u/cavatica/opentarget.

Code availability

All original code and manuscript figures and tables are available at https://github.com/diskin-lab-chop/pbta-germline-somatic.

Author Contributions

Conceptualization: SJD, SM, AJW, MB, KAC, RJC, JLR, RSK

Methodology: RJC, RSK, ZV, MAB, SP, JLR, SJD, ASN, SW, JMD, AS

Software: MAB, SP, JMD, AS

Validation: ASN, AC, RSK, RJC, MAB, ZG, ECCC, PJS, JLR, SJD

Formal Analysis: RJC, RSK, MAB, SP, BZ, CZ, JLR

Investigation: RJC, RSK, SJD, JLR, MB, SM, SWM, KAC, ML, SW, AJW

Resources: PBS, ACR, SJD, JLR, AJW

Data Curation: JLR, JLM, RJC, ZG, EB

Writing - Original Draft: RJC, JLR, SWM

Writing - Review & Editing: JLR, SJD, AJW, SW, SP, MB, SM, RJC, KAC, EG

Visualization: RSK, RJC, JLR, ASN

Supervision: JLR, SJD

Project Administration: JLR, SJD

Funding Acquisition: SJD, JLR, ACR

Acknowledgements

The authors thank the patients and families for donating tissue to the pediatric oncology community and the Children's Brain Tumor Network and Pacific Pediatric Neuro-Oncology Consortium for generating the sequencing data used for this work. This work was funded in part by National Institutes of Health (NIH) grants R03CA287169 (SJD, JLR), R03OD036498 (JLR), U24OD038422 (SJD, JLR), R03CA230366 (SJD), the NIH Kids First Cloud Credits Program (SJD, JLR), the Children's Brain Tumor Network, the Chad Tough Foundation, and the anonymous private investors to the Children's National Hospital Brain Tumor Institute.

References

1. SEER Program (National Cancer Institute (U.S.)). *SEER, Surveillance, Epidemiology, and End Results Program*. (National Institutes of Health, National Cancer Institute, 2000).
2. Lu, V. M., Elarjani, T. & Niazi, T. N. Global, regional, and national incidence and mortality trends in pediatric central nervous system tumors over the past 2 decades. *World Neurosurg.* **179**, e568–e574 (2023).
3. Fiala, E. M. *et al.* Prospective pan-cancer germline testing using MSK-IMPACT informs clinical translation in 751 patients with pediatric solid tumors. *Nat Cancer* **2**, 357–365 (2021).
4. Zhang, J. *et al.* Germline Mutations in Predisposition Genes in Pediatric Cancer. *N. Engl. J. Med.* **373**, 2336–2346 (2015).
5. Wong, M. *et al.* Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer. *Nat. Med.* **26**, 1742–1753 (2020).
6. Mody, R. J. *et al.* Integrative Clinical Sequencing in the Management of Refractory or Relapsed Cancer in Youth. *JAMA* **314**, 913–925 (2015).
7. Oberg, J. A. *et al.* Implementation of next generation sequencing into pediatric hematology-oncology practice: moving beyond actionable alterations. *Genome Med.* **8**, 133 (2016).

8. Parsons, D. W. *et al.* Diagnostic Yield of Clinical Tumor and Germline Whole-Exome Sequencing for Children With Solid Tumors. *JAMA Oncol* **2**, 616–624 (2016).
9. Newman, S. *et al.* Genomes for kids: The scope of pathogenic mutations in pediatric cancer revealed by comprehensive DNA and RNA sequencing. *Cancer Discov.* **11**, 3008–3027 (2021).
10. Nakano, Y., Rabinowicz, R. & Malkin, D. Genetic predisposition to cancers in children and adolescents. *Curr. Opin. Pediatr.* **35**, 55–62 (2023).
11. Akhavanfard, S., Padmanabhan, R., Yehia, L., Cheng, F. & Eng, C. Comprehensive germline genomic profiles of children, adolescents and young adults with solid tumors. *Nat. Commun.* **11**, 2206 (2020).
12. Mardis, E. R. *et al.* Germline susceptibility from broad genomic profiling of pediatric brain cancers. *Neurooncol. Adv.* **6**, vdae099 (2024).
13. Kim, J. *et al.* Frequency of pathogenic germline variants in cancer-susceptibility genes in the Childhood Cancer Survivor Study. *JNCI Cancer Spectr.* **5**, kab007 (2021).
14. Jovanović, A. *et al.* Germline variants in cancer predisposition genes in pediatric patients with central nervous system tumors. *Int. J. Mol. Sci.* **24**, (2023).
15. Muskens, I. S. *et al.* Germline genetic landscape of pediatric central nervous system tumors. *Neuro. Oncol.* **21**, 1376–1388 (2019).
16. Hottinger, A. F. & Khakoo, Y. Update on the management of familial central nervous system tumor syndromes. *Curr. Neurol. Neurosci. Rep.* **7**, 200–207 (2007).
17. Rahman, N. Realizing the promise of cancer predisposition genes. *Nature* **505**, 302–308 (2014).
18. Shukla, N. *et al.* Feasibility of whole genome and transcriptome profiling in pediatric and young adult cancers. *Nat. Commun.* **13**, 2485 (2022).
19. Qing, T. *et al.* Germline variant burden in cancer genes correlates with age at diagnosis and somatic mutation burden. *Nat. Commun.* **11**, 2438 (2020).
20. Chatrath, A., Ratan, A. & Dutta, A. Germline variants that affect tumor progression. *Trends Genet.* **37**, 433–443 (2021).
21. Chatrath, A. *et al.* The pan-cancer landscape of prognostic germline variants in 10,582 patients. *Genome Med.* **12**, 15 (2020).

22. Waszak, S. M. *et al.* Spectrum and prevalence of genetic predisposition in medulloblastoma: a retrospective genetic study and prospective validation in a clinical trial cohort. *Lancet Oncol.* **19**, 785–798 (2018).
23. Waszak, S. M. *et al.* Germline Elongator mutations in Sonic Hedgehog medulloblastoma. *Nature* **580**, 396–401 (2020).
24. Lilly, J. V. *et al.* The children's brain tumor network (CBTN) - Accelerating research in pediatric central nervous system tumors through collaboration and open science. *Neoplasia* **35**, 100846 (2023).
25. Shapiro, J. A., Gaonkar, K. S., Spielman, S. J. & Savonen, C. L. OpenPBTA: The Open Pediatric Brain Tumor Atlas. *Cell Genomics* (2022).
26. Geng, Z. *et al.* The Open Pediatric Cancer Project. *bioRxivorg* (2024) doi:10.1101/2024.07.09.599086.
27. Louis, D. N. *et al.* The 2021 WHO classification of tumors of the Central Nervous System: A summary. *Neuro. Oncol.* **23**, 1231–1251 (2021).
28. Chen, S. *et al.* A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**, 92–100 (2024).
29. Kim, J. *et al.* AutoGVP: a dockerized workflow integrating ClinVar and InterVar germline sequence variant classification. *Bioinformatics* **40**, (2024).
30. Geoffroy, V. *et al.* AnnotSV: an integrated tool for structural variations annotation. *Bioinformatics* **34**, 3572–3574 (2018).
31. Gurbich, T. A. & Ilinsky, V. V. ClassifyCNV: a tool for clinical annotation of copy-number variants. *bioRxiv* (2020) doi:10.1101/2020.07.20.213215.
32. Bahubeshi, A. *et al.* Germline DICER1 mutations and familial cystic nephroma. *J. Med. Genet.* **47**, 863–866 (2010).
33. Knijnenburg, T. A. *et al.* Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell Rep.* **23**, 239-254.e6 (2018).
34. Stundon, J. L. *et al.* Alternative lengthening of telomeres (ALT) in pediatric high-grade gliomas can occur without ATRX mutation and is enriched in patients with pathogenic germline mismatch repair (MMR) variants. *Neuro. Oncol.* **25**, 1331–1342 (2023).

35. Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* **2017**, (2017).

36. Sepp, T., Yates, J. R. & Green, A. J. Loss of heterozygosity in tuberous sclerosis hamartomas. *J. Med. Genet.* **33**, 962–964 (1996).

