Characterization of aberrant splicing in pediatric central nervous system tumors reveals *CLK1* as a candidate oncogenic dependency

Ammar S. Naqvi[1,2], Ryan J. Corbett[1,2,3], Priyanka Seghal[4], Karina L. Conkrite[3], Komal S. Rathi[1,2], Brian M. Ennis[1,2], Katharina E Hayer[3,5], Bo Zhang[1,2], Miguel A. Brown[1,2], Daniel P. Miller[1,2], Adam A. Kraya[1,2], Joseph M. Dybas[1,2], Zhuangzhuang Geng[1,2], Christopher Blackden[1,2], Shehbeel Arif[1,2], Antonia Chroni[1,2], Aditya Lahiri[5], Madison L. Hollawell[1,2], Phillip B. Storm[1,2], Jessica B. Foster[3,6], Mateusz Koptyra[1,2], Peter J. Madsen[1,2], Sharon J. Diskin[3,5,6], Andrei Thomas-Tikhonenko[3,4,6,7], Adam C. Resnick[1,2], Jo Lynne Rokita[1,2,*]

[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] Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, PA

[4] Division of Cancer Pathobiology, Children's Hospital of Philadelphia, Philadelphia, PA, USA

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

[6] Department of Pediatrics, University of Pennsylvania, Philadelphia, PA, USA

[7] Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA

*Correspondence:

Jo Lynne Rokita, Ph.D.

Children's Hospital of Philadelphia

Philadelphia, PA 19104

E-mail: rokita@chop.edu

Highlights

- Splice variation across pediatric CNS tumors from 1,434 patients is heterogeneous
- In high-grade gliomas, 12,145 splice events in 5,424 genes contain functional sites
- *CLK1* exon 4 inclusion adds two phosphorylation binding sites and drives expression
- *CLK1* splicing impacts cancer genes, phenotypes, and is therapeutically targetable

Summary

Pediatric brain cancer is the leading cause of disease-related mortality in children, and many aggressive tumors still lack effective treatment strategies. Despite extensive studies characterizing these tumor genomes, alternative transcriptional splicing patterns remain underexplored. Here, we systematically characterized aberrant alternative splicing across pediatric brain tumors, identifying pediatric high-grade gliomas (HGGs) among the most heterogeneous. Through integration with UniProt Knowledgebase annotations, we identified 12,145 splice events in 5,424 genes, leading to functional changes in protein activation, folding, and localization. We discovered that the master splicing factor and cell-cycle modulator, *CDC-like kinase 1* (*CLK1*), is aberrantly spliced in HGGs to include exon 4, resulting in a gain of two phosphorylation sites and subsequent activation of CLK1. Inhibition of *CLK1* with Cirtuvivint in the pediatric HGG KNS-42 cell line significantly decreased both cell viability and proliferation in a dose-dependent manner. Morpholino-mediated depletion of *CLK1* exon 4 splicing reduced RNA expression, protein abundance, and cell viability. Notably, KNS-42 cells treated with the *CLK1* exon 4 morpholino demonstrated differential expression 78 genes and differential splicing with loss or gain of a functional site in 193 genes annotated as oncogene or tumor suppressor genes (TSGs). These genes were enriched for cancer-associated pathways, with 20 identified as significant gene dependencies in pediatric HGGs. Our findings highlight a dependency of pediatric HGGs on *CLK1* and its roles contributing to tumor splicing heterogeneity through

transcriptional dysregulation of splicing factors and transcriptional modulation of oncogenes. Overall, aberrant splicing in HGGs and other pediatric brain tumors represents a potentially targetable oncogenic pathway contributing to tumor growth and maintenance.

Introduction

Pediatric brain cancer is the number one cause of disease-related death in children[1]. Furthermore, pediatric high-grade gliomas (HGGs) are largely resistant to chemotherapy and are often surgically unresectable, making them exceptionally challenging[2]. Despite decades of clinical trials in pediatrics, patients with certain tumor types such as diffuse intrinsic pontine glioma (DIPG) or diffuse midline glioma (DMG) will succumb to their disease, with a median overall survival of 9.73 months in patients undergoing surgical biopsy[3]. Even with optimal multimodal therapy, median overall survival for non-DMG patients with a HGG is 14-20 months[4]. Within the last decade, surgical biopsies and post-mortem tissue collection has enabled genomic profiling and somatic characterization of pediatric HGGs. While somatic drivers such as mutations, fusions, and copy number amplifications and/or deletions have been well-characterized, the splicing landscape of pediatric HGGs remains underexplored. Two reports have shown rare, but mutually exclusive, alterations in spliceosome-related factors such as *SF3B1* and *SF3B2* in pediatric HGGs. These alterations dysregulated additional cellular processes involved in maintenance of DNA replication, genome integrity, or transcriptional fidelity[5,6]. More recently, Siddaway, et. al. describe alternative aberrant splicing as a novel mechanism of oncogenic pathway activation in pediatric HGGs. In this study, cancer drivers, including members of the RAS/MAPK pathway, were differentially spliced leading to the activation of the RAS/MAPK pathway and worse glioblastoma patient survival[7]. Alternative splicing plays critical roles in numerous fundamental biological processes, for example, increasing proteomic diversity, stabilizing or destabilizing mRNA transcripts, lowering protein steady-state levels, and influencing critical protein functionality such as enzymatic activity,

protein folding, and localization[8,9]. Splicing plays a vital role in generating tissue-specific transcriptomes. The mammalian brain expresses the most complex and conserved alternative splicing programs in relation to other tissues and their disruption can result in a variety of neurological diseases and disorders[10,11]. The splicing process is regulated by a balance of multiple trans-acting RNA-binding proteins (RBP), such as the Serine-rich Splicing Factors (SRSFs), that bind to splicing regulatory elements across preliminary mRNA transcripts. Disturbing that balance can be detrimental to cellular functionality, and has been associated with oncogenic transformation[12,13].

Here, we perform a large-scale analysis of genome-wide splicing across pediatric CNS tumors and show widespread dysregulation of pre-mRNA alternative splicing across pediatric CNS tumors. Our findings reveal that alternative splicing can result in the removal or addition of known functional sites predicted to have significant downstream effects on various cellular functions potentially leading to cell viability and cell-cycle dysregulation. We identify and experimentally characterize *CLK1* as a candidate oncogenic dependency in pediatric HGGs and propose that targeting of aberrant splicing and/or its resulting downstream proteins may offer additional therapeutic avenues for precision cancer therapy.

Results

Pediatric brain tumors display heterogeneous global patterns of aberrant splicing

We utilized replicate Multivariate Analysis of Transcript splicing (rMATs)[14,15] to quantify transcriptome-wide alternative splicing events across primary pediatric brain tumors (N = 1,434) from the Open Pediatric Cancer (OpenPedCan) project[16]. These broad tumor types shown in **Figure 1A** include Atypical Teratoid Rhabdoid Tumor (ATRT, N = 56), choroid plexus tumor (CPT, N = 32), craniopharyngioma (CPG, N = 51), diffuse intrinsic pontine glioma or diffuse midline glioma (DIPG or DMG, N = 131), ependymoma (N = 110), germ cell tumor (N = 16), low-grade glioma (LGG, N = 337), medulloblastoma (MB, N = 200), meningioma (N = 29),

mesenchymal tumor (N = 27), mixed neuronal-glial tumor (GNT, N = 112), neurofibroma plexiform (N = 12), non-neoplastic tumor (N = 45), other CNS embryonic tumor (N = 16), other high-grade glioma (other HGG, N = 212), schwannoma (N = 21), and other rare brain tumors (N = 27). Associated demographic and clinical data for each patient and tumor in this study are available in **Table S1**. We examined four types of splicing events (single exon: SE, alternative 5' splice site: A5SS, alternative 3' splice site: A3SS, and retained intron: RI) and observed that SE splicing is overwhelmingly the most frequently observed (**Figure 1B**) for both increased inclusion and skipping events, consistent with a previous report in pediatric HGGs[7]. Importantly, we did not observe an effect of RNA library on percent spliced in (PSI) predictions and found that the sample aliquots sequenced with a poly-A or stranded library had strong positive correlations for genome-wide PSI values (Pearson's R = 0.98, p-value < 2.2e-16, **Figure S1A-B**).

Due to the diverse biological drivers of these CNS cancers and their molecular subtypes, we hypothesized that we might observe histology- or subtype-specific splicing patterns. Indeed, our assessment of recurrent exon skipping (ES) or exon inclusion (EI) events (N ≥ 2) revealed histology-specific splicing events in some tumor types, shared events across some histologies, as well as lack of recurrent events in other histologies when focusing on the top 40 sets that can comprise one histology or multiple histologies (**Fig 1C**, **Figure S1B**). Within these histologies, we found that MB, LGG, and HGG tumors exhibited the highest number of recurrent, unique skipping events (MB = 2,644, LGG = 1,860, other HGG = 1,610, DMG = 687) and inclusion events (MB = 1,235, LGG = 814, other HGG = 677). To ensure this observation was not skewed by the number of patient tumors per histology grouping, we normalized these events by the number of patients. We observed that MB and HGG tumors had the highest total number of unique, recurrent single exon events per patient (**Fig 1D**). We plotted the top 40 sets with a complete list of unique events per histology reported in **Table S2A-B**. Taken together, these

findings suggest that there are both tumor type-specific and shared splicing patterns in pediatric brain tumors.

We further investigated single exon-associated events and devised a metric called the splicing burden index (SBI) to compute the proportion of differential alternative splicing (AS) events in each sample compared to all other tumor samples (**Figure 1E**, see **Methods**). This metric allows for transcriptome-wide assessment of differential AS within a tumor sample. The median SBI across this cohort was 0.029 (or 2.9%). LGG tumors had the lowest median SBI (0.0150, 1.5%), while germ cell tumors had the highest (0.0802, 8.02%). Tumors with a low splicing burden variance (< 1st quantile variance across all tumors, or 0.0015561) include other tumors, other CNS embryonal tumors, meningiomas, choroid plexus tumors and schwannomas, while more heterogeneous tumors (splicing burden variance > 0.0022047) include DIPG or DMG, MB, mesenchymal tumors, neurofibroma plexiform, and germ cell tumors. We performed a similar analysis on the other splicing cases (A5SS, A3SS, and RI) and observed that LGGs and schwannomas continued to exhibit the lowest median SBI across splice types, while other tumors and germ cell tumors (GCTs) maintained the highest median SBI. However, the ordering of tumors shifted in some instances, suggesting that certain types of splice events may be more prevalent in some histologies compared to others. For example, DIPG or DMG tumors changed from having a low median SBI for SE events to a higher median SBI for other event types, suggesting an increased splicing burden for RI, A5SS, and A3SS splicing changes (**Figures S1D-F**). We next assessed the proportion of low and high SBI tumors in the cohort by histology. The majority (> 70%, N = 156/213) of LGGs consisted of tumors with low SBI, while the majority of germ cell tumors (N = 11/12) had a high SBI (**Figure 1F**). Moreover, greater than 25% of tumors within the following histologies had a high SBI: other HGG, MB, ATRT, CPG, other tumors, non-neoplastic tumors, GCT, mesenchymal tumors, and neurofibroma plexiform while greater than 25% of tumors within the following histologies had a low SBI: LGG, GNT, DIPG or

DMG, and CPG. These results further highlight the heterogeneous splicing landscape across pediatric CNS tumors.

We next hypothesized that tumors with a low tumor mutation burden (TMB) might have a higher splicing burden index as an alternate mechanism driving tumorigenesis. Interestingly, we did not find an overall correlative relationship between TMB and SBI when analyzing the full CNS tumor cohort together (all tumors: Pearson's R = 0.037, p-value = 0.19 and with hyper-mutant tumors removed: R = 0.018, p-value = 0.53, **Figure S1G-H**), but when assessing the relationship by histology (**Figure S1I**), we found significant negative correlations between TMB and SBI in LGG (Pearson's R = -0.15, p-value = 0.0059), DIPG or DMG (Pearson's R = -0.24, p-value = 0.025), and Schwannomas (Pearson's R = -0.77, p-value = 0.00018). As an alternate approach, we asked whether TMB was different between tumors with high and low SBI within histologies (**Figure 1F**). We found a significant inverse relationship between SBI and TMB in only CPTs, LGGs, GNTs, and schwannomas (Wilcoxon p-value < 0.05), indicating that aberrant splicing may serve as an alternative reservoir for uncovering oncogenic mechanisms in mutationally silent tumors.

Splicing drives novel biological clusters and splicing burden differentiates key splicing factors in pediatric high-grade gliomas

To determine whether CNS tumors share transcriptional splicing biology, we performed consensus clustering of PSI values across all primary CNS tumors using the Partition Around Medoids algorithm with Canberra distance metric, which revealed 12 clusters spanning histologies (**Figure 2A**). A full list of parameters and statistics are detailed in **Table S2C**. Clusters 1 and 12 were predominantly composed of MB (**Figure 2B**) with both containing all four MB subtypes: SHH, WNT, Group 3, and Group 4 (**Figure S2C**), suggesting unique splice-driven biological underpinnings even within molecularly-defined subtypes. Although Cluster 11 was dominated by EPNs (**Figure 2B**), it comprised all EPN molecular subtypes (**Figure S2C**).

While Cluster 4 contained the majority of LGGs in the cohort, LGGs spanned nine of 12 clusters (**Figure S2C**) and clusters were not subtype-specific. Notably, HGGs, including DIPG or DMG, exhibited the highest degree of splicing heterogeneity, spanning across all 12 clusters (**Figure 2B and S2C**). Taken together, the transcriptional mechanisms underlying these clusters' formation are not solely driven by molecular subtype. A full list of samples with associated cluster membership information is outlined in **Table S2D**. We further assessed cluster membership of tumors with high or low SBI and found that tumors with high SBI were present in ten clusters, while the low SBI tumors grouped almost entirely with Cluster 4, followed by Clusters 3, 5, and 11 (**Figure 2C**). We aimed to understand the biology driving cluster formation and used gene set variation analysis (GSVA) to identify enriched cancer-associated signaling pathways represented by splice events in each of the 12 clusters. **Figure 2D** displays the top pathways differentially-regulated between pairs of clusters (Bonferroni-adjusted p-value < 0.05). Strikingly, the spliceosome pathway was significantly up- or down-regulated in all cluster groups when performing pairwise comparisons to all other groups (Bonferroni-adjusted p-value < 1.28^{e-46} - 1.01^{e-4}). Further, we show a significant positive correlation (**Figure 2E**, Pearson's R = 0.55, p-value < 2.2e-16) between GSVA spliceosome pathway scores and SBI, both validating the use of the SBI metric to measure splicing activity and suggesting that splicing factors themselves are mis-spliced, likely contributing to the transcriptome-wide differential splicing we are observing within pediatric brain tumors. Other pathways that were enriched for certain clusters included DNA repair, mitotic spindle, and KRAS signaling. For example, Cluster 1 was dominated by MB subtype Group 4 (**Figure S2C**), which also showed significant dysregulation in the KRAS signaling pathway, corroborating early characterization of these tumor types[17]. These results reveal that although we observe transcriptome-wide splicing variation in all tumors, distinct pathways and genes are targeted in each cluster. Taken together, we show that each cluster may have unique transcriptional underpinnings influencing distinct pathways which may, in turn, contribute to tumorigenesis.

Given their cluster and SBI heterogeneity and high number of unique, recurrent splice events, we narrowed our focus to all pediatric HGGs. Since we sought to understand the mechanisms underlying the widespread and pervasive splicing, we first assessed somatic alterations in splicing factor-encoding genes or splicing regulators in our cohort. Interestingly, a recent splicing study found that 34% of pediatric HGGs had a somatic alteration, mutation and/or CNV, in a gene in the HUGO spliceosome complex[7]. Since few of these genes overlap with the splicing factors and their canonical regulators assessed in our study, we combined our gene lists and investigated somatic alterations in these genes. We removed hypermutant tumors (≥ 10 Mut/Mb), filtered for putative deleterious mutations (defined by SIFT and/or PolyPhen), and found that 128 HGG tumors from 195 patients (65%) harbored at least one somatic SNVs/InDel, fusion, or CNV in a gene in the spliceosome complex or that regulates alternative splicing (**Table S3A, Supplemental Figure S2D**). However, the mutation frequencies in each gene only ranged from 1-8%, and there was no significant enrichment in any gene based on SBI status. It has been previously shown that in the absence of splicing factor gene mutations, RNA expression changes in these genes can cause downstream splicing changes to promote tumor formation [18–20]. Thus, we performed differential gene expression (DE) analysis between high vs low SBI HGGs for known splicing factors and related genes[21] (**Figure 2F**) and found 44.3% (N = 77/174) to be significantly differentially expressed (adjusted p-value < 0.05 and $\log_2$-fold change > |2|, **Table S2**). Specifically, 57% (16/28) genes encoding the serine/arginine-rich splicing factor (SRSF) and heterogeneous nuclear ribonucleoproteins (hnRNP) families of trans-acting splicing factors known to directly influence exon-associated splicing were significantly DE between high vs low SBI HGGs. (**Figure 2G)**. Since changes in gene expression may not necessarily result in corresponding changes at the protein level, this finding prompted us to investigate proteomic alterations in these splicing factors. We integrated

gene expression and proteogenomic (N = 188) data from pediatric brain tumors obtained from the Clinical Proteomic Tumor Analysis Consortium (CPTAC)[22]. We observed that mRNA and protein expression for these splicing factors are tightly correlated across pediatric CNS tumors (**Figure 2H**), supporting previous reports that differential mRNAs are better correlates to protein levels[23]. As such, mRNA levels of these splicing factor genes can be used as surrogate measurements for protein abundance and function.