37. Garcia-Linares, C. *et al.* Dissecting loss of heterozygosity (LOH) in neurofibromatosis type 1-associated neurofibromas: Importance of copy neutral LOH. *Hum. Mutat.* **32**, 78–90 (2011).

38. Hadfield, K. D. *et al.* Rates of loss of heterozygosity and mitotic recombination in NF2 schwannomas, sporadic vestibular schwannomas and schwannomatosis schwannomas. *Oncogene* **29**, 6216–6221 (2010).

39. D'Angelo, F. *et al.* The molecular landscape of glioma in patients with Neurofibromatosis 1. *Nat. Med.* **25**, 176–187 (2019).

40. Pei, J. *et al.* Bridging the gap between sequence and structure classifications of proteins with AlphaFold Models. *J. Mol. Biol.* **436**, 168764 (2024).

41. Paysan-Lafosse, T. *et al.* The Pfam protein families database: embracing AI/ML. *Nucleic Acids Res.* (2024) doi:10.1093/nar/gkae997.

42. Uludağ Alkaya, D., Lissewski, C., Yeşil, G., Zenker, M. & Tüysüz, B. Expanding the clinical phenotype of RASopathies in 38 Turkish patients, including the rare LZTR1, RAF1, RIT1 variants, and large deletion in NF1. *Am. J. Med. Genet. A* **185**, 3623–3633 (2021).

43. Hodges, T. R. *et al.* Mutational burden, immune checkpoint expression, and mismatch repair in glioma: implications for immune checkpoint immunotherapy. *Neuro. Oncol.* **19**, 1047–1057 (2017).

44. Kim, J. *et al.* Germline pathogenic variants in neuroblastoma patients are enriched in BARD1 and predict worse survival. *J. Natl. Cancer Inst.* **116**, 149–159 (2024).

45. Qian, M. *et al.* TP53 germline variations influence the predisposition and prognosis of B-cell acute lymphoblastic leukemia in children. *J. Clin. Oncol.* **36**, 591–599 (2018).

46. Mirabello, L. *et al.* Frequency of pathogenic germline variants in cancer-susceptibility genes in patients with osteosarcoma. *JAMA Oncol.* **6**, 724–734 (2020).

47. Koczkowska, M. *et al.* Analysis of 200 unrelated individuals with a constitutional NF1 deep intronic pathogenic variant reveals that variants flanking the alternatively spliced NF1 exon 31 [23a] cause a

classical neurofibromatosis type 1 phenotype while altering predominantly NF1 isoform type II. *Hum. Genet.* **142**, 849–861 (2023).

48. Li, M. M. *et al.* Points to consider for reporting of germline variation in patients undergoing tumor testing: a statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **22**, 1142–1148 (2020).
49. Le, D. T. *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409–413 (2017).
50. Randall, M. P. *et al.* BARD1 germline variants induce haploinsufficiency and DNA repair defects in neuroblastoma. *J. Natl. Cancer Inst.* **116**, 138–148 (2024).
51. Morrissy, A. S. *et al.* Divergent clonal selection dominates medulloblastoma at recurrence. *Nature* **529**, 351–357 (2016).
52. Johnson, B. E. *et al.* Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* **343**, 189–193 (2014).
53. Shiraishi, Y. *et al.* Systematic identification of intron retention associated variants from massive publicly available transcriptome sequencing data. *Nat. Commun.* **13**, 5357 (2022).
54. You, S. *et al.* PTCH1, a receptor of Hedgehog signaling pathway, is correlated with metastatic potential of colorectal cancer. *Ups. J. Med. Sci.* **115**, 169–175 (2010).
55. Yang, H. *et al.* Exosomal miR-423-5p targets SUFU to promote cancer growth and metastasis and serves as a novel marker for gastric cancer. *Mol. Carcinog.* **57**, 1223–1236 (2018).
56. Rokita, J. L. *et al. D3b-Center/OpenPedCan-Analysis: V2.5.1 - OpenPedCan Manuscript Release*. (Zenodo, 2025). doi:10.5281/ZENODO.15848007.
57. Komori, T. MPC-17 2021 WHO Classification of Tumors of the CNS, 5th ed. *Neurooncol. Adv.* **3**, vi18–vi18 (2021).
58. Huang, Z., Ren, X. & Hu, J. Drosha: a new tumor suppressor in pineoblastoma. *Genes Dev.* **39**, 677–678 (2025).
59. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013).

60. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

61. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).

62. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).

63. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

64. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* 201178 (2018) doi:10.1101/201178.

65. Kim, J. *et al.* Germline pathogenic variants in neuroblastoma patients are enriched in *BARD1* and predict worse survival. *J. Natl. Cancer Inst.* **116**, 149–159 (2024).

66. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).

67. Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **21**, 974–984 (2011).

68. Corbett, R. J. *et al.* Genetic ancestry superpopulations show distinct prevalence and outcomes across pediatric central nervous system tumors from the PBTA and PNOC. *Neuro. Oncol.* (2025) doi:10.1093/neuonc/noaf017.

69. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* **12**, e1004873 (2016).

70. Boeva, V. *et al.* Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423–425 (2012).

71. Khanna, A. *et al.* Bam-readcount - rapid generation of basepair-resolution sequence metrics. *Journal of Open Source Software* **7**, 3722 (2022).

72. Naqvi, A. S. *et al.* Characterization of aberrant splicing in pediatric central nervous system tumors reveals CLK1 as a candidate oncogenic dependency. *bioRxivorg* 2024.08.03.606419 (2024).

73. Capper, D. *et al.* DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469–474 (2018).

74. Petralia, F. *et al.* Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer. *Cell* **183**, 1962-1985.e31 (2020).

75. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).

76. Wickham, H. Ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* **3**, 180–185 (2011).

77. Therneau, T. M. & Grambsch, P. M. *Modeling Survival Data: Extending the COx Model*. (Springer Science & Business Media, 2000).

78. Therneau, T. M. *A Package for Survival Analysis in R*. https://CRAN.R-project.org/package=survival (2023).

79. Rokita, J. L. *et al. D3b-Center/OpenPedCan-Analysis: V2.4.0 - Interim Module and Code Updates*. (Zenodo, 2024). doi:10.5281/ZENODO.13836051.

FIGURE LEGENDS

Figure 1. Pediatric CNS tumor cohort germline analysis summary. A. Overview of germline and integrative somatic analyses. **B.** Circos plot summarizing distribution of tumor histologies and availability of matched tumor RNA-seq, proteomics, and DNA methylation data modalities in 830 pediatric CNS tumor patients. Atypical Teratoid Rhabdoid Tumors (AT/RT, N=27); choroid plexus tumors (CPT, N=18); craniopharyngiomas (N=39); diffuse intrinsic pontine glioma or diffuse midline glioma (DIPG or DMG, N=53); ependymoma (EPN, N=71); germ cell tumors (GCT, N=13); low-grade gliomas (LGG, N=221); medulloblastomas (MB, N=96); meningiomas (MNG, N=25); mesenchymal tumors (N=19); mixed glioneuronal and neuronal tumors (GNT,

N=79); neurofibroma plexiforms (NFP, N=15); other non-AT/RT, non-MB CNS embryonal tumors (N=12); other high-grade gliomas (HGG, N=76); and schwannomas (SWN, N=13). All other benign tumors and non-cancerous lesions were assigned to a “non-neoplastic tumor” category (N=28), and other rare CNS tumors of low sample size (N<10) were grouped into an “Other tumor” category (N=25). **C.** Sex predicted genetic ancestry superpopulation, and matched tumor event distributions among cohort. AFR=sub-Saharan African, AMR=admixed American, EAS=East Asian, EUR=European, SAS=South Asian.

Figure 2. Prevalence of cancer predisposition gene germline P/LP variants in pediatric CNS tumor patients. A. Number of patients with identified germline P/LP variants by cancer predisposition gene (CPG) and tumor histology. **B.** Number of CPG germline P/LP SNVs and InDels identified by call and call source. **C.** Number of patients with identified P/LP structural variants by gene, tumor histology, and type. DEL=deletion, DUP=duplication. **.** Number of patients harboring P/LP variants in syndrome-associated genes, and clinical definition. RTPS=Rhabdoid Tumor Predisposition Syndrome, FAP=Familial Adenomatous Polyposis, MAS=melanoma astrocytoma syndrome, BAP1 TPS=BAP1 Tumor Predisposition Syndrome. **E.** Percent of patients harboring a CPG P/LP variant by tumor histology. **F.** Odds ratios and associated p-values of CPG P/LP variant burden among pediatric CNS tumor cohort relative to PMBB and gnomAD cancer-free control cohorts. **G-H.** Odds ratios and associated p-values of P/LP variant burden in DNA repair genes from Knijnenburg *et al.* 2018 among **G)** entire cohort and **H)** pediatric high-grade glioma cohort (excluding DIPG and DMG) relative to control cohorts. Dashed lines in p-value plots indicate Bonferroni-adjusted $p<0.05$.