Recurrent splicing aberrations alter known proteomic functional sites, including the gain of phosphorylation binding sites in splicing regulator *CLK1*

In order to further elucidate the aberrant splicing landscape of pediatric HGGs (N = 343), we developed a robust and adaptable workflow to prioritize recurrent single exon events with predicted functional impact (**Figure 3A**). We first applied a threshold of ≥ 10 junction read counts to identify expressed splice events. Next, we prioritized histology-specific recurrent (N ≥ 2) events that were differentially spliced (ΔPSI z-score > |2|) in a sample compared to the whole cohort. We identified a total of 56,550 recurrent differential splicing events in HGGs. Subsequently, we annotated these events for overlap with known Uniprot functional sites and prioritized those leading to gain or loss of a functional site affecting disulfide bonding, localization signaling, amino acid modifications, or others. The Uniprot annotated “other” category is user-defined, but includes sites for ion-binding, calcium binding, PDZ-binding-motif and more[24]. This reduced the number of prioritized splice events to 12,145 events in 5,424 genes with a putative functional effect. Among these predicted functional spliced sites illustrated in **Figure 3B**, the majority favored increased exon inclusion (N = 12,337), with smaller subsets favoring increased exon skipping (N = 2,380) or displaying a mixed pattern (N = 662). Shown in **Figure 3C**, these functional sites included changes to disulfide bonding (N_{EI} = 95, N_{ES} = 915), localization signaling (N_{EI} = 93, N_{ES} = 426), amino acid modification (N_{EI} = 489, N_{ES} = 4,093), and other functional sites (N_{EI} = 1,909, N_{ES} = 8,015). Each event is listed in **Table S4**. The

remaining splicing events were associated with un-annotated sites, untranslated, and/or non-coding regions. To identify potentially targetable events, we selected functional splice events in kinases and performed over-representation pathway analysis which revealed MAPK, ERBB, and PI3K-AKT MTOR as the top cancer-related pathways significantly over-represented (Bonferroni-adjusted p-value < 0.05, **Figure S3B-C**).

Strikingly, the gene encoding protein kinase *CDC Like Kinase 1 (CLK1)*, an oncogenic factor and known master modulator of alternative splicing[12], was amongst this subset of differentially spliced kinase genes in HGGs (**Figure 3D**). The majority of tumors showed very high levels of *CLK1* exon 4 inclusion (**Figure 3E**) with mean PSI of 0.7657 (or 76.57%) and thus the inclusion event was not differential in most HGGs. In contrast, 12 tumors demonstrated significant skipping (decreased inclusion) of exon 4 (ΔPSI <= -2 z-scores of mean PSI), thus driving the differential splicing observed for *CLK1* exon 4 (**Figure 3D**). Additionally, we observed *CLK1* exon 4 associated transcript expression heterogeneity across non-tumor brain controls from the Genotype Tissue Expression Project (**Figure S3C**). *CLK1* regulates the SR (Serine aRginine) family of splicing factor proteins through hyper-phosphorylation of the SR-rich peptide regions of SR proteins to induce cooperative RNA binding and increased activity[25–27]. Moreover, the differential splicing of *CLK1* resulted in differences of exon usage across HGGs (**Figure 3E-F**). *CLK1* exon 4 contains two catalytic sites, Thr138 and Ser140 (**Figure 3G**), and these have been described previously to be associated with increased protein abundance[26]. *CLK1* exon 4 inclusion was significantly positively correlated with expression of total *CLK1* mRNA (Pearson's R = 0.29, p-value = 4.1e-5, **Figure 3H**), supporting the hypothesis that inclusion of these phosphorylation sites in exon 4 increases canonical *CLK1* expression. Further, *CLK1* expression is significantly positively correlated with expression of *Serine/Arginine-rich protein-specific kinase 1* (*SRPK1*), a kinase that cooperates with SR-bound *CLK1* to facilitate SR phosphorylation, U1 exchange of *CLK1*, and subsequent splicing[28] (**Figure 3I**, Pearson's R = 0.69, p-value < 2.2e-16). High *CLK1* exon 4 inclusion was not unique to HGGs though they

were the most heterogeneous. Indeed, we observed widespread high median *CLK1* exon 4 inclusion levels across pediatric brain tumors, suggesting these tumors contain active *CLK1* (**Figure 3J**).

CLK1 is an oncogenic dependency in pediatric HGGs

We sought to further examine the role of *CLK1* as a potential oncogene in HGGs. We investigated the cancer Dependency Map (DepMap) portal and database and found that CNS and brain tumor cell lines with high expression of the exon 4 included transcript of *CLK1* (≥ third quantile mRNA expression of ENST00000321356) have significantly higher CRISPR dependency (lower scores) compared to *CLK1* low expressing cell lines (≤ first quantile) (Wilcoxon p = 0.034, **Figure 4A**). This observation was significant and unique only to cell lines derived from CNS tumor and myeloid malignancies (**Figure S4A**), suggesting tissue- and tumor-specific regulation of *CLK1*. Across all DepMap profiled cell lines, we found that the pediatric brain tumor cell line KNS-42 had a strong dependency on *CLK1* (low CRISPR dependency score) (**Figure 4B**) and chose it for further *in vitro* testing. We next tested the impact of CLK1 inhibition in KNS-42 cells using the pan-Dyrk/Clk inhibitor Cirtuvivint (SM08502)[29]. Using the IncuCyte to monitor real-time proliferation, we observed a significant reduction in cell growth at multiple concentrations over a 6-day period (**Figure 4C**). Additionally, we observed a dose-dependent decrease in cell viability using CellTiter-Glo at three days (**Figure 4D**) and six days (**Figure S4B**) post-treatment of 0.5, 1, 5, and 10 μM Cirtuvivint.

Based on these findings and to rule out off-target effects of the pan-inhibitor, we selected KNS-42 along with two additional cell lines from our pediatric brain tumor cohort with high *CLK1* exon 4 PSI (7316-1763 and 7316-1769) to experimentally validate the exon 4 splice event identified from short-read RNA-Seq. We performed long-read RNA-seq using Oxford Nanopore Technologies (ONT) and validated the presence of two major full-length *CLK1* mRNA isoforms that either included or skipped exon 4 across these three patient-derived cell lines (**Figure 4E**).

We therefore postulated that the gain of *CLK1* phosphorylation sites on exon 4 increases mRNA and subsequent protein production in HGGs. To directly test this hypothesis, we modulated *CLK1* exon 4 splicing using targeted morpholino oligomers (see **Methods**), in which we forced exon 4 skipping in the KNS-42 cell line. We performed qRT-PCR and observed a near total loss of the *CLK1* exon 4 inclusion transcript at both 5 and 10 μM of exon 4 targeted morpholino, evidenced by reduced expression of the exon 3-4 junction. Likewise, at these same concentrations, we observed increased *CLK1* exon 4 skipping using primers targeting the exon 3-5 junction (**Figure 4F**). Importantly, forced *CLK1* exon 4 skipping resulted in ablation of CLK1 protein at 5 and 10 μM (**Figure 4G),** corroborating previous work that *CLK1* exon 4 is required for full-length and catalytically active *CLK1*[30–32]. Next, we assessed the functional impact of *CLK1* exon 4 splicing using CellTiter-Glo and confirmed that cells with high *CLK1* exon 4 skipping (*CLK1* exon 4 targeting morpholino) exhibited significantly decreased viability compared to those with *CLK1* exon 4 inclusion (non-targeting morpholino) at 24, 72, and 96 hours ($p \leq 0.01$, within-time Student's *t-test*, **Figure 4H**). Taken together, we demonstrate that *CLK1* is a dependency in pediatric HGGs required for cellular growth and viability and CLK1 mRNA and protein is maintained through increased exon 4 inclusion.

To identify bona-fide *CLK1* targets mediated by exon 4 splicing, we performed RNA-seq from KNS-42 cells treated with morpholino oligomers (N = 3 controls, N = 3 targeted to skip exon 4). We performed differential gene expression (DE) analysis and identified 296 genes with differential expression (193 upregulated, 103 downregulated) between the treated and untreated populations (**Figure 4I, Table S5A**). Next, we quantified differential alternative splicing (**Methods**, **Table S5B**) and applied the same downstream computational workflow from **Figure S3A** to prioritize splice events affecting functional sites. We identified a total of 2,006 unique differential splicing (DS) events within 1,467 genes predicted to alter functional sites (SE = 1,905, A5SS = 196, A3SS = 272, and RI = 388, **Figure 4J** and **Table S5C-F**). These dysregulated genes included TSGs and oncogenes involved in RNA-binding, epigenetics,

transcription factors, and kinases (**Figures 4K-L** and **Table S5G**). These genes were over-represented in G2M checkpoint, mitotic spindle, and nucleotide excision repair pathways (**Figure 4M**). To further investigate the impact on DNA repair and other pathways, we performed gene-set enrichment analyses of DNA repair and cancer signaling pathways on these DS oncogenes and TSGs and found that depletion of CLK1 leads to upregulation of TNFA, PI3K/AKT/MTOR, IL6/JAK/STAT3, and apoptosis pathway expression and downregulation of multiple DNA repair pathways (**Figure S4C-F**). Moreover, of the cancer genes with putative functional consequences driven by *CLK1* splicing, we discovered that 2.3% (N = 6) had dysregulation at the level of both splicing and expression (**Figure 4N**), indicating these may impact the tumor's proteome. The DE genes were significantly over-represented (Bonferroni-adjusted p-value <0.05) for KRAS upregulation and drug metabolism pathways (**Figure S5A**) while DE genes were significantly over-represented (**Figure S5B**) in G2M checkpoint, mitotic spindle, and nucleotide excision repair pathways, suggesting a potential for these events to impact cellular functions, contribute to the cancer disease state, and/or play a role in regulatory mechanisms of gene expression.

Finally, we asked whether *CLK1* splicing affects any of the essential oncogenes defined by the pediatric gene dependency maps of the Childhood Cancer Model Atlas[33]. We observed 15 of these genes also exhibit significant gene dependencies (GD) in established pediatric HGG cell lines (**Figure 4O-P, Table S5B**), including *CDK4*, *FGFR1*, *FGFR2*, *EZH2*, *RAF1*, and *SRC*. For instance, the expression levels of mRNAs encoding proto-oncogene *SRC*[34,35] are higher in cells with high *CLK1* exon 4 (non-targeting morpholino), indicating that *CLK1* may enhance or promote *SRC* expression. The differential splicing cases are more complex as they affect multiple transcripts, but taken together, these data suggest that transcript-level changes mediated through *CLK1* could be contributing to some of these dependencies, particularly given the association of aberrant splicing with cancer progression[36–38].

Discussion

Pediatric brain cancer remains the leading cause of disease-related mortality in children, and HGGs present formidable challenges due to their resistance to chemotherapy and surgical limitations. In this study, we conducted a large-scale analysis of aberrant alternative splicing across pediatric CNS tumors, revealing widespread dysregulation of pre-mRNA alternative splicing. We developed an analytical framework to prioritize and predict the consequences of splicing events. Our study demarcates intricate splicing patterns across various tumor types and introduces the splicing burden index (SBI) as a novel metric to quantify differential splicing events at the sample level without requiring a normal control. We discovered significant negative correlations between tumor mutation burden (TMB) and SBI in CPT, LGG, GNT, and schwannoma, suggesting that aberrant splicing may serve as a compensatory mechanism for tumorigenesis in these mutationally silent tumors. Further analysis of skipped exon events unveiled novel biological clusters driven by splicing variations across histologies and molecular subtypes, highlighting extensive splicing heterogeneity in pediatric HGGs. Of note, molecular subtypes were not cluster-specific and specifically for MB, this differs from a previous report which showed that subgroups WNT, SHH, Group 3, and Group 4 can be clustered using splicing information[39]. This is due to fundamental differences in clustering procedure and study goals. We performed unsupervised clustering to identify novel groupings, whereas Dubuc and Morrissy, et. al performed supervised clustering following identification of differentially expressed splice events among the four subgroups in order to classify groups using splice events[39]. Utilizing the UniProt Knowledgebase, we identified splice variants in HGGs that alter functional sites, potentially impacting protein functions such as activation, folding, and localization.

A key finding was differential splicing of *CDC-like kinase 1 (CLK1)*, a critical splicing factor and cell-cycle modulator in pediatric HGGs. This splicing event led to the inclusion of phosphorylation sites in exon 4, promoting increased protein abundance. Experimental

modulation of *CLK1*, either through inhibition or morpholino-directed exon 4 depletion in the KNS-42 cell line, resulted in significantly reduced cell proliferation and/or viability. Splicing modulation to deplete exon 4 ablated CLK1 RNA and protein levels, altogether supporting *CLK1* as a gene dependency in pediatric HGGs. Additionally, we identified transcriptional dysregulation of essential cancer genes mediated by aberrant *CLK1* splicing. *CLK1* is currently being targeted therapeutically with the Pan-Clk/Dyrk Inhibitor Cirtuvivint (SM08502) in heme malignancies[29] and non-CNS solid tumors such as castrate-resistant prostate cancer, colorectal cancer, and non-small cell lung cancer[40–42]. Here, our study suggests that *CLK1* may also represent a therapeutic vulnerability in CNS malignancies including pediatric HGGs.

This investigation enhances our understanding of the splicing landscape in pediatric brain tumors and proposes that aberrant splicing may be a viable target for therapeutic intervention. Further, we openly share the splicing data for all pediatric CNS tumors and believe this can be a valuable resource for the oncology community. Our approach to characterizing splicing aberrations and their functional consequences paves the way for future research into mRNA splicing-based mechanisms of tumorigenesis, the identification and development of therapies targeting aberrant splice events, and may even guide splicing-based diagnostics, all of which have the potential to improve the therapeutic landscape for pediatric brain cancers.

Limitations of the study

In this study, splicing quantifications were primarily performed using short-read RNA-Seq technology, which limits the interpretation of the full spectrum of splicing variation, particularly larger multi-exon transcripts or with genes that contain a high number of transcripts. While there are proteomics data for over 200 matched pediatric brain tumors publicly available, we were limited in sample size for DIPG or DMG and other HGGs, so it will be important to validate our findings in larger datasets as they become available. Additionally, the lack of pediatric normal tissue RNA-Seq necessitated use of non-tumor controls restricted to adult samples from GTEx,

and may not necessarily represent the splicing landscape of pediatric tissues. Further, GTEx does not contain tissue of origin for all pediatric brain tumors (eg: pons for DIPG or brainstem for DMG), further limiting the comparison. Additionally, within histologies (eg: LGG), the primary site of the tumor can vary widely depending on diagnosis and it would be ideal to match each tumor to a normal one by one. Although this is not yet possible with the normal RNA-Seq available, the creation of the upcoming developmental GTEx will be critical in the future. We mitigated these normal tissue limitations through the use of the SBI metric, in which we compared each tumor to every other tumor in the cohort. Finally, this RNA-Seq cohort contained samples with many different library preparation strategies (poly-A, stranded, exome capture, stranded poly-A), which ultimately leads to batch effects in certain expression analyses. Using two samples sequenced by different library preparation strategies, we demonstrated that SBI was largely unaffected by library type (**Figure S1A-B**). Therefore, to mitigate batch effects with gene expression measurements, we used the entire pooled cohort for splicing analyses (PSI, SBI) but used only stranded samples in other analyses involving gene expression values such as correlations, differential expression, and/or over-representation analyses. Overall, we employed robust statistical techniques, cross-validated our findings with external datasets, and utilized orthogonal approaches and experimental methods where possible.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgements

We graciously thank all of the patients and families of the Children's Brain Tumor Network (CBTN) and the Pacific Pediatric Neuro-oncology Consortium for donating tissue which enabled this research. The authors thank the following collaborators who contributed experiments, analyses, code, and/or code review that were ultimately not included in the

manuscript: Poonam Sonawane, Cullen Wilson, Krutika S. Gaonkar, and Run Jin. This work was funded in part by the National Cancer Institute (R03OD036498 to JLR), the Chad Tough Foundation, the Children's Hospital of Philadelphia Division of Neurosurgery, and the CBTN.

Author Contributions

Conceptualization: ASN, JLR, ACR, RJC

Methodology: ASN, JLR, RJC, ACR, KSR, AK, SA, KLC

Software: BE, KSR, AK

Validation: PS, KC, RJC, KLC, KEH, ASN, AL

Formal Analysis: ASN, JLR, RJC, KLC, KSR, BE

Investigation: ASN, JLR, RJC, KSR

Resources: JLR, ACR, ATT, PBS, CB, SJD

Data Curation: JLR

Writing - Original Draft: ASN, JLR, KSR, PS, KLC

Writing - Review & Editing: ASN, JLR, RJC, SJD, KLC, JBF, JMD

Visualization: ASN, JLR, KSR, RJC, KLC, PS

Supervision: JLR, ACR, ATT

Project Administration: JLR, ACR

Funding Acquisition: JLR, ACR, PBS, ATT, SJD

STAR Methods

RESOURCE AVAILABILITY

Lead Contact

Requests for access to raw data and/or specimens may be directed to and will be fulfilled by Jo Lynne Rokita (rokita@chop.edu).

Materials availability

This study did not create new, unique reagents.

Data and code availability

PBTA patient genomic data

All pediatric brain tumor raw data are available upon request from the database of Genotypes and Phenotypes (dbGAP), accession number phs002517.v2.p2, and/or from the Children's Brain Tumor Network (https://cbtn.org) and the Pacific Pediatric Neuro-Oncology Consortium (pnoc.us) for data not immediately available in dbGaP. All processed data used in this study were derived from the OpenPedCan project[16] v13 data release at https://github.com/d3b-center/OpenPedCan-analysis. All code for the manuscript analyses and figures are openly available at https://github.com/d3b-center/pbta-splicing.

GTEX non-tumor tissue control RNA-Seq

We utilized OpenPedCan[16] release v13 processed RNA-Seq data from the Genotype Tissue Expression (GTEx) project which had harmonized gene symbols to GENCODE v39 using the custom script at: https://github.com/d3b-center/D3b-DGD-Collaboration/blob/main/scripts/update_gene_symbols.py.

CLK1 morpholino RNA-Seq

RNA-sequencing data from the *CLK1* morpholino experiment has been deposited in GSE273841.

Merged primary and summary data

Merged primary matrices and summary files utilized in this manuscript were derived from are openly accessible via the download script in the https://github.com/d3b-center/pbta-splicing repository. To compare RNA-Seq from *CLK1* exon 4 morpholino-treated cells vs control morpholino-treated cells, we ran rMATs with three biological replicates for each condition `--b1 –b2`. This paired mode analysis calculated ΔPSI, p-values, and FDR statistics for each splice event.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Study participants

Study participants include pediatric brain tumor patients whose genomic data was deposited into, and obtained from, the OpenPedCan[16] project.

Patient-derived cell line models

The high-grade glioma patient-derived cell lines 7316-1736 and 7316-1739 were obtained by CBTN request and the KNS-42 cell line was obtained commercially as noted in the Key Resources table.

METHOD DETAILS

Primary data analyses

Somatic primary workflows were implemented by the Kids First Data Resource Center as described in the Open Pediatric Brain Tumor Atlas (OpenPBTA)[43] and OpenPedCan[16] projects. The code for these workflows, including RNA-seq quantification, fusion identification, RNA splicing, and SNV, INDEL, CNV, SV calling, can be found at https://github.com/d3b-center/OpenPedCan-workflows. Sample-level data can be found through the Kids First Portal at https://kidsfirstdrc.org/.

CLK1 exon 4 related visualizations and correlations

To visualize the *CLK1* exon 4 splice event, we utilized the R package *ggsashimi*[44]. We correlated *CLK1* exon 4 PSI values with *CLK1-201* or *CLK1* TPM. We computed Pearson correlation coefficients and p-values of this plot using the R package *ggpubr*[45]. High *CLK1* exon 4 inclusion tumors were defined as those with PSI values above the 75th percentile, while low SBI samples were those with PSI values below the 25th percentile comparing across all samples.

Protein Visualizations

Protein visualizations were obtained from the PhosphoSitePlus web portal[46], emphasizing protein domains, residue numbers, and sites of phosphorylation binding.

Cell Culture

The pediatric HGG cell line KNS-42 was cultured in DMEM-F12 (GIBCO, 11320033) supplemented with 10% FBS (GIBCO, 26140079), 2 mmol/L L-glutamine (GIBCO, 25030081), and 1X penicillin/streptomycin (GIBCO, 15140122) at 37°C and 5% CO_2. The cell line was authenticated by Guardian Forensic Sciences (Abington, PA) using the GenePrint 24 (Promega, B1870) short tandem repeat kit. Cells tested negative for mycoplasma using the EZ-PCR

Mycoplasma Detection Kit (Biological Industries, 20-700-20) and were used for a maximum of 12 passages post thaw.

Morpholino Treatments

A Vivo-Morpholino ACTCTTCTGGAAACGTCAAGTGGGC (Gene Tools, LLC) targeting the intron 3-exon 4 splice junction was used to skip exon 4 in *CLK1*. Cells were treated with 1, 5, and 10 μM concentrations of *CLK1* morpholino and 10 μM of Control morpholino. 48 hours post-treatment, cells were harvested for PCR and immunoblots.

RNA Extraction and Quantitative Real-time PCR (qRT-PCR)

Total RNA was isolated and treated with DNAse using the Maxwell RSC simplyRNA Cells kit (Promega, AS1390) with the Maxwell RSC48 Instrument (Promega) per the manufacturer's instructions. Next, 2 μg of RNA were reverse-transcribed using SuperScript IV (Invitrogen, 18090010). Primers used for *CLK1* mRNA transcript quantification are listed in **TableS5H**. qRT-PCR was performed using PowerSYBR Green PCR Master Mix (Invitrogen, 4367659) on an Applied Biosystems Viia7 machine. The amplification was performed using the following settings: denaturation at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 s and annealing at 60°C for 1 min. The comparative cycle threshold (CT) method was applied to quantify the expression levels of *CLK1*. The fold change of gene expression was calculated by the equation 2ΔΔCT, with *HPRT* (Thermo Fisher, 4453320, assay ID: Hs02800695_m1) used as the housekeeping gene.

Protein Extraction

Cultured cells were washed once in chilled D-PBS (pH 7.4) and lysed in RIPA buffer containing 50 mM Tris□HCl, pH 7.4, NP□40 (1%), deoxycholate (0.25%), 150 mM NaCl, 1 mM EDTA pH 8.0, 1x protease and phosphatase inhibitor cocktail (Pierce Halt Inhibitor Cocktail,

Thermo Fisher Scientific, 78446), and SDS (0.1%). Total protein in the lysate was estimated by the DC Protein assay (BioRad Laboratories, 5000111).

Detection of Proteins Using Immunoblot Analysis

70 µg of total protein were mixed with 5X SDS loading dye (Biorad, 161-0374) and resolved on 10% SDS polyacrylamide gel. The protein was transferred onto a PVDF membrane (Immobilin-P, Millipore, IPVH00010) and probed with α-CLK1 mouse monoclonal primary antibody (Santa Cruz, sc-515897) and HRP conjugated secondary antibody (Cell Signaling Technology, 7076S). Bands were detected using enhanced chemiluminescence (Millipore, WBKLS0500) and captured by a Chemiluminescence imager (GE Healthcare). β-actin was used as the loading control and probed with α-β-actin rabbit monoclonal antibody (Cell Signaling Technology, 12262S).

Cell Viability Assay

Cell viability was measured using the CellTitre-Glo (CTG) luminescent cell viability assay (Promega, G7570). Cells were seeded in white 96-well flat-bottom plates at a density of 24,000 cells per well and treated the following day with either 7.5 µM control or CLK1 exon 4 targeted morpholino. Luminescence was measured using a Biotek Synergy 2 plate reader at 24, 48, 72, and 96 hours.

pan-DYRK/CLK1 inhibitor Cirtuvivint (SM08502) experiments

The KNS-42 cell line was cultured in DMEM-F12 (Gibco, 11330032) supplemented with 10% FBS (Corning, MT3501CV, lot 003322001) and additional L-glutamine (Thermo Fisher, 25030081) to a final concentration of 4.5 mM. Dissociation was performed with Trypsin-EDTA (0.05%, Thermo 25300054) and counted on a DeNovix Cell Drop cell counter.

For growth kinetics, 10,000 (3 day assay) or 6,000 (6 day assay) cells were plated per well into a 96-well plate (Greiner Bio-One, 655098) in a 200 uL total volume per well. Plates were placed into an Incucyte SX5 device and scanned every 2 hours for several days to measure growth via a mask designed uniquely for this cell type. At the end point of the assay, cell viability was analyzed with CellTiter Glo 2.0 reagent (Promega, G9242) by replacing half the media with reagent and reading on a Promega GloMax device.

Cirtuvivint (MedChem Express, HY-137435) was resuspended in 100% DMSO (Sigma, D2650-5X5ML) to 1 mM and stored in aliquots at -80 C. Dosing was optimized via serial dilution at a range of 20 uM to 0.02 uM against a vehicle control equivalent to the highest dosing of drug. Cells were plated and at 24 hours, 100 uL of media were removed from each well and replaced with drug media for a final dose range of 0.01, 0.05, 0.5, 0.1, 1, 5, and 10 uM. Cells were untouched for 3 days total while growth was monitored via Incucyte.

QUANTIFICATION AND STATISTICAL ANALYSIS

Splicing identification and quantification

To detect alternative splicing, we ran rMATS turbo (v. 4.1.0)[14] with GENCODE v39 GFF annotations, as described by the Kids First RNA-Seq workflow (https://github.com/d3b-center/OpenPedCan-workflows). We filtered for alternative splicing events with ≥ 10 junction read counts. These results were then used for all downstream processing throughout the manuscript.

Splicing burden index (SBI) calculation

The following describes the SBI calculation used in the manuscript.

Let X be the list of all samples, where X_{ij} represents the j-th item in the i-th sample.

Let n be the number of items in each sample.

Let SE be the splice event of interest.

Let SE_i be the number of splice events in the i-th sample.

Let $mean_{SE}$ be the mean of the splice event across all samples.

Let σ_{SE} be the standard deviation of the splice event across all samples.

Let SBI be the proportion of splice events that have z-scores > |2| out of the total number of splice events in a particular sample.

Then the equation for SBI is:

$$SBI_i = \frac{\sum_{j=1}^{n} I\left(\left|\frac{X_{i,j} - mean_{SE}}{\sigma_{SE}}\right| > |2|\right)}{n \cdot len(X)},$$

where i = 1 to len(X) and j = 1 to n

We compared PSI values of each primary tumor against all other tumors in the cohort. We first computed mean and standard deviation metrics for each alternative splicing event observed in at least one sample. Then for each sample in each group or histology, we identified the proportion of genes that underwent aberrant splicing as defined by a z-score > |2| across the entire transcriptome that undergoes alternative splicing.

Consensus clustering

We first preprocessed the splicing PSI matrix, restricting it to one splice event per gene by choosing the splice event with the maximum PSI value in a given gene. To reduce the dimensionality of the input matrix, we applied a feature selection using Hartigans' dip test[47]. This test identifies dips in the distribution of input features and selects features that have a bi- or multi-modal distribution across the input samples. These “dips” in the distribution may

correspond to differences within underlying clinical variables of interest. A total of 6999 features passed the test for multi-modality (p-value < 0.05) and were used for downstream clustering. Next, we applied all combinations of the following clustering algorithms (PAM, K-means, and Hierarchical) and distance methods (Pearson, Spearman, Euclidean, Manhattan, Binary, Maximum, Canberra, and Minkowski) available in the R package *ConsensusClusterPlus*[48]. For each combination, we evaluated a minimum *k* value of 2 and a maximum k value of 17. This resulted in a total of 272 clustering solutions corresponding to the different input combinations.

To identify the optimal clustering solution, we first evaluated the cluster performance using the R package *fpc*[49]. Using a given input data matrix and clustering solution, the function `fpc::clusterstats` computes the metrics silhouette width, entropy, purity, and Dunn index that represent separation between different clusters and closeness of data points within a cluster. The silhouette score defines the compactness of individual clusters (intra-cluster distance) and separation amongst clusters (inter-cluster distance) to measure an overall representative score. The entropy and purity evaluate the stability of the cluster. The higher the purity, the more stable the cluster is and the smaller the entropy, the better the clustering performance. The Dunn index is the ratio of the smallest inter-cluster distance and the largest intra-cluster distance. A higher Dunn Index will indicate compact, well-separated clusters, while a lower index will indicate less compact or less well-separated clusters. We used the R package *COINr* to assign weights to each metric and compute a composite score representing the overall “cluster quality”. The composite score was calculated by assigning a directional and weighted scoring mechanism. Ranks were assigned to each evaluated combination, with the highest composite score or cluster quality being assigned a rank of 1, 2, etc. For our dataset, the highest cluster quality, i.e. top ranking method, was the combination of PAM clustering algorithm with Canberra distance measure and a value of 12.

Clustering-based differential expression or pathway enrichment

We identified differentially expressed genes per cluster of interest and conducted pre-ranked pathway enrichment using *limma*[50], *fgsea*[51], and *GSVA*[52] on those genes. We interrogated KEGG spliceosome and HALLMARK cancer pathways. We visualized these clusters using the R package *pheatmap*[53] labeling rows with histology and calculated cluster information.

Differential expression and visualization

Differential expression was performed based on a model using the negative binomial distribution, a method employed by the R package *DeSeq2*[54]. Those differential genes that had a p-value < 0.05 were deemed as significantly up or down-regulated. Volcano plots were generated by the *EnhancedVolcano* R package. Bar plots were generated using the R package *ggplot2*[55]. Note: differential expression analyses were limited to stranded-only RNA-seq samples in order to limit batch effects.

Identification of recurrent functional differential splicing variants in pediatric HGGs

To identify differential or aberrant alternative splicing events, we assessed the percent spliced in (PSI) value of each splice event relative to the median PSI value of splice event across all samples. Splicing events with a ΔPSI exceeding |2| z-scores from the median PSI value were classified as differential or aberrant. For these events, we computed average ΔPSIs and generated bed files for each mis-spliced exon event. We then obtained bed files of known functional annotations as defined by Uniprot release 2024_03[24] from UCSC Genome Browser web server. We ran bedtools v2.30[56] to find the overlap between mis-spliced exons and functional features using the command `bedtools intersect -wo -a`. We then plotted summary data by functional category (disulfide bonding sites, localization signals, amino acid modifications, and other).

Upset R and Volcano plots

To visualize the intersections of multiple sets, we employed the UpSetR[57] plot in R. The input data consisted of differential and recurrent splicing events, if it was > 2 z-scores from the meanPSI and 2% of the histology-specific cohort. Volcano plots were generated by the *EnhancedVolcano* R package.

Splicing burden index and tumor mutation burden correlations

We identified samples with available data for both SBI (RNA-Seq) and WGS or WXS tumor mutation burden (TMB) from OpenPedCan[16]. Using the R package *ggscatter*, we performed a Pearson correlation analysis to examine the relationship between SBI and TMB. To ensure robustness, we repeated this analysis after excluding hyper-mutated samples (defined as those with TMB ≥ 10). Subsequently, we compared the distribution of TMB between high SBI and low SBI tumor samples using the Wilcoxon rank-sum test. High SBI samples were defined as those with SBI values above the 75th percentile, while low SBI samples were those with SBI values below the 25th percentile. The analyses were conducted across all samples and further stratified according to `plot_group`, as specified in the histologies clinical file.

Pathway over-representation analysis (ORA) and gene set variation analysis (GSVA)

We conducted over-representation analysis (ORA) using the R package clusterProfiler[58] and pathway data from the msigdbr package[59], including “CP:KEGG", "CP:BIOCARTA", “CP:HALLMARK”, and “TFT:GTRD.” After inputting the genes of interest (eg. differentially spliced), we applied a p-value cutoff of 0.05 and used the Benjamini-Hochberg (BH) method for p-value adjustment. For visualization of the over-represented pathways, we employed the `enrichplot::dotplot()` function, displaying the gene ratio and the count of genes in each pathway.

To perform Gene set variation analysis (GSVA) we utilized the R packages `GSVA` and `msidbr`. Expression data for our samples, sourced from OpenPedCan v13[16], were used to compute gene-set enrichment scores. Genes with zero variance were excluded from the analysis. We then assessed enrichment in Hallmark, KEGG, and custom pathways from Knijnenburg et al[60]. Gaussian-distributed scores were calculated using *gsvaParam* function in R. The results were visualized using heatmaps of GSVA scores, generated with the R packages *ComplexHeatmap* and *circlize*.