Figure 3. Identification of somatic second hit events in matched tumors. A. Oncoprint of somatic alterations (SNVs/InDels, copy number variation, loss of heterozygosity, differential gene and protein expression) in matched tumors of patients harboring CPG P/LP variants. **B.** Lollipop plots displaying germline P/LP variants ("G") and oncogenic SNVs/InDels ("T") in matched tumors of *APC*, *NF1*, and *SUFU* P/LP carriers. **C.** Tumor versus germline variant allele frequency (VAF) of all P/LP variants detected in matched tumors. **D.** Violin plot of cancer predisposition gene (CPG) transcripts per million (TPM) z-scores in matched tumors of CPG P/LP carriers by histology. Vertical dashed lines indicate z-score thresholds for significantly increased and decreased gene expression. N=18 and N=6 indicate number of cases exhibiting significant transcript loss and gain, respectively. Dot colors correspond to tumor histology-specific color assignments defined in Figure 1B. **E.** TPM z-scores of CPGs in P/LP carriers with identified somatic second hits versus those with no second hits. P-value is derived from a Wilcoxon rank sum test. **F.** TPM z-scores by histology and CPG in P/LP carriers versus non-carriers, with Wilcoxon rank sum test p-values.

Figure 4. Landscape of germline P/LP variant-associated alternative splicing events in matched tumors. A. Germline P/LP variant-proximal splice event percent splice index (PSI) z-scores in matched tumors, grouped by proximity to germline variant and class of variant (splice vs. non-splice). P-values are derived from Wilcoxon rank-sum tests **B.** Transcripts per million (TPM) z-scores against P/LP-variant-proximal splice event PSI z-scores among frameshift variant, splice variant, and stop gained variant P/LP carriers. R values represent Pearson correlation coefficients. **C.** Significant P/LP variant-proximal alternative splicing events in matched tumors of P/LP carriers by splice event type and histology, plotted by percent spliced in (PSI) difference relative to other tumors of same histology. SE=single exon, RI=retained intron, A3SS=alternative 3' splice site. **D.** A germline *PTCH1* splice acceptor P/LP carrier with SHH-activated medulloblastoma (SHH-MB) exhibits increased exon 2 skipping relative in matched tumor relative to other SHH-MB tumors. **E.** A *LZTR1* stop gained P/LP carrier with schwannoma exhibits increased upstream intron 14 retention in matched tumor relative to other schwannomas. **F.** A *TSC2* splice polypyrimidine tract P/LP variant carrier with subependymal giant cell astrocytoma (SEGA) exhibits increased exon 11 skipping in matched tumor relative to other SEGAs. **G.** *LZTR1* TPM z-scores against intron 14 PSI z-scores in

schwannoma cases. **H.** *TSC2* TPM z-scores against exon 11 PSI z-scores in low-grade gliomas. **I.** Two *GBA* splice donor P/LP carriers with *KIAA1549::BRAF* fusion positive LGG and SHH-MB exhibit increased exon 2 skipping in matched tumors relative to other tumors of the same histology. **J.** Indexed tumor mutation burdens in *KIAA1549::BRAF* fusions-positive LGG with *GBA* P/LP carrier PT_EQ5C5TEA highlighted. **K.** *GBA* sashimi plot highlighting exon 2 skipping in initial and metastatic second malignancy of PT_EQ5C5TEA.

Figure 5. Germline P/LP variants in DNA repair genes are associated with distinct molecular profiles in high-grade glioma (HGG). A. Count of DNA repair variant P/LP carriers within CNS tumor histology cohorts by DNA repair pathway, with Fisher's exact test-derived odds ratios in parentheses. Cells are colored by odds ratio weight. *=FDR<0.05, ^=p<0.05. BER=base excision repair, HR=homologous recombination, MMR=mismatch repair, NER=nucleotide excision repair, NHEJ=non-homologous end joining. "Other" indicates genes in the DNA repair list that are not in any of the five pathways. **B.** SBS3 mutational signature exposure weights in HR gene P/LP carriers versus non-carriers with H3-wildtype HGG. **C-D.** MMR deficiency mutational signatures exposure weights **(C)** and tumor mutation burden **(D)** in MMR gene P/LP carriers versus non-carriers with H3-wildtype HGG. **E.** SBS30 exposure weights in BER gene P/LP carriers versus non-carriers with mesenchymal tumors. **F.** Mean beta value from all probes on Infinium MethylationEPIC array in MMR gene P/LP carriers versus non-carriers with H3-wildtype HGG, with constitutional mismatch repair deficiency (CMMRD) syndrome patient indicated by label. P-values are derived from Wilcoxon rank sum tests. **G.** Clustering of H3 wildtype HGG samples by 20k most variable promoter-annotated probes on the Infinium MethylationEPIC array. Gene symbols above heatmap indicate annotation of germline P/LP variants, and *MSH6/MSH6* indicates homozygous P/LP carrier. Hierarchical clustering was performed using the Euclidean distance measure and ward D2 clustering method. **H-I.** A germline P/LP deletion in *MSH2* **(H)** and P/LP SNP in *MSH6* **(I)** are associated with significant hypermethylation of respective promoter probes in H3-wildtype HGG cases. The red box and asterisk in gene models indicate positions of germline P/LP deletion and SNP, respectively. **=|beta value z-score| $\geq$ 2.

Figure 6. P/LP carrier status-associated survival differences in pediatric CNS tumor patients. A. Hazard

ratios of event-free and overall survival (EFS and OS, respectively) in germline CPG P/LP carriers versus non-carriers by tumor histology or molecular subtype. Statistics were derived from cox proportional hazards models that included age at diagnosis covariate for all tumor histologies, and molecular subtype and extent of tumor resection covariates for relevant tumor histologies. **B.** Kaplan-Meier survival curves for EFS and OS in patients with *KIAA1549::BRAF* fusion-positive LGG, stratified by CPG P/LP carrier status. **C.** Cox proportional hazards model forest plot of EFS in patients with *KIAA1549::BRAF* fusion positive LGG, including covariates for extent of surgical resection, age at diagnosis, and interaction between pathology diagnosis and CPG P/LP carrier status. **D.** Kaplan-Meier survival curves for EFS and OS in patients with medulloblastoma, stratified by CPG P/LP carrier status. **E.** Logistic regression model forest plot of metastasis in patients with medulloblastoma. **F.** Tumor mutation burden in patients with medulloblastoma by CPG P/LP carrier status and molecular subtype. P-values are derived from Wilcoxon rank sum tests. **G-H.** Heatmap of patient distribution among **H)** SHH-MB subgroups and **I)** Group 4 MB methylation subgroups by CPG P/LP carrier status, with Fisher's exact test-derived odds ratios in parentheses. *=FDR<0.05. NR=not reportable.