Oxford Nanopore Technologies (ONT) Targeted Long-Read RNA-Sequencing

We designed primers to bind to exons present in all isoforms of *CLK1* to ensure full coverage of all alternative splicing events. 5 ng of cDNA were amplified with LongAmp Taq 2X Master Mix (M0287S, New England Biolabs) for 25 cycles. The resulting amplicons were subjected to amplicon-seq (SQKNBD112.24, ONT) library preparation, loaded into a Spot-ON flow cell R9 Version (FLO-MIN112, ONT), and sequenced in a MinION Mk1C device (ONT) until at least 1,000 reads per sample were obtained. Results were aligned using Minimap2 version 2.24-r1122 and visualized in IGV version 2.12.3.

DepMap and CRISPR dependency analyses

Datasets comprising gene transcript expression, cell line information, and CRISPR dependency scores were downloaded from DepMap (version 24Q2). The expression of *CLK1* ENST00000321356 (exon 4 containing transcript) was categorized into high and low TPM expression, defined by values above the 75th quantile and below the 25th quantile, respectively. CRISPR dependency scores were plotted on the y-axis, and Wilcoxon tests were conducted to compare high versus low TPM expression groups. These were stratified for each cell line type. Additionally, CRISPR dependency scores for all CNS/brain cell lines were plotted, with KNS-42 highlighted in red. For the Childhood Cancer Model Atlas CRISPR dependency analyses, we

acquired data from the Childhood Cancer Model Atlas[33]. We plotted CRISPR dependency scores (z) on the y-axis for each gene in CBTN pediatric HGG cell lines, either as median scores or stratified by individual patients with genes of interest highlighted.

Proteogenomic analysis

Pediatric proteomics, phosphoproteomics, and RNA data were obtained from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) via the ProTrack: Pediatric Brain Tumor open-source web portal. Data and z-scores were computed using the methods described by Petralia et al.[22]. Correlation plots of mRNA expression and proteomics were generated using these computed z-scores. Additional processed proteogenomic data utilized for correlation analyses were obtained from OpenPedCan v15 release[16].

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
α-CLK1 (F-12) mouse mAb	Santa Cruz	sc-515897
α-β-Actin (8H10D10) Mouse mAb (HRP Conjugate)	Cell Signaling Technology	12262S
Anti-mouse IgG, HRP-linked Antibody	Cell Signaling Technology	7076S
Drugs		
DMSO	Sigma	D2650-5X5ML
Cirtuvivint (SM08502)	MedChem Express	HY-13743
Critical commercial assays		
GenePrint 24	Promega	B1870
EZ-PCR Mycoplasma Detection Kit	Biological Industries	20-700-20
Maxwell RSC simplyRNA Cells Kit	Promega	AS1390
SuperScript IV	Invitrogen	18090010
CellTitre-Glo luminescent cell viability assay	Promega	G7570

Deposited data		
PBTA raw WGS, WXS, Panel, RNA-Seq	dbGAP	phs002517.v2.p2
PBTA harmonized WGS, WXS, Panel, RNA-Seq	OpenPedCan[16]	https://github.com/d3b-center/OpenPedCan-workflows
PBTA merged summary files and downstream analyses	This project	10.5281/zenodo.13362856
Morpholino treated KNS42 RNA-seq	This project	GSE273841
Experimental models: Cell lines		
PBTA patient-derived cell lines 7316-1763 and 7316-1769	Children's Brain Tumor Network	Children's Brain Tumor Network
KNS-42 human glioma cell line	Accegen	ABC-TC0532
Oligonucleotides		
CLK1 intron 3-exon 4 splice junction targeting morpholino - ACTCTTCTGGAAACGTCAAGTGGGC	Gene Tools, LLC, This Project	
CLK1 Ex3-Ex4 (inclusion of Exon4) F- GGACATCGCCAAAGAGACCA	Integrated DNA Technologies IDT, This Project	
CLK1 Ex3-Ex4 (inclusion of Exon4) R- TCCTTCGGTGACTCTTCCCA	Integrated DNA Technologies IDT, This Project	
CLK1 Ex3-Ex5 (inclusion of Exon4) F- ATCGCCAAAGAGACCATGAAAG	Integrated DNA Technologies IDT, This Project	
CLK1 Ex3-Ex5 (inclusion of Exon4) R- GTATCAACAATTTCATCCCATGTGA	Integrated DNA Technologies IDT, This Project	
CLK1 Ex3-Ex5 (inclusion of Exon4) F- CCATGAAAGCCGGTATCAGAAC	Integrated DNA Technologies IDT, This Project	
CLK1 Ex3-Ex5 (inclusion of Exon4) R- ACCTAAAGTATCAACAATTTCATCCCA	Integrated DNA Technologies IDT, This Project	
Software and algorithms		
Replicate Multivariate Analysis of Transcript Splicing (rMATS)	https://github.com/Xinglab/rmats-turbo	rMATS-turbo: an efficient and flexible computational tool for alternative splicing analysis

		of large-scale RNA-seq data[14]
ggsashami	https://github.com/guigolab/ggsashimi	ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization[44]
OpenPedCan analysis repository	OpenPedCan[16]	https://github.com/d3b-center/OpenPedCan-analysis
Project repository	This Project	https://github.com/d3b-center/pbta-splicing

Figure Legends

Figure 1: Pediatric brain tumors display heterogeneous global patterns of aberrant splicing

(A) Circos plot of CNS tumors used in this study, categorized by histology, tumor location, and reported gender. Non-neoplastic tumors consist of benign tumors and/or cysts. **(B)** Lollipop plot illustrating the total number of splicing events across the cohort, classified by splicing type (SE: single exon, RI: retain intron, A3SS: alternative 3' splice site, A5SS: alternative 5' splice site). **(C)** UpsetR plot of recurrent differential splicing events that prefer exon skipping (N ≥ 2 of samples within a histology). **(D)** Barplots of the number of histology-specific recurrent events per patient. Histologies are reverse ordered by total number of unique events (skipping + inclusion). **(E)** Cumulative distribution plots of splicing burden index (SBI) by histology. **(F)** Barplots displaying percent of tumors with high (≥ third quantile) and low (≤ first quantile) SBI in each histology. **(G)** Boxplots of tumor mutation burden (TMB, $\log_{10}$) stratified by high or low SBI by

histology. Within-histology Wilcoxon p-values are shown. All boxplots represent the 25th and 75th percentile and the bar represents the median.

Figure 2: Splicing drives novel biological clusters and splicing burden differentiates key splicing factors in pediatric high-grade gliomas

(A) Consensus clustering heatmap of PSI values for all expressed genes (junction read counts ≥ 10) with a multi-modal distribution across tumors (see Methods). **(B)** Stacked barplot showing histology sample membership in each cluster. **(C)** Stacked barplot of the number of tumors with high or low SBI within each cluster. **(D)** Heatmap of top cancer-related enriched pathways by cluster (GSVA scores represented by blue/orange color). **(E)** Pearson's correlation scatterplot of $\log_2$ SBI and KEGG Spliceosome GSVA score (R = 0.55, p-value < $2.2e^{-16}$). **(F)** Volcano plot illustrating the expression direction of splicing factor genes in HGGs with high SBI compared to those with low SBI (NS = not significant, FC = fold change, colored dots represent $\log_2$FC > |.5| and/or Benjamini and Hochberg adjusted p-value < 0.05). **(G)** Barplot presenting members of the hnRNP and SRSF families of primary splicing factors that are differentially expressed (Benjamini and Hochberg adjusted p-value < 0.05) with directionality (+ or -). **(H)** Heatmap displaying these splicing factors and known regulators *CLK1* and *SRPK1* for all pediatric brain tumors with available RNA-seq and whole-cell proteomics data from the CPTAC portal.

Figure 3: Recurrent splicing aberrations alter known proteomic functional sites in pediatric high-grade gliomas, including phosphorylation sites in splicing regulator protein kinase *CLK1* (A) Workflow to identify 12,145 differential exon-level splicing events that alter UniProt-defined functional sites in HGGs. **(B)** Stacked bar plots showing the fraction of exon inclusion, skipping, or mixed splicing events categorized by predicted impact. **(C)** Boxplots of splice events resulting from gain or loss of functional sites categorized by UniProt annotation. Wilcoxon between-group p-values are shown. **(D)** Boxplots of predicted functional splice events

affecting known kinases with *CLK1* highlighted. **(E)** Stacked barplot of *CLK1* exon 4 inclusion and skipping isoform fraction in HGGs. Dotted line represents the mean PSI of 0.7657. **(F)** Sashimi plot of two representative tumor samples with either high (BS_HRJ9145M) or low (BS_XM1AHBDJ) *CLK1* exon 4 inclusion. **(G)** PhosphositePlus[46] CLK1 protein visual highlighting the two phosphorylation binding sites in exon 4. **(H)** Pearson's correlation scatter plot of *CLK1* exon 4 PSI and RNA expression in HGG tumors ($R = 0.29$, $p = 4.1e^{-5}$). **(I)** Pearson's correlation scatter plot of *CLK1* exon 4 PSI and *SRPK1* RNA expression in HGG tumors ($R = 0.69$, $p = 2.2e^{-16}$). **(J)** Boxplot of CLK1 Exon 4 PSI levels across all primary pediatric brain tumors. All boxplots represent the 25th and 75th percentile and the bar represents the median.

Figure 4. CLK1 aberrant splicing contributes to oncogenesis in brain tumor cell lines.

(A) Boxplot of DepMap dependency scores stratified by high or low *CLK1* exon 4 containing transcript expression in brain tumor cell lines. Wilcoxon p-value shown. **(B)** Ranked dotplot of DepMap dependency scores in brain tumor cell lines with pediatric line KNS-42 highlighted in red. **(C)** Proliferation of KNS-42 cells treated with increasing concentrations of pan-DYRK/CLK1 inhibitor Cirtuvivint over six days. **(D)** Day 6 cell viability of KNS-42 cells treated with increasing concentrations of Cirtuvivint. Stars denote Bonferroni-adjusted p-values following pairwise Student's *t-tests*. **(E)** Stacked bar plots of the percent inclusion and skipping of *CLK1* exon 4 transcripts in patient-derived cell lines (7316-1763 and 7316-1769 from the CBTN) and KNS-42 (commercial) derived using either long (ONT) or short RNA-seq strategies. **(F)** Barplot showing the RNA expression fold-change in cells treated with control morpholino or morpholino targeting the *CLK1* exon 3-4 junction or exon 3-5 junction **(G)** Western blot of CLK1 with increasing morpholino treatment of 1, 5, and 10 μM. **(H)** Cell viability of cells treated with *CLK1* exon 4 morpholino or non-targeting morpholino. Stars denote within-time paired Student's *t-tests*. **(I)** Volcano plot illustrating genes differentially-expressed in KNS-42 cells treated with *CLK1* exon 4

targeting morpholino compared to cells treated with non-targeting morpholino. **(J)** Boxplot of |ΔPSI| of significantly differential splicing events comparing KNS-42 cells treated with *CLK1* exon 4 targeting morpholino vs. non-targeting morpholino (ΔPSI ≥ |.10|, p-value < 0.05, FDR < 0.05). Plot shows Uniprot-defined functional sites which are gained/lost categorized by splicing case (A3SS, A5SS, RI, and SE). **(K)** Barplots displaying number of differentially expressed (DE) genes or **(L)** differentially spliced (DS) genes affecting functional sites categorized by gene family. **(M)** Over-representation analysis using ClusterProfiler of DS cancer genes that result in gain/loss of functional sites. **(N)** Venn diagram depicting overlap of DS and DE genes from K and L **(O)** Venn diagram depicting overlap of DS and DE genes from K and L and significant (Wald FDR < 0.05, z-score < -1.5) essential genes identified in matched CBTN HGG cell lines through CRISPR dependency experiments from the Childhood Cancer Model Atlas (CCMA v3). **(P)** Ranked dotplot of significant CRISPR gene dependency mean z-scores for pediatric HGG cell lines with *CLK1* expression and splicing-based target genes highlighted in red and blue respectively.

Supplemental Figure Legends

Figure S1. Pearson correlation plots of genome-wide poly-A and stranded PSI values for SE events in PT_RYMG3M91 (**A**) and PT_W5GP3F6B (**B**) showing high concordance across RNA library types. (**C**) UpsetR plot showing recurrent differential splicing events (N ≥ 2) that prefer exon inclusion. Cumulative distribution plots of splicing burden index (SBI) by histology for RI (**D**), A5SS (**E**), and A3SS (**F**) events. Correlation plots for SBI vs TMB across the entire cohort including (**G**) and excluding (**H**) hypermutant and ultra-hyper mutant tumors, or within histology (**I**) excluding hypermutant and ultra-hypermutant tumors. Pearson's R and p-values are shown.

Figure S2. (**A**) Stacked barplot showing tumor histology membership in each cluster stratified by molecular subtype for EPN, HGG, LGG and MB histologies. (**B**) Oncoprint displaying mutation frequencies of splicing factors and spliceosome component genes in HGGs sorted by splicing burden index. Additional annotations include gender, molecular subtype, CNS region, and tumor mutation status.

Figure S3. (A) Over-representation analysis of mis-spliced kinase genes that result in gain/loss of functional sites stratified by **(A)** exon skipping and **(B)** exon skipping. **(C)** Boxplot of *CLK1* exon 4-containing transcript expression in HGGs compared to GTEx normal brain tissues. **(D)** Oncoprint displaying mutation frequencies of key brain cancer genes in HGGs with annotations for *CLK1* exon 4 PSI, *CLK1-201* expression, total *NF1*, *NF1-215* PSI, NF1 pS864 and pS2796 phosphoproteomic abundance, and total NF1 protein abundance z-scores. Additional annotations include gender, molecular subtype, CNS region, and tumor mutation status. All boxplots represent the 25th and 75th percentile and the bar represents the median.

Figure S4. (A) Boxplot of dependency scores stratified by high vs low *CLK1* exon 4 containing transcript expression across all available DepMap brain tumor cell lines. Within histology Wilcoxon p-values are shown. All boxplots represent the 25th and 75th percentile and the bar represents the median. **(B)** Cell viability assay after three days of treatment of KNS-42 cells with increasing concentrations of pan-DYRK/CLK1 inhibitor Cirtuvivint. **(C)** Heatmap displaying single-sample HALLMARK GSVA scores for DS genes affecting functional sites in cells treated with *CLK1* exon 4 morpholino or non-targeting morpholino. **(D)** Barplots illustrating the mean GSVA scores in panel **D** (n = 3 replicates per treatment). **(E)** Heatmap presenting single-sample DNA repair pathway GSVA scores for DS genes affecting functional sites in cells treated with *CLK1* exon 4 morpholino or non-targeting morpholino. **(F)** Barplots displaying mean DNA repair pathway GSVA scores from panel **E** (n = 3 replicates per treatment).

Figure S5. (A) Venn diagram showing the overlap of the total number of DS and DE genes. **(B)** Over-representation analysis of DE genes or **(C)** DS cancer genes that result in gain/loss of functional sites. **(D)** Ranked dotplot of significant CCMA v3 CRISPR gene dependency z-scores in individual CBTN cell lines with *CLK1* expression (red) and splicing-based (blue) target genes highlighted for HGG or **(E)** DMG patient-derived cell lines.

Supplementary Table Legends

Table S1: Sample metadata. (1) Readme and feature definitions. (2) Sample information with clinical metadata and demographics (3) CNS region definition from OpenPedCan[16].

Table S2: Histology-specific splicing events. (1) Exon inclusion related differential splicing events. (2) Exon skipping related differential splicing events. SpliceID includes gene name, mis-spliced exon start and end coordinates, upstream and downstream exon start and end coordinates. (3) Optimal clustering metrics and scoring method information. (4) Cluster membership for each sample.

Table S3: (1) Splicing factor and spliceosome component gene list. (2) DeSeq 2 results comparing high vs low SBI HGG tumors for splicing factors.

Table S4: Differential splicing events impacting functional sites in HGGs. (1) Exon skipping related differential splicing variants. (2) Exon inclusion related differential splicing variants. (3) Functional splice variants subsetted for known kinases.