Table 1. Demographic and clinical features of PBTA cohort by CPG P/LP carrier status

	CPG P/LP	No CPG P/LP	p
Number Cohort	193 (23.3%)	637 (76.8%)	
Sex			0.21
Female	93 (48.2%)	262 (45.1%)	
Male	96 (49.7%)	338 (53.4%)	
Unknown	4 (2.1%)	37 (5.8%)	
Race			0.34
American Indian or Alaska Native	0 (0%)	5 (0.8%)	
Asian	2 (1%)	25 (3.9%)	
Black or African American	14 (7.3%)	53 (8.3%)	
More Than One Race	2 (1%)	9 (1.4%)	
Native Hawaiian or Other Pacific Islander	0 (0%)	1 (0.2%)	
Other/Unknown Race	36 (18.7%)	99 (15.5%)	
White	139 (72.0%)	445 (69.9%)	
Ethnicity			0.74
Hispanic or Latino	22 (11.4%)	62 (9.7%)	
Not Hispanic or Latino	163 (84.5%)	544 (85.4%)	
Other/Unknown Ethnicity	8 (4.1%)	31 (4.9%)	
Genetic Ancestry Superpopulation			0.84
AFR	18 (9.3%)	57 (8.9%)	
AMR	28 (14.5%)	73 (11.5%)	
EAS	4 (2.1%)	20 (3.1%)	
EUR	134 (69.4%)	435 (68.3%)	
SAS	5 (2.6%)	16 (2.5%)	
Unknown	4 (2.1%)	36 (5.7%)	
Tumor Histology			***5.0E-04***
ATRT	8 (4.1%)	19 (3.0%)	
Choroid plexus tumor	4 (2.1%)	14 (2.2%)	
Craniopharyngioma	4 (2.1%)	35 (5.5%)	
DIPG or DMG	8 (4.1%)	45 (7.1%)	
Ependymoma	11 (5.7%)	60 (9.4%)	
Germ cell tumor	4 (2.1%)	9 (1.4%)	
Low-grade glioma	47 (24.4%)	174 (27.3%)	
Medulloblastoma	28 (14.5%)	68 (10.7%)	
Meningioma	8 (4.1%)	17 (2.7%)	
Mesenchymal tumor	5 (2.6%)	14 (2.2%)	
Mixed neuronal-glial tumor	11 (5.7%)	68 (10.7%)	
Neurofibroma plexiform	11 (5.7%)	4 (0.6%)	
Non-neoplastic tumor	8 (4.1%)	20 (3.1%)	
Other CNS embryonal tumor	0 (0%)	12 (1.9%)	
Other high-grade glioma	26 (13.5%)	50 (7.8%)	

Other tumor	6 (3.1%)	19 (3.0%)	
Schwannoma	4 (2.1%)	9 (1.4%)	
Median age at diagnosis, years	7.93	8.50	0.47
Median tumor mutation burden	0.39	0.36	0.12

Figure 1

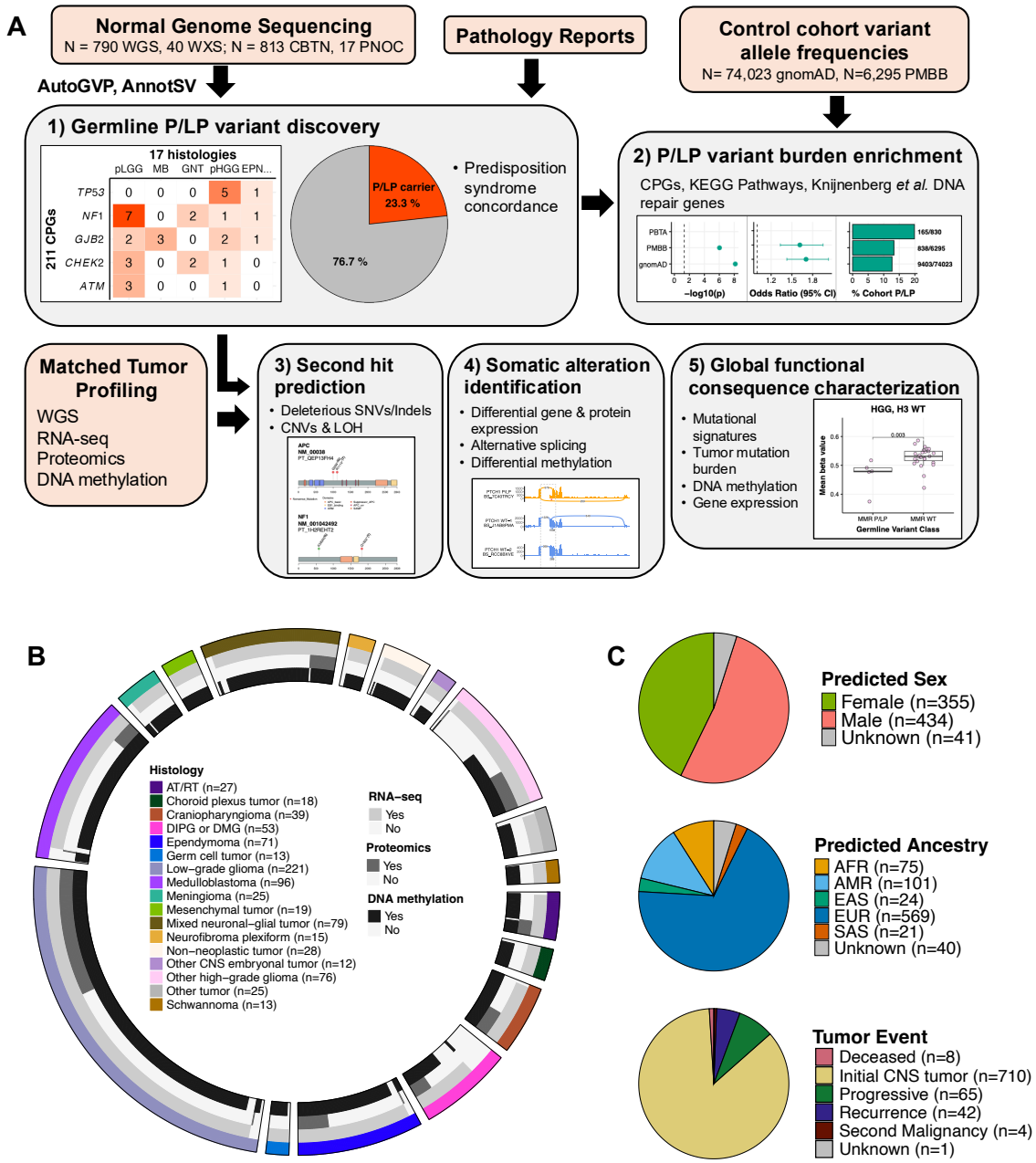

A
Normal Genome Sequencing
N = 790 WGS, 40 WXS; N = 813 CBTN, 17 PNOC
Pathology Reports
Control cohort variant allele frequencies
N= 74,023 gnomAD, N=6,295 PMBB
AutoGVP, AnnotSV
1) Germline P/LP variant discovery
17 histologies
pLGG MB GNT pHGG EPN...
211 CPGs
TP53 0 0 0 5 1
NF1 7 0 2 1 1
GJB2 2 3 0 2 1
CHEK2 3 0 2 1 0
ATM 3 0 0 1 0
P/LP carrier 23.3 %
76.7 %
• Predisposition syndrome concordance
2) P/LP variant burden enrichment
CPGs, KEGG Pathways, Knijnenberg et al. DNA repair genes
PBTA
PMBB
gnomAD
−log10(p)
Odds Ratio (95% CI)
% Cohort P/LP
165/830
838/6295
9403/74023
Matched Tumor Profiling
WGS
RNA-seq
Proteomics
DNA methylation
3) Second hit prediction
• Deleterious SNVs/Indels
• CNVs & LOH
4) Somatic alteration identification
• Differential gene & protein expression
• Alternative splicing
• Differential methylation
5) Global functional consequence characterization
• Mutational signatures
• Tumor mutation burden
• DNA methylation
• Gene expression
HGG, H3 WT
Mean beta value
MMR P/LP
MMR WT
Germline Variant Class
B
Histology
AT/RT (n=27)
Choroid plexus tumor (n=18)
Craniopharyngioma (n=39)
DIPG or DMG (n=53)
Ependymoma (n=71)
Germ cell tumor (n=13)
Low−grade glioma (n=221)
Medulloblastoma (n=96)
Meningioma (n=25)
Mesenchymal tumor (n=19)
Mixed neuronal−glial tumor (n=79)
Neurofibroma plexiform (n=15)
Non−neoplastic tumor (n=28)
Other CNS embryonal tumor (n=12)
Other high−grade glioma (n=76)
Other tumor (n=25)
Schwannoma (n=13)
RNA−seq
Yes
No
Proteomics
Yes
No
DNA methylation
Yes
No
C
Predicted Sex
Female (n=355)
Male (n=434)
Unknown (n=41)
Predicted Ancestry
AFR (n=75)
AMR (n=101)
EAS (n=24)
EUR (n=569)
SAS (n=21)
Unknown (n=40)
Tumor Event
Deceased (n=8)
Initial CNS tumor (n=710)
Progressive (n=65)
Recurrence (n=42)
Second Malignancy (n=4)
Unknown (n=1)

Figure 2

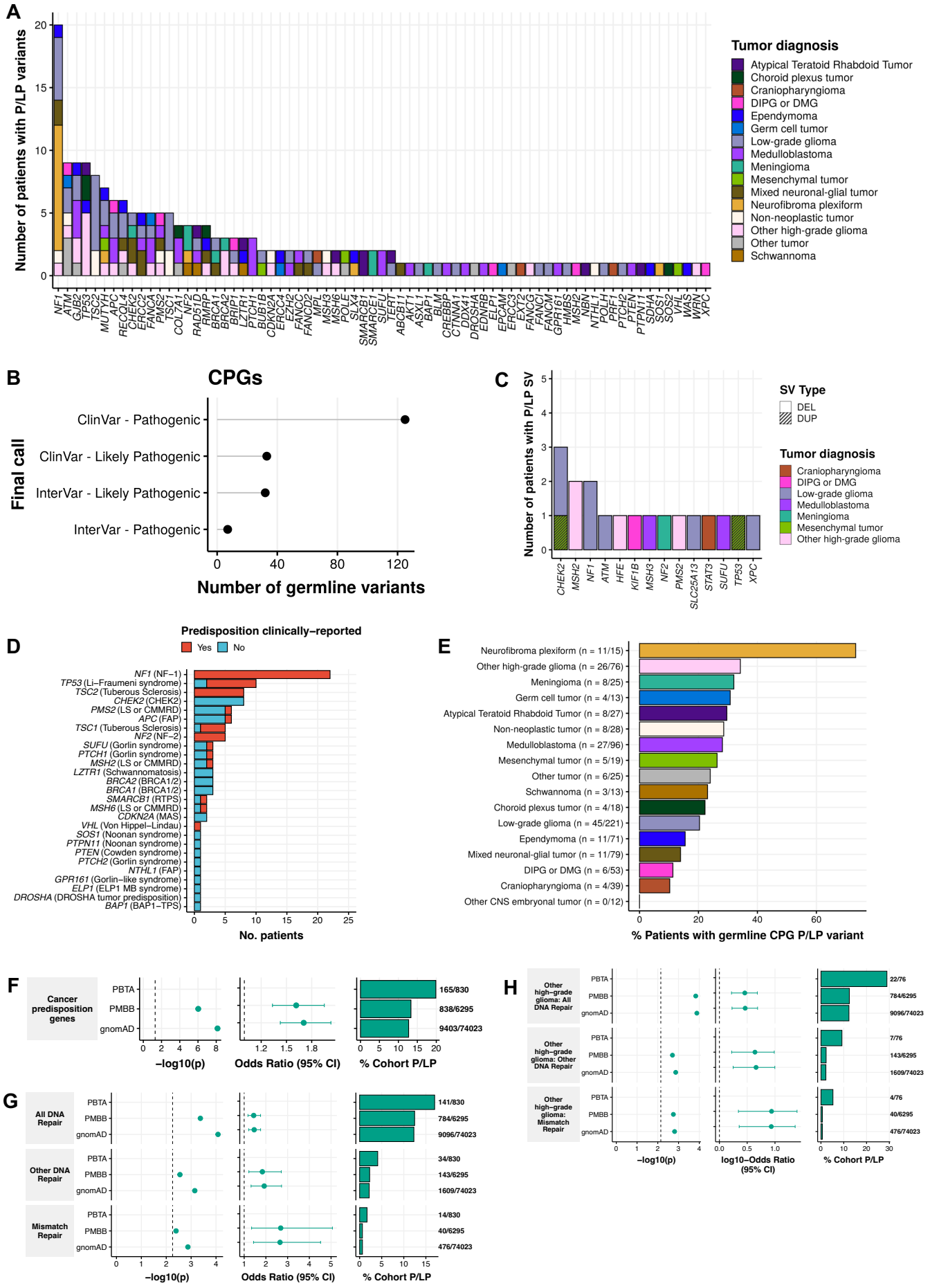

A
Number of patients with P/LP variants
Tumor diagnosis
Atypical Teratoid Rhabdoid Tumor
Choroid plexus tumor
Craniopharyngioma
DIPG or DMG
Ependymoma
Germ cell tumor
Low-grade glioma
Medulloblastoma
Meningioma
Mesenchymal tumor
Mixed neuronal-glial tumor
Neurofibroma plexiform
Non-neoplastic tumor
Other high-grade glioma
Other tumor
Schwannoma
B
CPGs
Final call
ClinVar - Pathogenic
ClinVar - Likely Pathogenic
InterVar - Likely Pathogenic
InterVar - Pathogenic
Number of germline variants
C
Number of patients with P/LP SV
SV Type
DEL
DUP
Tumor diagnosis
Craniopharyngioma
DIPG or DMG
Low-grade glioma
Medulloblastoma
Meningioma
Mesenchymal tumor
Other high-grade glioma
D
Predisposition clinically-reported
Yes
No
No. patients
E
% Patients with germline CPG P/LP variant
F
Cancer predisposition genes
PBTA
PMBB
gnomAD
–log10(p)
Odds Ratio (95% CI)
% Cohort P/LP
G
All DNA Repair
Other DNA Repair
Mismatch Repair
H
Other high-grade glioma: All DNA Repair
Other high-grade glioma: Other DNA Repair
Other high-grade glioma: Mismatch Repair
log10-Odds Ratio (95% CI)