Table S5: *CLK1* morpholino analyses. (1) Differential gene expression results from DeSeq2 and (2) rMATs results comparing treated with *CLK1* exon 4 morpholino and non-targeting

morpholino. (3) Differential splicing events associated with SE that correspond to known Uniprot functional sites. (4) Differential splicing events associated with A5SS that correspond to known Uniprot functional sites. (5) Differential splicing events associated with A3SS that correspond to known Uniprot functional sites. (6) Differential splicing events associated with RI that correspond to known Uniprot functional sites. (7) Differential splicing of known cancer genes that correspond to known Uniprot functional sites. (8) CLK1 exon 3-4, exon 3-5 and exon 3-5 junction forward and reverse primer sets. (9) Differentially expressed or differentially spliced CLK1 target genes overlapping essential oncogenes defined by CCMA v3.

Bibliography

1. Thorbinson, C. & Kilday, J.-P. Childhood Malignant Brain Tumors: Balancing the Bench and Bedside. *Cancers* **13**, (2021).
2. Bavle, A. & Chintagumpala, M. Pediatric high-grade glioma: a review of biology, prognosis, and treatment. *J. Radiat. Oncol.* **7**, 7–15 (2018).
3. Dalmage, M. *et al.* Survival and neurological outcomes after stereotactic biopsy of diffuse intrinsic pontine glioma: a systematic review. *J. Neurosurg. Pediatr.* **32**, 665–672 (2023).
4. Hatoum, R. *et al.* Extent of Tumor Resection and Survival in Pediatric Patients With High-Grade Gliomas: A Systematic Review and Meta-analysis. *JAMA Netw Open* **5**, e2226551 (2022).
5. Wu, G. *et al.* The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. *Nat. Genet.* **46**, 444–450 (2014).
6. Mackay, A. *et al.* Integrated Molecular Meta-Analysis of 1,000 Pediatric High-Grade and Diffuse Intrinsic Pontine Glioma. *Cancer Cell* **32**, 520–537.e5 (2017).
7. Siddaway, R. *et al.* Splicing is an alternate oncogenic pathway activation mechanism in glioma. *Nat. Commun.* **13**, 588 (2022).
8. Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437–451 (2017).
9. Ule, J. & Blencowe, B. J. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol. Cell* **76**, 329–345 (2019).
10. Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nat. Rev. Genet.* **17**, 19–32 (2016).
11. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. *Science* **338**, 1593–1599 (2012).
12. Chen, S. *et al.* CLK1/SRSF5 pathway induces aberrant exon skipping of METTL14 and

Cyclin L2 and promotes growth and metastasis of pancreatic cancer. *J. Hematol. Oncol.* **14**, 60 (2021).

13. Fuentes-Fayos, A. C. *et al.* Splicing machinery dysregulation drives glioblastoma development/aggressiveness: oncogenic role of SRSF3. *Brain* **143**, 3273–3293 (2020).
14. Shen, S. *et al.* rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E5593–601 (2014).
15. Mehmood, A. *et al.* Systematic evaluation of differential splicing tools for RNA-seq studies. *Brief. Bioinform.* **21**, 2052–2065 (2020).
16. Geng, Z. *et al.* The Open Pediatric Cancer Project. *bioRxiv* (2024) doi:10.1101/2024.07.09.599086.
17. Schönholzer, M. T. *et al.* Real-time sensing of MAPK signaling in medulloblastoma cells reveals cellular evasion mechanism counteracting dasatinib blockade of ERK activation during invasion. *Neoplasia* **22**, 470–483 (2020).
18. Karni, R. *et al.* The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat. Struct. Mol. Biol.* **14**, 185–193 (2007).
19. Kaida, D., Schneider-Poetsch, T. & Yoshida, M. Splicing in oncogenesis and tumor suppression. *Cancer Sci.* **103**, 1611–1616 (2012).
20. Golan-Gerstl, R. *et al.* Splicing factor hnRNP A2/B1 regulates tumor suppressor gene splicing and is an oncogenic driver in glioblastoma. *Cancer Res.* **71**, 4464–4472 (2011).
21. Sebestyén, E. *et al.* Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. Preprint at https://doi.org/10.1101/023010.
22. Petralia, F. *et al.* Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer. *Cell* **183**, 1962–1985.e31 (2020).
23. Koussounadis, A., Langdon, S. P., Um, I. H., Harrison, D. J. & Smith, V. A. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft

model system. *Sci. Rep.* **5**, 10775 (2015).

24. Wu, C. H. *et al.* The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.* **34**, D187–91 (2006).
25. Ninomiya, K., Kataoka, N. & Hagiwara, M. Stress-responsive maturation of Clk1/4 pre-mRNAs promotes phosphorylation of SR splicing factor. *J. Cell Biol.* **195**, 27–40 (2011).
26. Dominguez, D. *et al.* An extensive program of periodic alternative splicing linked to cell cycle progression. *Elife* **5**, (2016).
27. Aubol, B. E. *et al.* N-terminus of the protein kinase CLK1 induces SR protein hyperphosphorylation. *Biochem. J* **462**, 143–152 (2014).
28. Aubol, B. E. *et al.* Release of SR Proteins from CLK1 by SRPK1: A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing. *Mol. Cell* **63**, 218–228 (2016).
29. Bossard, C. *et al.* The Pan-Clk/Dyrk Inhibitor Cirtuvivint (SM08502) Exposes Mechanistic Underpinnings of Alternative Splicing As a Therapeutic Vulnerability in Heme Malignancies. *Blood* **138**, 2950 (2021).
30. Hillman, R. T., Green, R. E. & Brenner, S. E. An unappreciated role for RNA surveillance. *Genome Biol.* **5**, R8 (2004).
31. Uzor, S. *et al.* Autoregulation of the human splice factor kinase CLK1 through exon skipping and intron retention. *Gene* **670**, 46–54 (2018).
32. Duncan, P. I., Stojdl, D. F., Marius, R. M. & Bell, J. C. In vivo regulation of alternative pre-mRNA splicing by the Clk1 protein kinase. *Mol. Cell. Biol.* **17**, 5996–6001 (1997).
33. Sun, C. X. *et al.* Generation and multi-dimensional profiling of a childhood cancer cell line atlas defines new therapeutic opportunities. *Cancer Cell* **41**, 660–677.e7 (2023).
34. Dehm, S. M. & Bonham, K. SRC gene expression in human cancer: the role of transcriptional activation. *Biochem. Cell Biol.* **82**, 263–274 (2004).
35. Georgescu, M.-M. *et al.* Global activation of oncogenic pathways underlies therapy resistance in diffuse midline glioma. *Acta Neuropathol Commun* **8**, 111 (2020).

36. Zhang, J. & Manley, J. L. Misregulation of pre-mRNA alternative splicing in cancer. *Cancer Discov.* **3**, 1228–1237 (2013).

37. Nomura, S. *et al.* FGF10/FGFR2 signal induces cell migration and invasion in pancreatic cancer. *Br. J. Cancer* **99**, 305–313 (2008).

38. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).

39. Dubuc, A. M. *et al.* Subgroup-specific alternative splicing in medulloblastoma. *Acta Neuropathol.* **123**, 485–499 (2012).

40. Tolcher, A. *et al.* Abstract CT112: Initial results from a Phase 1 trial of a first-in-class pan-CDC-like kinase inhibitor (SM08502) with proof of mechanism in subjects with advanced solid tumors. *Cancer Res.* **81**, CT112–CT112 (2021).

41. ClinicalTrials.Gov. https://www.clinicaltrials.gov/study/NCT05084859.

42. A. Scott, J. Call, S. Chandana, E. Borazanci, G. Falchook, R. Bordoni, S. Richey, A. Starodub, V. Chung, N. Lakhani, E. Lam, K. Schaffer, J. Wang, G. Shapiro, J. Sachdev, D. Beaupre and A. Tolcher. Preliminary Evidence of Clinical Activity from Phase 1 and 1b Trials of the CLK/DYRK Inhibitor Cirtuvivint (CIRT) in Subjects with Advanced Solid Tumors. (2022).

43. Shapiro, J. A. *et al.* OpenPBTA: The Open Pediatric Brain Tumor Atlas. *Cell Genomics* **3**, (2023).

44. Garrido-Martín, D., Palumbo, E., Guigó, R. & Breschi, A. ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization. *PLoS Comput. Biol.* **14**, e1006360 (2018).

45. ggpubr: ‘ggplot2’ Based Publication Ready Plots. https://rdrr.io/cran/ggpubr/ (2023).

46. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–20 (2015).

47. Maechler, M. Package ‘diptest’. *R Package Version 0. 75--5. R: a language and*

environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing (2013).

48. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
49. Hennig, C. & Imports, M. Package 'fpc'. *Flexible procedures for clustering* **1176**, (2015).
50. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
51. Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv* (2016) doi:10.1101/060012.
52. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
53. Kolde, R. & Kolde, M. R. Package 'pheatmap'. *R package* **1**, 790 (2015).
54. Love, M., Anders, S. & Huber, W. Differential analysis of count data--the DESeq2 package. *Genome Biol.* **15**, 10–1186 (2014).
55. Wickham, H. Ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* **3**, 180–185 (2011).
56. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
57. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).
58. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).
59. Dolgalev, I. msigdbr: MSigDB gene sets for multiple organisms in a tidy data format. *R package version* **7**, (2020).
60. 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).

Figure 1

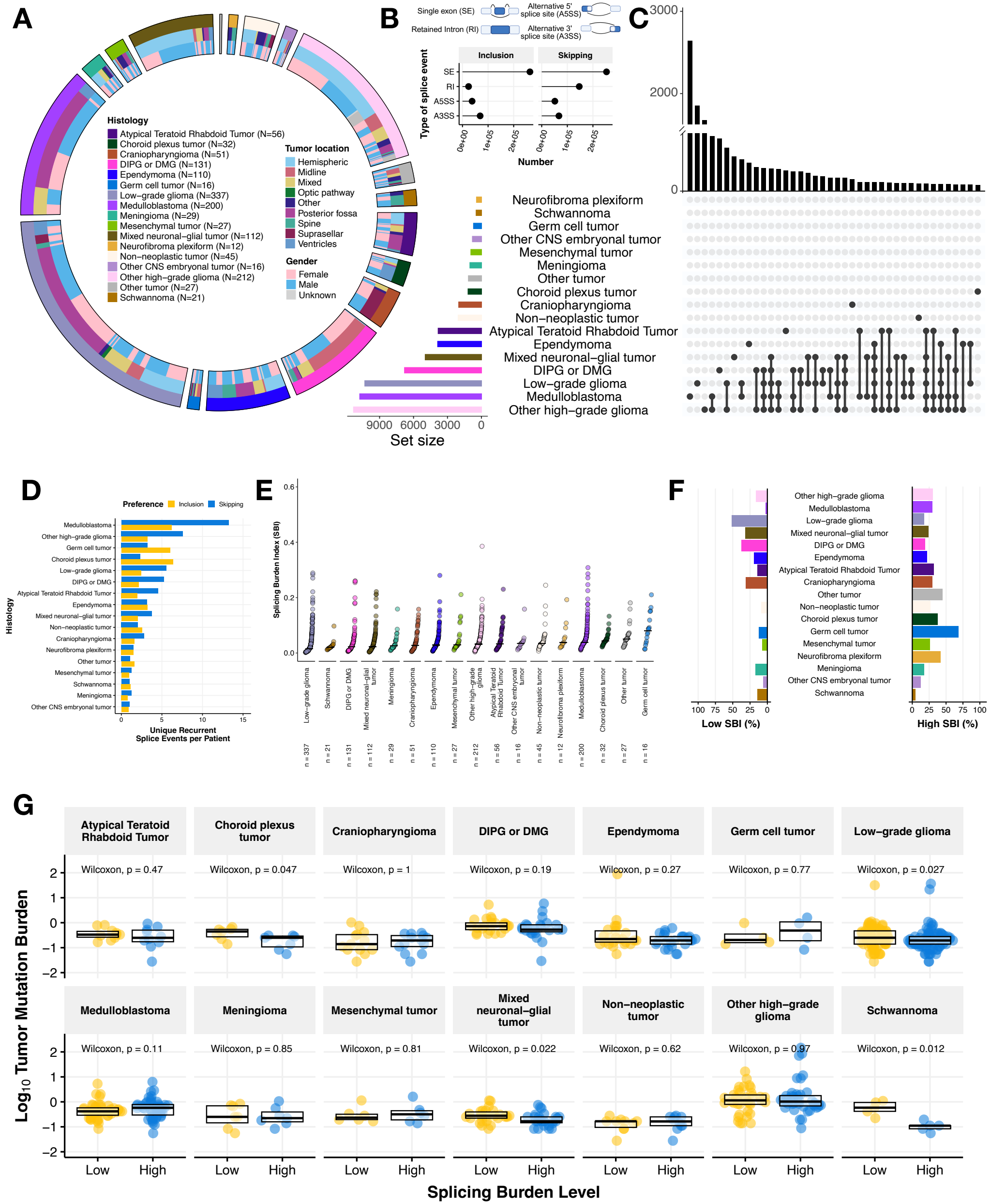

A
Histology
Atypical Teratoid Rhabdoid Tumor (N=56)
Choroid plexus tumor (N=32)
Craniopharyngioma (N=51)
DIPG or DMG (N=131)
Ependymoma (N=110)
Germ cell tumor (N=16)
Low−grade glioma (N=337)
Medulloblastoma (N=200)
Meningioma (N=29)
Mesenchymal tumor (N=27)
Mixed neuronal−glial tumor (N=112)
Neurofibroma plexiform (N=12)
Non−neoplastic tumor (N=45)
Other CNS embryonal tumor (N=16)
Other high−grade glioma (N=212)
Other tumor (N=27)
Schwannoma (N=21)
Tumor location
Hemispheric
Midline
Mixed
Optic pathway
Other
Posterior fossa
Spine
Suprasellar
Ventricles
Gender
Female
Male
Unknown
B
Single exon (SE)
Retained Intron (RI)
Alternative 5' splice site (A5SS)
Alternative 3' splice site (A3SS)
Inclusion
Skipping
Type of splice event
SE
RI
A5SS
A3SS
0e+00
1e+05
2e+05
Number
C
3000
2000
0
Neurofibroma plexiform
Schwannoma
Germ cell tumor
Other CNS embryonal tumor
Mesenchymal tumor
Meningioma
Other tumor
Choroid plexus tumor
Craniopharyngioma
Non−neoplastic tumor
Atypical Teratoid Rhabdoid Tumor
Ependymoma
Mixed neuronal−glial tumor
DIPG or DMG
Low−grade glioma
Medulloblastoma
Other high−grade glioma
9000 6000 3000 0
Set size
D
Preference
Inclusion
Skipping
Histology
Medulloblastoma
Other high−grade glioma
Germ cell tumor
Choroid plexus tumor
Low−grade glioma
DIPG or DMG
Atypical Teratoid Rhabdoid Tumor
Ependymoma
Mixed neuronal−glial tumor
Non−neoplastic tumor
Craniopharyngioma
Neurofibroma plexiform
Other tumor
Mesenchymal tumor
Schwannoma
Meningioma
Other CNS embryonal tumor
0
5
10
15
Unique Recurrent Splice Events per Patient
E
Splicing Burden Index (SBI)
0.6
0.4
0.2
0.0
Low−grade glioma
Schwannoma
DIPG or DMG
Mixed neuronal−glial tumor
Meningioma
Craniopharyngioma
Ependymoma
Mesenchymal tumor
Other high−grade glioma
Atypical Teratoid Rhabdoid Tumor
Other CNS embryonal tumor
Non−neoplastic tumor
Neurofibroma plexiform
Medulloblastoma
Choroid plexus tumor
Other tumor
Germ cell tumor
n = 337
n = 21
n = 131
n = 112
n = 29
n = 51
n = 110
n = 27
n = 212
n = 56
n = 16
n = 45
n = 12
n = 200
n = 32
n = 27
n = 16
F
Other high−grade glioma
Medulloblastoma
Low−grade glioma
Mixed neuronal−glial tumor
DIPG or DMG
Ependymoma
Atypical Teratoid Rhabdoid Tumor
Craniopharyngioma
Other tumor
Non−neoplastic tumor
Choroid plexus tumor
Germ cell tumor
Mesenchymal tumor
Neurofibroma plexiform
Meningioma
Other CNS embryonal tumor
Schwannoma
100 75 50 25 0
Low SBI (%)
0 25 50 75 100
High SBI (%)
G
Atypical Teratoid Rhabdoid Tumor
Wilcoxon, p = 0.47
Choroid plexus tumor
Wilcoxon, p = 0.047
Craniopharyngioma
Wilcoxon, p = 1
DIPG or DMG
Wilcoxon, p = 0.19
Ependymoma
Wilcoxon, p = 0.27
Germ cell tumor
Wilcoxon, p = 0.77
Low−grade glioma
Wilcoxon, p = 0.027
Medulloblastoma
Wilcoxon, p = 0.11
Meningioma
Wilcoxon, p = 0.85
Mesenchymal tumor
Wilcoxon, p = 0.81
Mixed neuronal−glial tumor
Wilcoxon, p = 0.022
Non−neoplastic tumor
Wilcoxon, p = 0.62
Other high−grade glioma
Wilcoxon, p = 0.97
Schwannoma
Wilcoxon, p = 0.012
Log10 Tumor Mutation Burden
2
1
0
−1
−2
Low
High
Splicing Burden Level

Figure 2

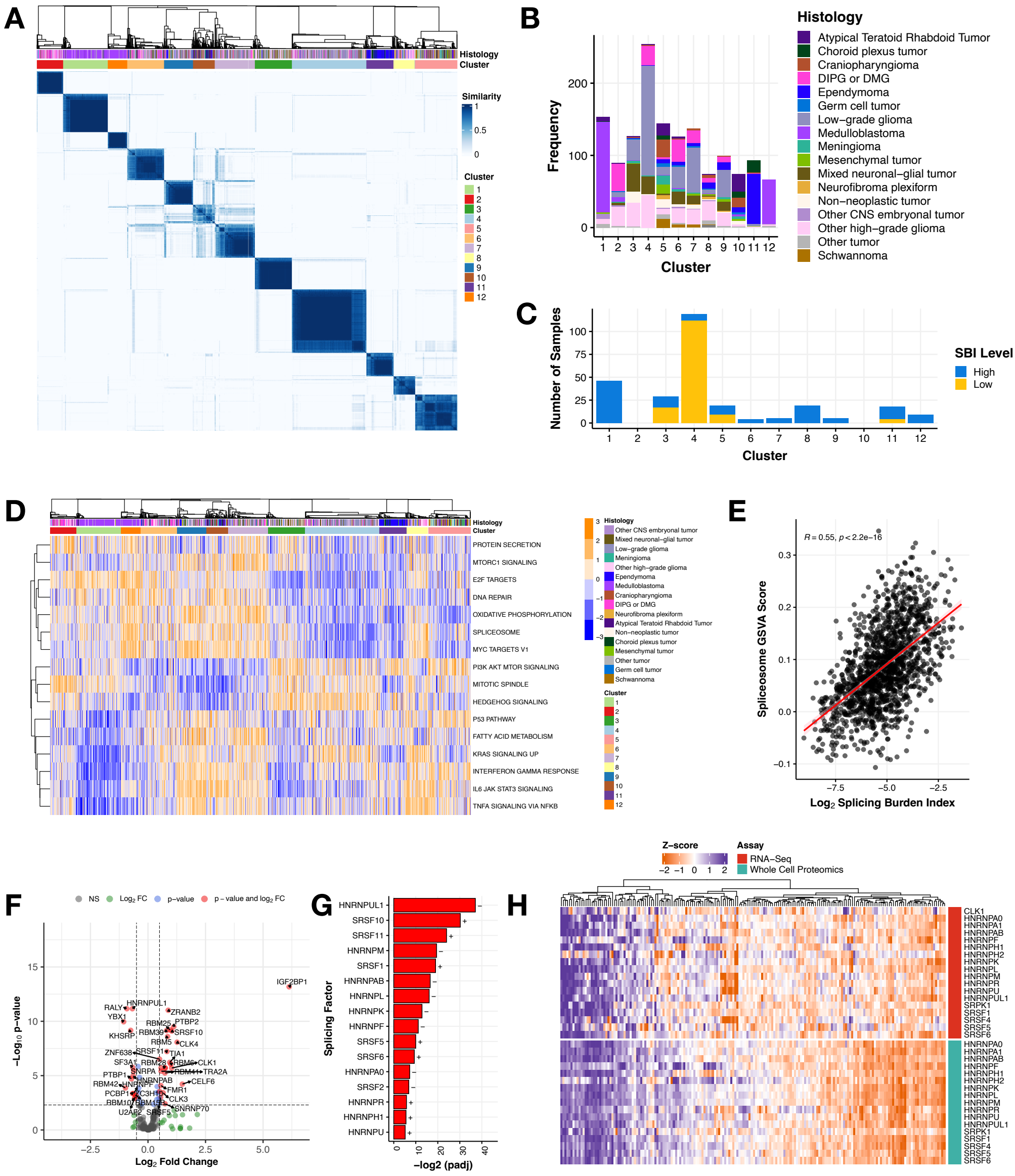

A
Histology
Cluster
Similarity
1
0.5
0
Cluster
1
2
3
4
5
6
7
8
9
10
11
12
B
Histology
Atypical Teratoid Rhabdoid Tumor
Choroid plexus tumor
Craniopharyngioma
DIPG or DMG
Ependymoma
Germ cell tumor
Low−grade glioma
Medulloblastoma
Meningioma
Mesenchymal tumor
Mixed neuronal−glial tumor
Neurofibroma plexiform
Non−neoplastic tumor
Other CNS embryonal tumor
Other high−grade glioma
Other tumor
Schwannoma
Frequency
Cluster
C
Number of Samples
Cluster
SBI Level
High
Low
D
Histology
Cluster
PROTEIN SECRETION
MTORC1 SIGNALING
E2F TARGETS
DNA REPAIR
OXIDATIVE PHOSPHORYLATION
SPLICEOSOME
MYC TARGETS V1
PI3K AKT MTOR SIGNALING
MITOTIC SPINDLE
HEDGEHOG SIGNALING
P53 PATHWAY
FATTY ACID METABOLISM
KRAS SIGNALING UP
INTERFERON GAMMA RESPONSE
IL6 JAK STAT3 SIGNALING
TNFA SIGNALING VIA NFKB
E
R = 0.55, p < 2.2e−16
Spliceosome GSVA Score
Log2 Splicing Burden Index
F
NS
Log2 FC
p−value
p − value and log2 FC
−Log10 p−value
Log2 Fold Change
IGF2BP1
HNRNPUL1
RALY
YBX1
ZRANB2
PTBP2
RBM25
KHSRP
RBM39
SRSF10
RBM5
CLK4
ZNF638
SRSF11
TIA1
SF3A1
RBM28
RBM6
CLK1
SNRPA
RBM41
TRA2A
PTBP1
HNRNPAB
CELF6
RBM42
HNRNPF
FMR1
PCBP1
ZC3H10
CLK3
RBM10
RBM15B
SNRNP70
U2AF2
SRSF5
G
Splicing Factor
HNRNPUL1
SRSF10
SRSF11
HNRNPM
SRSF1
HNRNPAB
HNRNPL
HNRNPK
HNRNPF
SRSF5
SRSF6
HNRNPA0
SRSF2
HNRNPR
HNRNPH1
HNRNPU
−log2 (padj)
H
Z−score
Assay
RNA−Seq
Whole Cell Proteomics
CLK1
HNRNPA0
HNRNPA1
HNRNPAB
HNRNPF
HNRNPH1
HNRNPH2
HNRNPK
HNRNPL
HNRNPM
HNRNPR
HNRNPU
HNRNPUL1
SRPK1
SRSF1
SRSF4
SRSF5
SRSF6

Figure 3

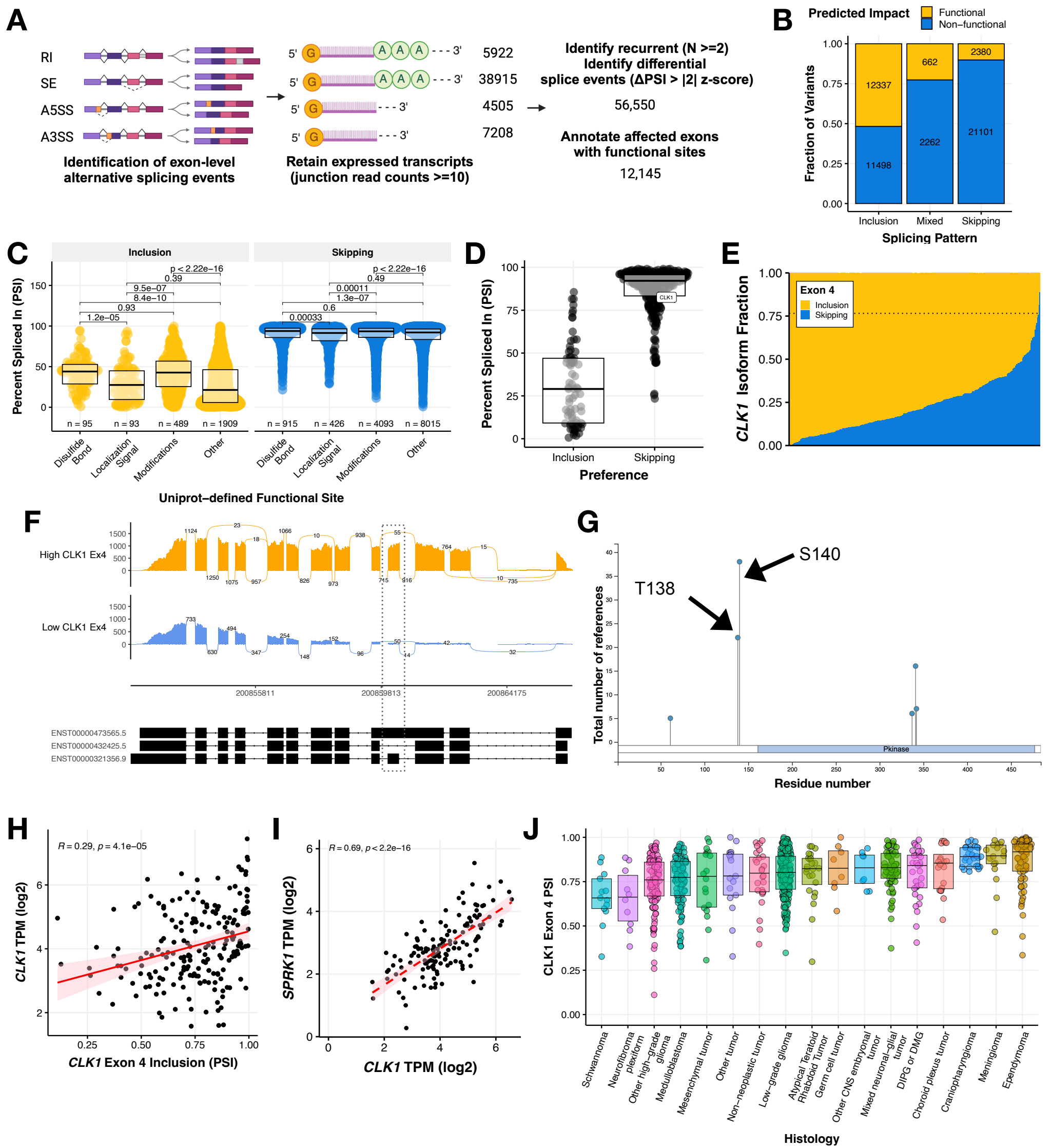

A
RI
SE
A5SS
A3SS
Identification of exon-level alternative splicing events
5'
3'
5922
38915
4505
7208
Retain expressed transcripts (junction read counts >=10)
Identify recurrent (N >=2) Identify differential splice events (ΔPSI > |2| z-score)
56,550
Annotate affected exons with functional sites
12,145
B
Predicted Impact
Functional
Non–functional
Fraction of Variants
12337
11498
662
2262
2380
21101
Inclusion
Mixed
Skipping
Splicing Pattern
C
Inclusion
Skipping
Percent Spliced In (PSI)
p < 2.22e–16
0.39
9.5e–07
8.4e–10
0.93
1.2e–05
0.49
0.00011
1.3e–07
0.6
0.00033
n = 95
n = 93
n = 489
n = 1909
n = 915
n = 426
n = 4093
n = 8015
Disulfide Bond
Localization Signal
Modifications
Other
Uniprot–defined Functional Site
D
Percent Spliced In (PSI)
CLK1
Inclusion
Skipping
Preference
E
CLK1 Isoform Fraction
Exon 4
Inclusion
Skipping
F
High CLK1 Ex4
Low CLK1 Ex4
200855811
200859813
200864175
ENST00000473565.5
ENST00000432425.5
ENST00000321356.9
G
Total number of references
T138
S140
Pkinase
Residue number
H
R = 0.29, p = 4.1e–05
CLK1 TPM (log2)
CLK1 Exon 4 Inclusion (PSI)
I
R = 0.69, p < 2.2e–16
SPRK1 TPM (log2)
CLK1 TPM (log2)
J
CLK1 Exon 4 PSI
Schwannoma
Neurofibroma plexiform
Other high–grade glioma
Medulloblastoma
Mesenchymal tumor
Other tumor
Non–neoplastic tumor
Low–grade glioma
Atypical Teratoid Rhabdoid Tumor
Germ cell tumor
Other CNS embryonal tumor
Mixed neuronal–glial tumor
DIPG or DMG
Choroid plexus tumor
Craniopharyngioma
Meningioma
Ependymoma
Histology