Figure 3

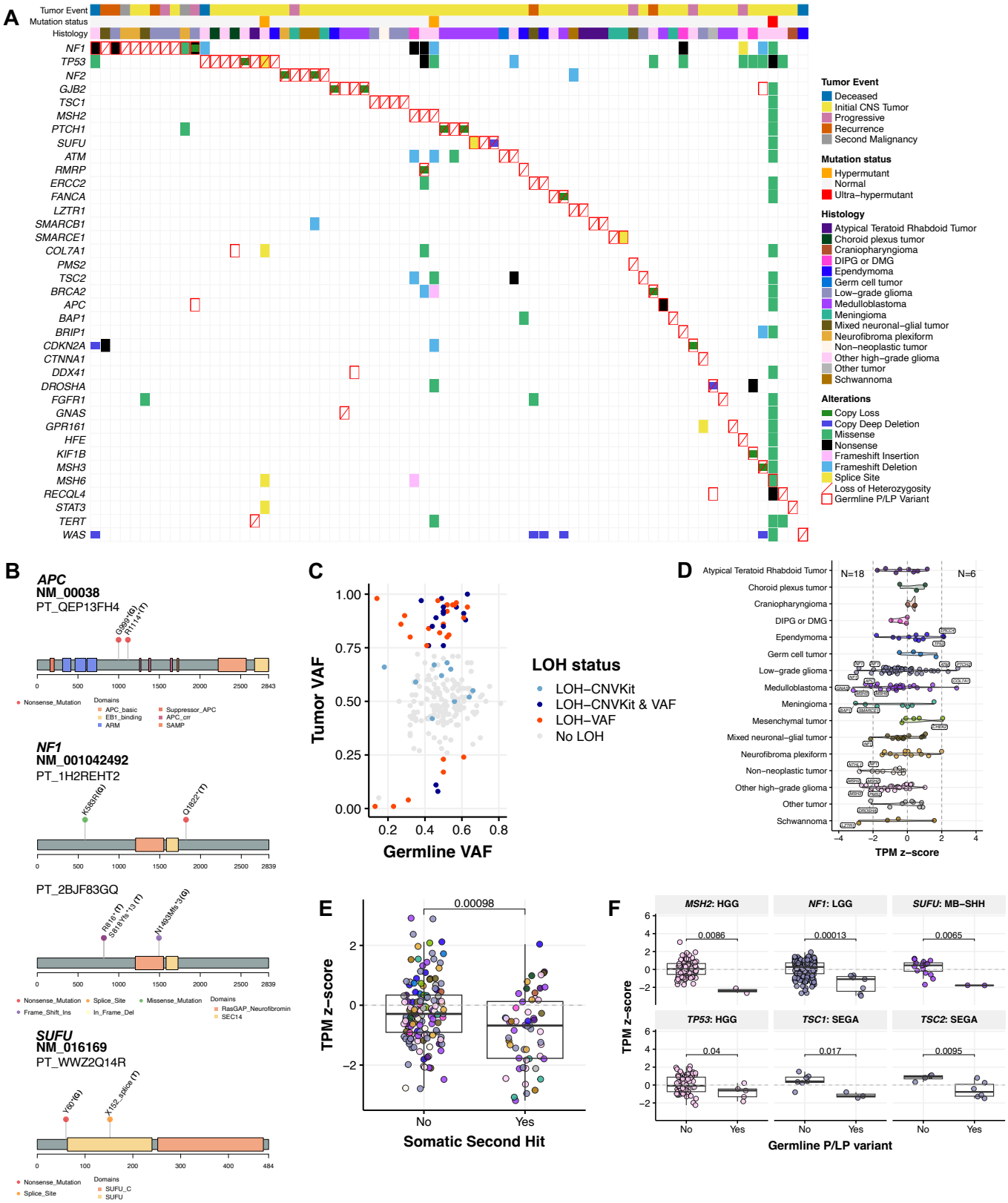

A
Tumor Event
Mutation status
Histology
NF1
TP53
NF2
GJB2
TSC1
MSH2
PTCH1
SUFU
ATM
RMRP
ERCC2
FANCA
LZTR1
SMARCB1
SMARCE1
COL7A1
PMS2
TSC2
BRCA2
APC
BAP1
BRIP1
CDKN2A
CTNNA1
DDX41
DROSHA
FGFR1
GNAS
GPR161
HFE
KIF1B
MSH3
MSH6
RECQL4
STAT3
TERT
WAS
Tumor Event
Deceased
Initial CNS Tumor
Progressive
Recurrence
Second Malignancy
Mutation status
Hypermutant
Normal
Ultra−hypermutant
Histology
Atypical Teratoid Rhabdoid Tumor
Choroid plexus tumor
Craniopharyngioma
DIPG or DMG
Ependymoma
Germ cell tumor
Low−grade glioma
Medulloblastoma
Meningioma
Mixed neuronal−glial tumor
Neurofibroma plexiform
Non−neoplastic tumor
Other high−grade glioma
Other tumor
Schwannoma
Alterations
Copy Loss
Copy Deep Deletion
Missense
Nonsense
Frameshift Insertion
Frameshift Deletion
Splice Site
Loss of Heterozygosity
Germline P/LP Variant
B
APC
NM_00038
PT_QEP13FH4
G869*(G)
R1114*(T)
Nonsense_Mutation
Domains
APC_basic
EB1_binding
ARM
Suppressor_APC
APC_crr
SAMP
NF1
NM_001042492
PT_1H2REHT2
K585R(G)
Q1822*(T)
PT_2BJF83GQ
R816*(T)
S818fs*13 (T)
N1493Mfs*3(G)
Nonsense_Mutation
Splice_Site
Missense_Mutation
Frame_Shift_Ins
In_Frame_Del
Domains
RasGAP_Neurofibromin
SEC14
SUFU
NM_016169
PT_WWZ2Q14R
Y60*(G)
X152_splice (T)
Nonsense_Mutation
Splice_Site
Domains
SUFU_C
SUFU
C
Tumor VAF
Germline VAF
LOH status
LOH−CNVKit
LOH−CNVKit & VAF
LOH−VAF
No LOH
D
Atypical Teratoid Rhabdoid Tumor
Choroid plexus tumor
Craniopharyngioma
DIPG or DMG
Ependymoma
Germ cell tumor
Low−grade glioma
Medulloblastoma
Meningioma
Mesenchymal tumor
Mixed neuronal−glial tumor
Neurofibroma plexiform
Non−neoplastic tumor
Other high−grade glioma
Other tumor
Schwannoma
N=18
N=6
TPM z−score
E
0.00098
TPM z−score
No
Yes
Somatic Second Hit
F
MSH2: HGG
NF1: LGG
SUFU: MB−SHH
TP53: HGG
TSC1: SEGA
TSC2: SEGA
0.0086
0.00013
0.0065
0.04
0.017
0.0095
TPM z−score
No
Yes
Germline P/LP variant