Figure 4

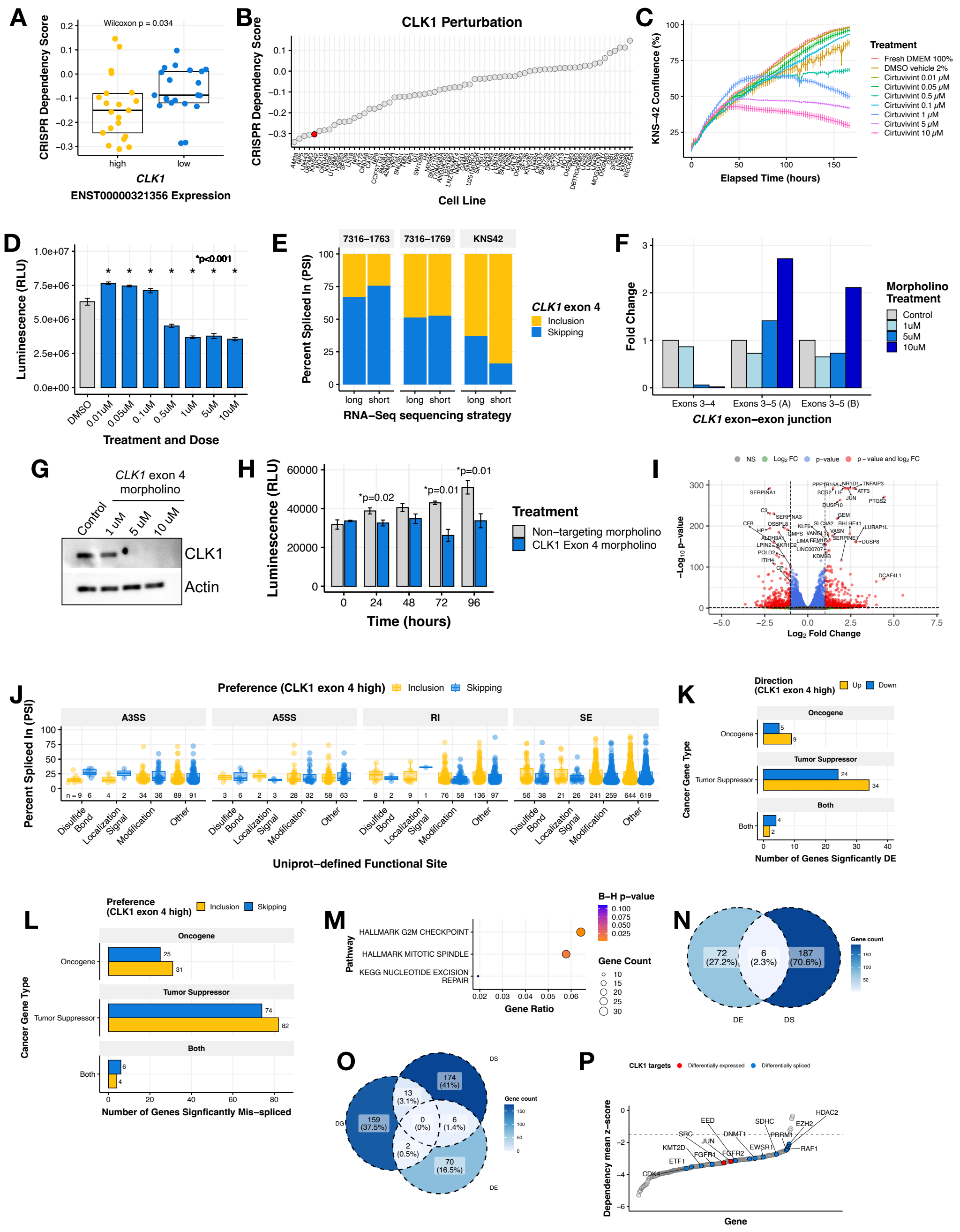

A
Wilcoxon p = 0.034
CRISPR Dependency Score
high
low
CLK1
ENST00000321356 Expression
B
CLK1 Perturbation
CRISPR Dependency Score
Cell Line
C
KNS−42 Confluence (%)
Elapsed Time (hours)
Treatment
Fresh DMEM 100%
DMSO vehicle 2%
Cirtuvivint 0.01 μM
Cirtuvivint 0.05 μM
Cirtuvivint 0.5 μM
Cirtuvivint 0.1 μM
Cirtuvivint 1 μM
Cirtuvivint 5 μM
Cirtuvivint 10 μM
D
Luminescence (RLU)
*p<0.001
DMSO
0.01uM
0.05uM
0.1uM
0.5uM
1uM
5uM
10uM
Treatment and Dose
E
7316−1763
7316−1769
KNS42
Percent Spliced In (PSI)
long
short
RNA−Seq sequencing strategy
CLK1 exon 4
Inclusion
Skipping
F
Fold Change
Exons 3−4
Exons 3−5 (A)
Exons 3−5 (B)
CLK1 exon−exon junction
Morpholino Treatment
Control
1uM
5uM
10uM
G
CLK1 exon 4 morpholino
Control
1 uM
5 uM
10 uM
CLK1
Actin
H
Luminescence (RLU)
*p=0.02
*p=0.01
*p=0.01
Time (hours)
Treatment
Non−targeting morpholino
CLK1 Exon 4 morpholino
I
NS
Log2 FC
p−value
p − value and log2 FC
−Log10 p−value
Log2 Fold Change
SERPINA1
C3
SERPINA3
CFB
OSBPL8
HP
ALDH3A1
LPIN2
AKR1D2
POLD2
ITIH4
CP
KLF6
GMPS
VANGL1
LIMA1
FEM1C
LINC00707
KDM6B
SLC3A2
VASN
SERPINE1
BHLHE41
LURAP1L
DUSP8
DCAF4L1
PPP1R15A
NR1D1
TNFAIP3
SCG2
LIF
ATF3
JUN
PTGS2
DUSP10
GEM
J
Preference (CLK1 exon 4 high)
Inclusion
Skipping
A3SS
A5SS
RI
SE
Percent Spliced In (PSI)
n = 9 6 4 2 34 36 89 91
3 6 2 3 28 32 58 63
8 2 9 1 76 58 136 97
56 38 21 26 241 259 644 619
Disulfide Bond
Localization
Signal
Modification
Other
Uniprot−defined Functional Site
K
Direction (CLK1 exon 4 high)
Up
Down
Oncogene
Tumor Suppressor
Both
Cancer Gene Type
Number of Genes Signficantly DE
L
Preference (CLK1 exon 4 high)
Inclusion
Skipping
Oncogene
Tumor Suppressor
Both
Cancer Gene Type
Number of Genes Signficantly Mis−spliced
M
B−H p−value
HALLMARK G2M CHECKPOINT
HALLMARK MITOTIC SPINDLE
KEGG NUCLEOTIDE EXCISION REPAIR
Pathway
Gene Ratio
Gene Count
N
72 (27.2%)
6 (2.3%)
187 (70.6%)
Gene count
DE
DS
O
DS
DG
DE
174 (41%)
13 (3.1%)
159 (37.5%)
0 (0%)
6 (1.4%)
2 (0.5%)
70 (16.5%)
Gene count
P
CLK1 targets
Differentially expressed
Differentially spliced
Dependency mean z−score
Gene
HDAC2
EED
SDHC
EZH2
SRC
DNMT1
PBRM1
JUN
KMT2D
EWSR1
RAF1
FGFR1
FGFR2
ETF1
CDK4

Figure S1

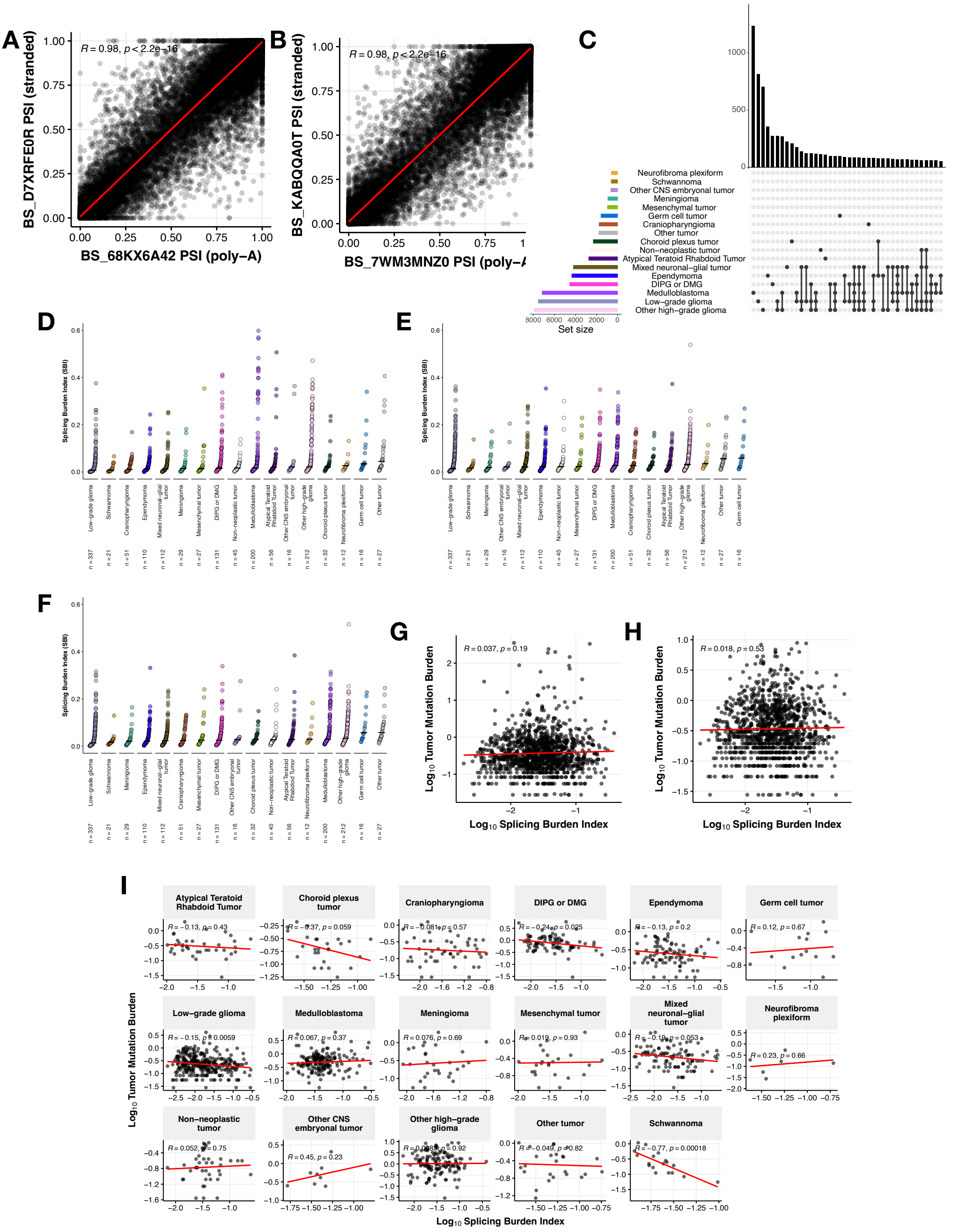
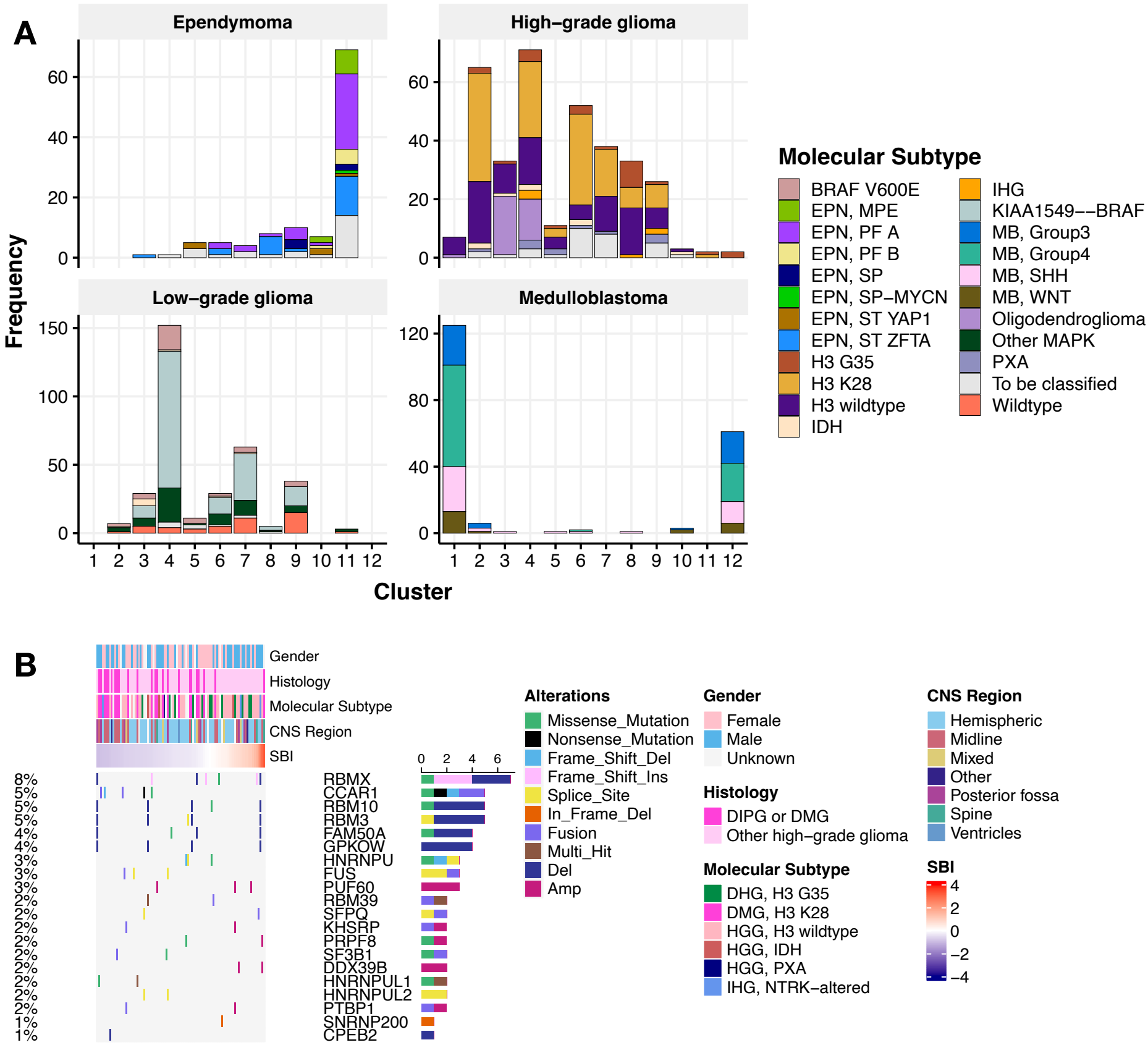

A
BS_D7XRFE0R PSI (stranded)
BS_68KX6A42 PSI (poly−A)
R = 0.98, p < 2.2e−16
B
BS_KABQQA0T PSI (stranded)
BS_7WM3MNZ0 PSI (poly−A)
R = 0.98, p < 2.2e−16
C
Neurofibroma plexiform
Schwannoma
Other CNS embryonal tumor
Meningioma
Mesenchymal tumor
Germ cell tumor
Craniopharyngioma
Other tumor
Choroid plexus tumor
Non−neoplastic tumor
Atypical Teratoid Rhabdoid Tumor
Mixed neuronal−glial tumor
Ependymoma
DIPG or DMG
Medulloblastoma
Low−grade glioma
Other high−grade glioma
Set size
D
E
F
Splicing Burden Index (SBI)
G
Log10 Tumor Mutation Burden
Log10 Splicing Burden Index
R = 0.037, p = 0.19
H
R = 0.018, p = 0.53
I
Atypical Teratoid Rhabdoid Tumor
R = −0.13, p = 0.43
Choroid plexus tumor
R = −0.37, p = 0.059
Craniopharyngioma
R = −0.081, p = 0.57
DIPG or DMG
R = −0.24, p = 0.025
Ependymoma
R = −0.13, p = 0.2
Germ cell tumor
R = 0.12, p = 0.67
Low−grade glioma
R = −0.15, p = 0.0059
Medulloblastoma
R = 0.067, p = 0.37
Meningioma
R = 0.076, p = 0.69
Mesenchymal tumor
R = 0.019, p = 0.93
Mixed neuronal−glial tumor
R = −0.19, p = 0.053
Neurofibroma plexiform
R = 0.23, p = 0.66
Non−neoplastic tumor
R = 0.052, p = 0.75
Other CNS embryonal tumor
R = 0.45, p = 0.23
Other high−grade glioma
R = 0.0083, p = 0.92
Other tumor
R = −0.049, p = 0.82
Schwannoma
R = −0.77, p = 0.00018
Log10 Tumor Mutation Burden
Log10 Splicing Burden Index

Figure S2

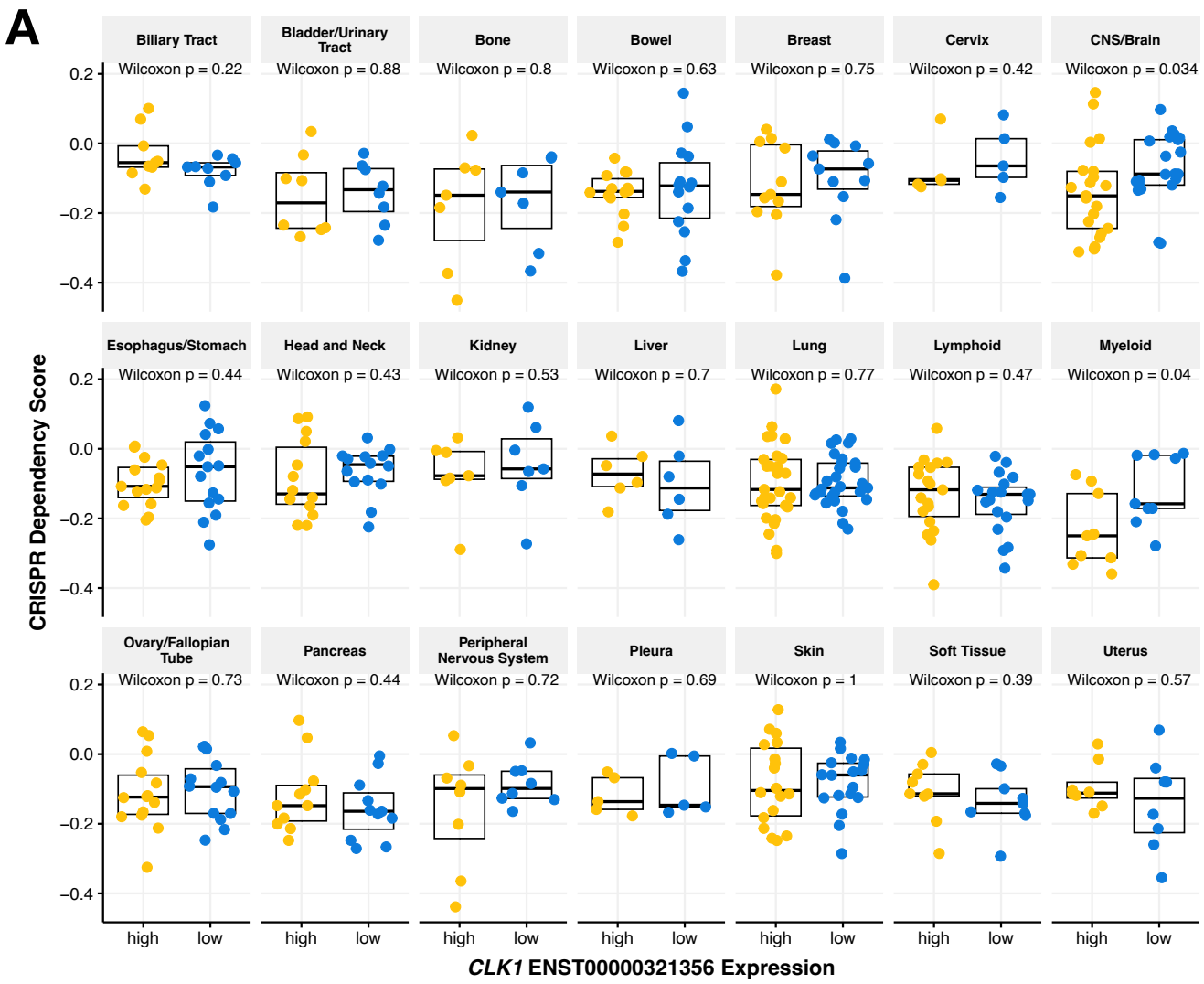

A
Ependymoma
High−grade glioma
Low−grade glioma
Medulloblastoma
Frequency
Cluster
Molecular Subtype
BRAF V600E
EPN, MPE
EPN, PF A
EPN, PF B
EPN, SP
EPN, SP−MYCN
EPN, ST YAP1
EPN, ST ZFTA
H3 G35
H3 K28
H3 wildtype
IDH
IHG
KIAA1549−−BRAF
MB, Group3
MB, Group4
MB, SHH
MB, WNT
Oligodendroglioma
Other MAPK
PXA
To be classified
Wildtype