Figure 4

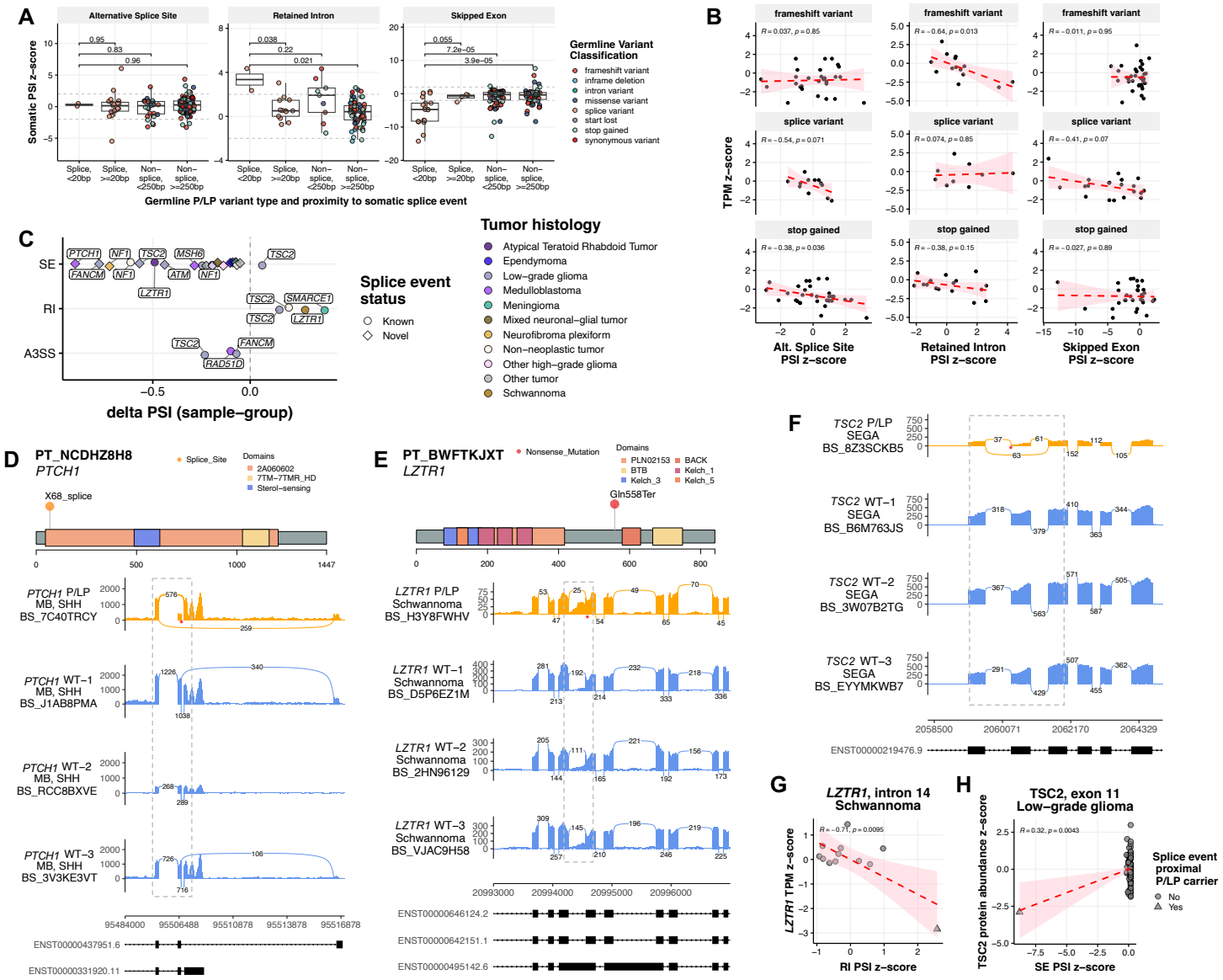

A
Alternative Splice Site
Retained Intron
Skipped Exon
Somatic PSI z−score
Germline P/LP variant type and proximity to somatic splice event
Splice, <20bp
Splice, >=20bp
Non−splice, <250bp
Non−splice, >=250bp
Germline Variant Classification
frameshift variant
inframe deletion
intron variant
missense variant
splice variant
start lost
stop gained
synonymous variant
B
frameshift variant
splice variant
stop gained
TPM z−score
Alt. Splice Site PSI z−score
Retained Intron PSI z−score
Skipped Exon PSI z−score
C
SE
RI
A3SS
delta PSI (sample−group)
Splice event status
Known
Novel
Tumor histology
Atypical Teratoid Rhabdoid Tumor
Ependymoma
Low−grade glioma
Medulloblastoma
Meningioma
Mixed neuronal−glial tumor
Neurofibroma plexiform
Non−neoplastic tumor
Other high−grade glioma
Other tumor
Schwannoma
D
PT_NCDHZ8H8
PTCH1
Splice_Site
Domains
2A060602
7TM−7TMR_HD
Sterol−sensing
X68_splice
PTCH1 P/LP MB, SHH BS_7C40TRCY
PTCH1 WT−1 MB, SHH BS_J1AB8PMA
PTCH1 WT−2 MB, SHH BS_RCC8BXVE
PTCH1 WT−3 MB, SHH BS_3V3KE3VT
ENST00000437951.6
ENST00000331920.11
E
PT_BWFTKJXT
LZTR1
Nonsense_Mutation
Domains
PLN02153
BACK
BTB
Kelch_1
Kelch_3
Kelch_5
Gln558Ter
LZTR1 P/LP Schwannoma BS_H3Y8FWHV
LZTR1 WT−1 Schwannoma BS_D5P6EZ1M
LZTR1 WT−2 Schwannoma BS_2HN96129
LZTR1 WT−3 Schwannoma BS_VJAC9H58
ENST00000646124.2
ENST00000642151.1
ENST00000495142.6
F
TSC2 P/LP SEGA BS_8Z3SCKB5
TSC2 WT−1 SEGA BS_B6M763JS
TSC2 WT−2 SEGA BS_3W07B2TG
TSC2 WT−3 SEGA BS_EYYMKWB7
ENST00000219476.9
G
LZTR1, intron 14 Schwannoma
LZTR1 TPM z−score
RI PSI z−score
H
TSC2, exon 11 Low−grade glioma
TSC2 protein abundance z−score
SE PSI z−score
Splice event proximal P/LP carrier
No
Yes

Figure 5

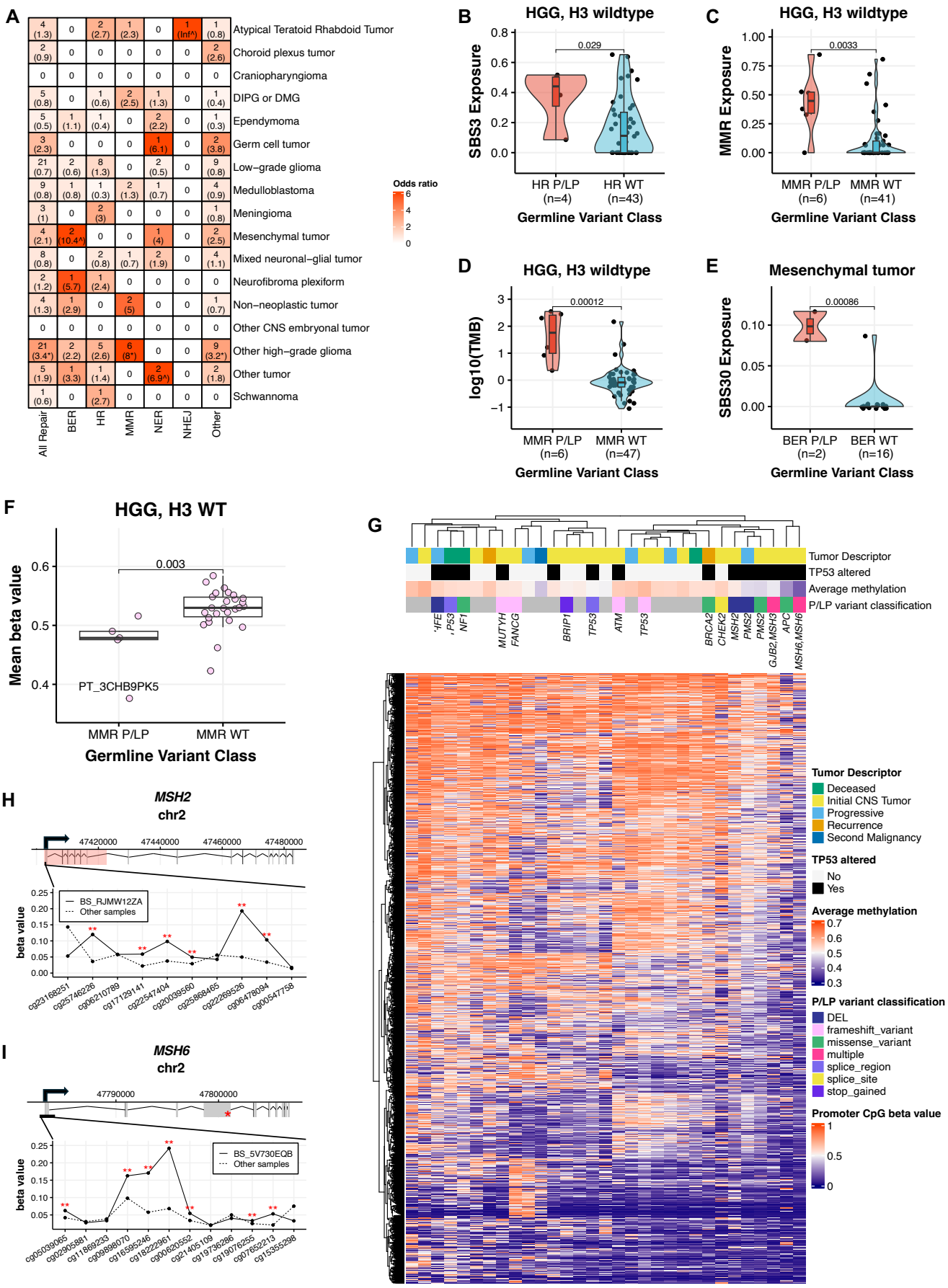
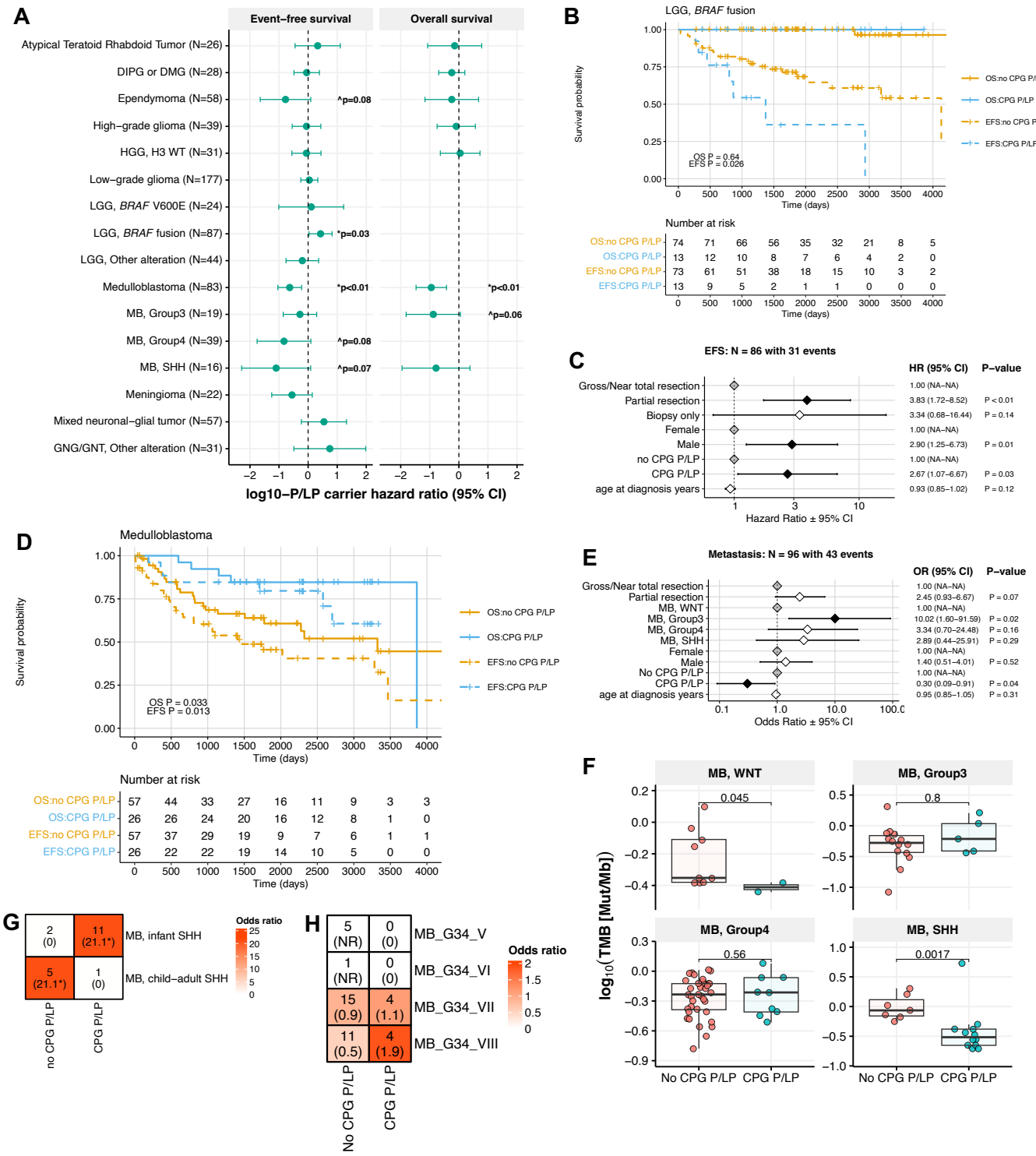

A
Atypical Teratoid Rhabdoid Tumor
Choroid plexus tumor
Craniopharyngioma
DIPG or DMG
Ependymoma
Germ cell tumor
Low−grade glioma
Medulloblastoma
Meningioma
Mesenchymal tumor
Mixed neuronal−glial tumor
Neurofibroma plexiform
Non−neoplastic tumor
Other CNS embryonal tumor
Other high−grade glioma
Other tumor
Schwannoma
All Repair
BER
HR
MMR
NER
NHEJ
Other
Odds ratio
B
HGG, H3 wildtype
SBS3 Exposure
0.029
HR P/LP (n=4)
HR WT (n=43)
Germline Variant Class
C
HGG, H3 wildtype
MMR Exposure
0.0033
MMR P/LP (n=6)
MMR WT (n=41)
Germline Variant Class
D
HGG, H3 wildtype
log10(TMB)
0.00012
MMR P/LP (n=6)
MMR WT (n=47)
Germline Variant Class
E
Mesenchymal tumor
SBS30 Exposure
0.00086
BER P/LP (n=2)
BER WT (n=16)
Germline Variant Class
F
HGG, H3 WT
Mean beta value
0.003
PT_3CHB9PK5
MMR P/LP
MMR WT
Germline Variant Class
G
Tumor Descriptor
TP53 altered
Average methylation
P/LP variant classification
MUTYH
FANCG
BRIP1
TP53
ATM
TP53
BRCA2
CHEK2
MSH2
PMS2
PMS2
GJB2,MSH3
APC
MSH6,MSH6
Tumor Descriptor
Deceased
Initial CNS Tumor
Progressive
Recurrence
Second Malignancy
TP53 altered
No
Yes
Average methylation
P/LP variant classification
DEL
frameshift_variant
missense_variant
multiple
splice_region
splice_site
stop_gained
Promoter CpG beta value
H
MSH2
chr2
47420000
47440000
47460000
47480000
BS_RJMW12ZA
Other samples
beta value
cg23168251
cg25746226
cg06210789
cg17129141
cg22547404
cg20039560
cg25868465
cg22269526
cg06478094
cg00547758
I
MSH6
chr2
47790000
47800000
BS_5V730EQB
Other samples
beta value
cg05039065
cg02905881
cg11869233
cg09898070
cg16595246
cg18222961
cg00620552
cg21405109
cg19736286
cg19076255
cg07652213
cg15355298

Figure 6


A
Event–free survival
Overall survival
Atypical Teratoid Rhabdoid Tumor (N=26)
DIPG or DMG (N=28)
Ependymoma (N=58)
High–grade glioma (N=39)
HGG, H3 WT (N=31)
Low–grade glioma (N=177)
LGG, BRAF V600E (N=24)
LGG, BRAF fusion (N=87)
LGG, Other alteration (N=44)
Medulloblastoma (N=83)
MB, Group3 (N=19)
MB, Group4 (N=39)
MB, SHH (N=16)
Meningioma (N=22)
Mixed neuronal–glial tumor (N=57)
GNG/GNT, Other alteration (N=31)
^p=0.08
*p=0.03
*p<0.01
*p<0.01
^p=0.06
^p=0.08
^p=0.07
log10–P/LP carrier hazard ratio (95% CI)
B
LGG, BRAF fusion
Survival probability
OS P = 0.64
EFS P = 0.026
Time (days)
OS:no CPG P/LP
OS:CPG P/LP
EFS:no CPG P/LP
EFS:CPG P/LP
Number at risk
OS:no CPG P/LP 74 71 66 56 35 32 21 8 5
OS:CPG P/LP 13 12 10 8 7 6 4 2 0
EFS:no CPG P/LP 73 61 51 38 18 15 10 3 2
EFS:CPG P/LP 13 9 5 2 1 1 0 0 0
C
EFS: N = 86 with 31 events
HR (95% CI) P–value
Gross/Near total resection 1.00 (NA–NA)
Partial resection 3.83 (1.72–8.52) P < 0.01
Biopsy only 3.34 (0.68–16.44) P = 0.14
Female 1.00 (NA–NA)
Male 2.90 (1.25–6.73) P = 0.01
no CPG P/LP 1.00 (NA–NA)
CPG P/LP 2.67 (1.07–6.67) P = 0.03
age at diagnosis years 0.93 (0.85–1.02) P = 0.12
Hazard Ratio ± 95% CI
D
Medulloblastoma
Survival probability
OS P = 0.033
EFS P = 0.013
Time (days)
OS:no CPG P/LP
OS:CPG P/LP
EFS:no CPG P/LP
EFS:CPG P/LP
Number at risk
OS:no CPG P/LP 57 44 33 27 16 11 9 3 3
OS:CPG P/LP 26 26 24 20 16 12 8 1 0
EFS:no CPG P/LP 57 37 29 19 9 7 6 1 1
EFS:CPG P/LP 26 22 22 19 14 10 5 0 0
E
Metastasis: N = 96 with 43 events
OR (95% CI) P–value
Gross/Near total resection 1.00 (NA–NA)
Partial resection 2.45 (0.93–6.67) P = 0.07
MB, WNT 1.00 (NA–NA)
MB, Group3 10.02 (1.60–91.59) P = 0.02
MB, Group4 3.34 (0.70–24.48) P = 0.16
MB, SHH 2.89 (0.44–25.91) P = 0.29
Female 1.00 (NA–NA)
Male 1.40 (0.51–4.01) P = 0.52
No CPG P/LP 1.00 (NA–NA)
CPG P/LP 0.30 (0.09–0.91) P = 0.04
age at diagnosis years 0.95 (0.85–1.05) P = 0.31
Odds Ratio ± 95% CI
F
MB, WNT
MB, Group3
MB, Group4
MB, SHH
0.045
0.8
0.56
0.0017
log10(TMB [Mut/Mb])
No CPG P/LP
CPG P/LP
G
2 (0)
11 (21.1*)
5 (21.1*)
1 (0)
MB, infant SHH
MB, child–adult SHH
no CPG P/LP
CPG P/LP
Odds ratio
H
5 (NR)
0 (0)
1 (NR)
0 (0)
15 (0.9)
4 (1.1)
11 (0.5)
4 (1.9)
MB_G34_V
MB_G34_VI
MB_G34_VII
MB_G34_VIII
No CPG P/LP
CPG P/LP
Odds ratio