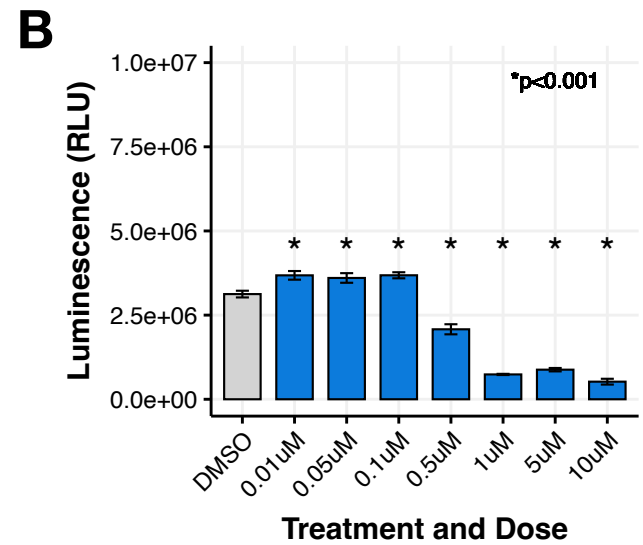
B
Gender
Histology
Molecular Subtype
CNS Region
SBI
RBMX
CCAR1
RBM10
RBM3
FAM50A
GPKOW
HNRNPU
FUS
PUF60
RBM39
SFPQ
KHSRP
PRPF8
SF3B1
DDX39B
HNRNPUL1
HNRNPUL2
PTBP1
SNRNP200
CPEB2
Alterations
Missense_Mutation
Nonsense_Mutation
Frame_Shift_Del
Frame_Shift_Ins
Splice_Site
In_Frame_Del
Fusion
Multi_Hit
Del
Amp
Gender
Female
Male
Unknown
Histology
DIPG or DMG
Other high−grade glioma
Molecular Subtype
DHG, H3 G35
DMG, H3 K28
HGG, H3 wildtype
HGG, IDH
HGG, PXA
IHG, NTRK−altered
CNS Region
Hemispheric
Midline
Mixed
Other
Posterior fossa
Spine
Ventricles
SBI

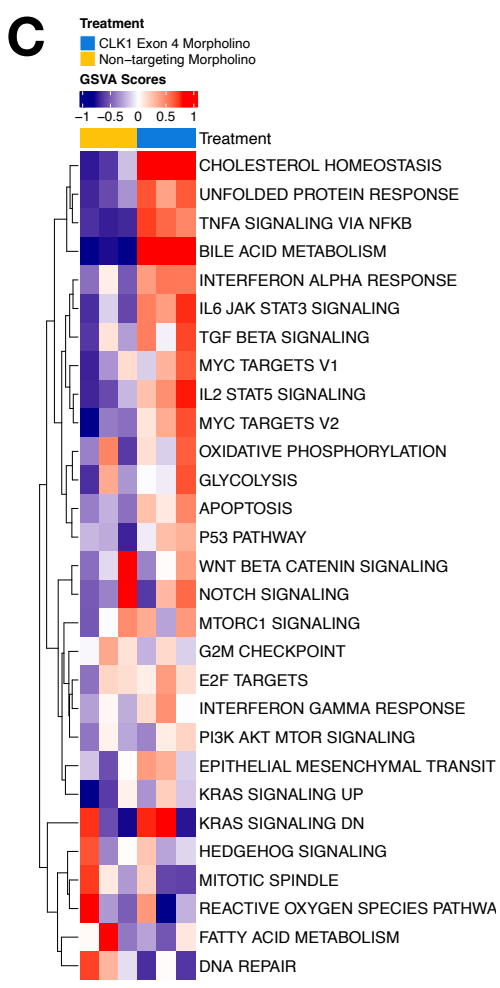
Figure S3

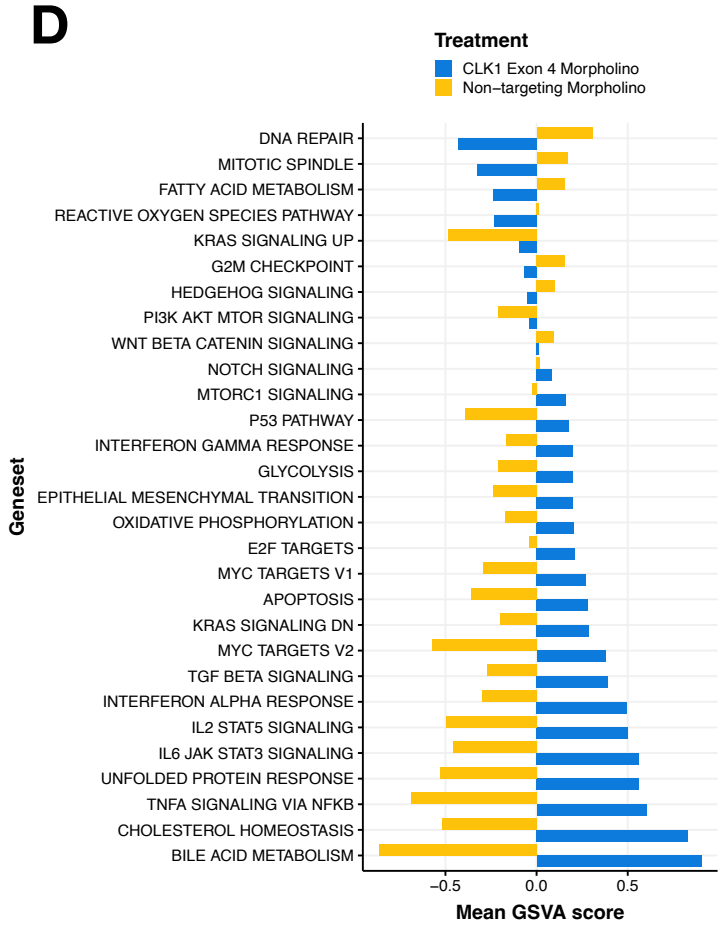
A
Exon Skipping
KEGG MAPK SIGNALING PATHWAY
BIOCARTA MAPK PATHWAY
HALLMARK PI3K AKT MTOR SIGNALING
KEGG INSULIN SIGNALING PATHWAY
KEGG CHEMOKINE SIGNALING PATHWAY
KEGG CALCIUM SIGNALING PATHWAY
KEGG ENDOCYTOSIS
KEGG ERBB SIGNALING PATHWAY
KEGG NEUROTROPHIN SIGNALING PATHWAY
KEGG AXON GUIDANCE
KEGG WNT SIGNALING PATHWAY
KEGG GNRH SIGNALING PATHWAY
KEGG HEDGEHOG SIGNALING PATHWAY
BIOCARTA MTOR PATHWAY
BIOCARTA EIF4 PATHWAY
Pathway
0.02
0.04
0.06
0.08
Gene Ratio
B–H p–value
0.005
0.004
0.003
0.002
0.001
Gene Count
5.0
7.5
10.0
B
Exon Inclusion
KEGG ERBB SIGNALING PATHWAY
KEGG MAPK SIGNALING PATHWAY
BIOCARTA MAPK PATHWAY
HALLMARK PI3K AKT MTOR SIGNALING
KEGG FOCAL ADHESION
KEGG GNRH SIGNALING PATHWAY
KEGG AXON GUIDANCE
KEGG REGULATION OF ACTIN CYTOSKELETON
BIOCARTA KERATINOCYTE PATHWAY
KEGG FC EPSILON RI SIGNALING PATHWAY
KEGG TOLL LIKE RECEPTOR SIGNALING PATHWAY
KEGG T CELL RECEPTOR SIGNALING PATHWAY
BIOCARTA AGPCR PATHWAY
BIOCARTA EIF4 PATHWAY
BIOCARTA HDAC PATHWAY
Pathway
0.050
0.075
0.100
0.125
Gene Ratio
Gene Count
2
3
4
5
6
B–H p–value
0.008
0.006
0.004
0.002
C
CLK1 ENST00000321356 TPM (log2)
150
100
50
0
Other high–grade glioma
Putamen (basal ganglia)
Anterior cingulate cortex (BA24)
Nucleus accumbens (basal ganglia)
Amygdala
Frontal Cortex (BA9)
Caudate (basal ganglia)
Cortex
Substantia nigra
Hypothalamus
Hippocampus
DIPG or DMG
Cerebellum
Spinal cord (cervical c–1)
Tissue
Tissue
DIPG or DMG
Other high–grade glioma
GTEx Brain
D
Gender
Predisposition
Histology
Molecular Subtype
CNS Region
Mutation Status
CLK1 Ex4 PSI
CLK1–201
Total CLK1 RNA
0 20 40 60 80
42% TP53
36% H3–3A
28% ATRX
16% CIC
16% PDGFRA
16% NF1
12% EGFR
12% KMT2C
9% PIK3CA
8% KIT
8% BCOR
8% SETD2
8% MET
8% BRAF
8% KDR
7% CDK4
7% PTEN
7% TSC2
6% PTPN11
6% FGFR1
6% ASXL1
6% ATM
6% CDKN2A
5% PPM1D
4% CDK6
Alterations
Missense_Mutation
Nonsense_Mutation
Frame_Shift_Del
Frame_Shift_Ins
Splice_Site
Translation_Start_Site
In_Frame_Del
Fusion
Multi_Hit
Del
Amp
Gender
Female
Male
Unknown
Predisposition
LFS
NF–1
Other
Histology
DIPG or DMG
Other high–grade glioma
Molecular Subtype
DHG, H3 G35
DMG, H3 K28
HGG, H3 wildtype
HGG, IDH
HGG, PXA
IHG, ALK–altered
IHG, NTRK–altered
IHG, ROS1–altered
To be classified
CNS Region
Hemispheric
Midline
Mixed
Other
Posterior fossa
Spine
Suprasellar
Ventricles
Mutation Status
Hypermutant
Normal
Ultra–hypermutant
Unknown
CLK1 Ex4 PSI
2
0
–2
–4
CLK1–201
4
2
0
–2
–4
Total CLK1 RNA
4
2
0
–2
–4

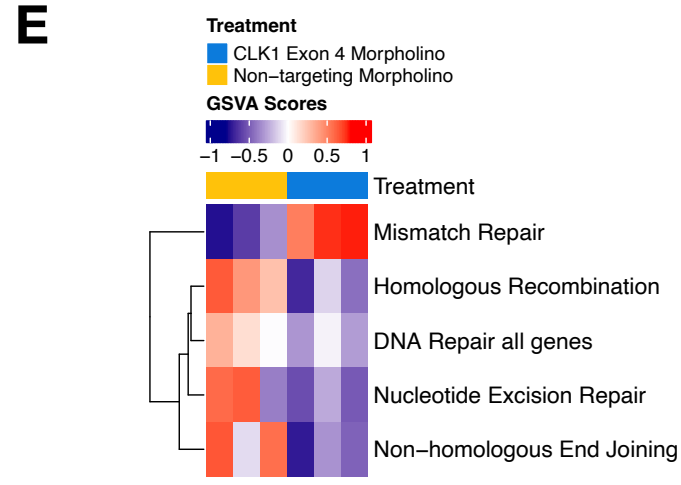
Figure S4


A
Biliary Tract
Wilcoxon p = 0.22
Bladder/Urinary Tract
Wilcoxon p = 0.88
Bone
Wilcoxon p = 0.8
Bowel
Wilcoxon p = 0.63
Breast
Wilcoxon p = 0.75
Cervix
Wilcoxon p = 0.42
CNS/Brain
Wilcoxon p = 0.034
Esophagus/Stomach
Wilcoxon p = 0.44
Head and Neck
Wilcoxon p = 0.43
Kidney
Wilcoxon p = 0.53
Liver
Wilcoxon p = 0.7
Lung
Wilcoxon p = 0.77
Lymphoid
Wilcoxon p = 0.47
Myeloid
Wilcoxon p = 0.04
Ovary/Fallopian Tube
Wilcoxon p = 0.73
Pancreas
Wilcoxon p = 0.44
Peripheral Nervous System
Wilcoxon p = 0.72
Pleura
Wilcoxon p = 0.69
Skin
Wilcoxon p = 1
Soft Tissue
Wilcoxon p = 0.39
Uterus
Wilcoxon p = 0.57
CRISPR Dependency Score
0.2
0.0
−0.2
−0.4
high
low
CLK1 ENST00000321356 Expression


B
Luminescence (RLU)
1.0e+07
7.5e+06
5.0e+06
2.5e+06
0.0e+00
*p<0.001
DMSO
0.01uM
0.05uM
0.1uM
0.5uM
1uM
5uM
10uM
Treatment and Dose


C
Treatment
CLK1 Exon 4 Morpholino
Non−targeting Morpholino
GSVA Scores
−1 −0.5 0 0.5 1
Treatment
CHOLESTEROL HOMEOSTASIS
UNFOLDED PROTEIN RESPONSE
TNFA SIGNALING VIA NFKB
BILE ACID METABOLISM
INTERFERON ALPHA RESPONSE
IL6 JAK STAT3 SIGNALING
TGF BETA SIGNALING
MYC TARGETS V1
IL2 STAT5 SIGNALING
MYC TARGETS V2
OXIDATIVE PHOSPHORYLATION
GLYCOLYSIS
APOPTOSIS
P53 PATHWAY
WNT BETA CATENIN SIGNALING
NOTCH SIGNALING
MTORC1 SIGNALING
G2M CHECKPOINT
E2F TARGETS
INTERFERON GAMMA RESPONSE
PI3K AKT MTOR SIGNALING
EPITHELIAL MESENCHYMAL TRANSIT
KRAS SIGNALING UP
KRAS SIGNALING DN
HEDGEHOG SIGNALING
MITOTIC SPINDLE
REACTIVE OXYGEN SPECIES PATHWA
FATTY ACID METABOLISM
DNA REPAIR


D
Treatment
CLK1 Exon 4 Morpholino
Non−targeting Morpholino
Geneset
DNA REPAIR
MITOTIC SPINDLE
FATTY ACID METABOLISM
REACTIVE OXYGEN SPECIES PATHWAY
KRAS SIGNALING UP
G2M CHECKPOINT
HEDGEHOG SIGNALING
PI3K AKT MTOR SIGNALING
WNT BETA CATENIN SIGNALING
NOTCH SIGNALING
MTORC1 SIGNALING
P53 PATHWAY
INTERFERON GAMMA RESPONSE
GLYCOLYSIS
EPITHELIAL MESENCHYMAL TRANSITION
OXIDATIVE PHOSPHORYLATION
E2F TARGETS
MYC TARGETS V1
APOPTOSIS
KRAS SIGNALING DN
MYC TARGETS V2
TGF BETA SIGNALING
INTERFERON ALPHA RESPONSE
IL2 STAT5 SIGNALING
IL6 JAK STAT3 SIGNALING
UNFOLDED PROTEIN RESPONSE
TNFA SIGNALING VIA NFKB
CHOLESTEROL HOMEOSTASIS
BILE ACID METABOLISM
−0.5
0.0
0.5
Mean GSVA score


E
Treatment
CLK1 Exon 4 Morpholino
Non−targeting Morpholino
GSVA Scores
−1 −0.5 0 0.5 1
Treatment
Mismatch Repair
Homologous Recombination
DNA Repair all genes
Nucleotide Excision Repair
Non−homologous End Joining

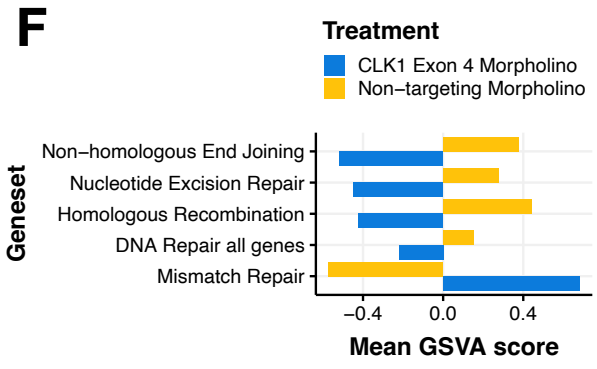
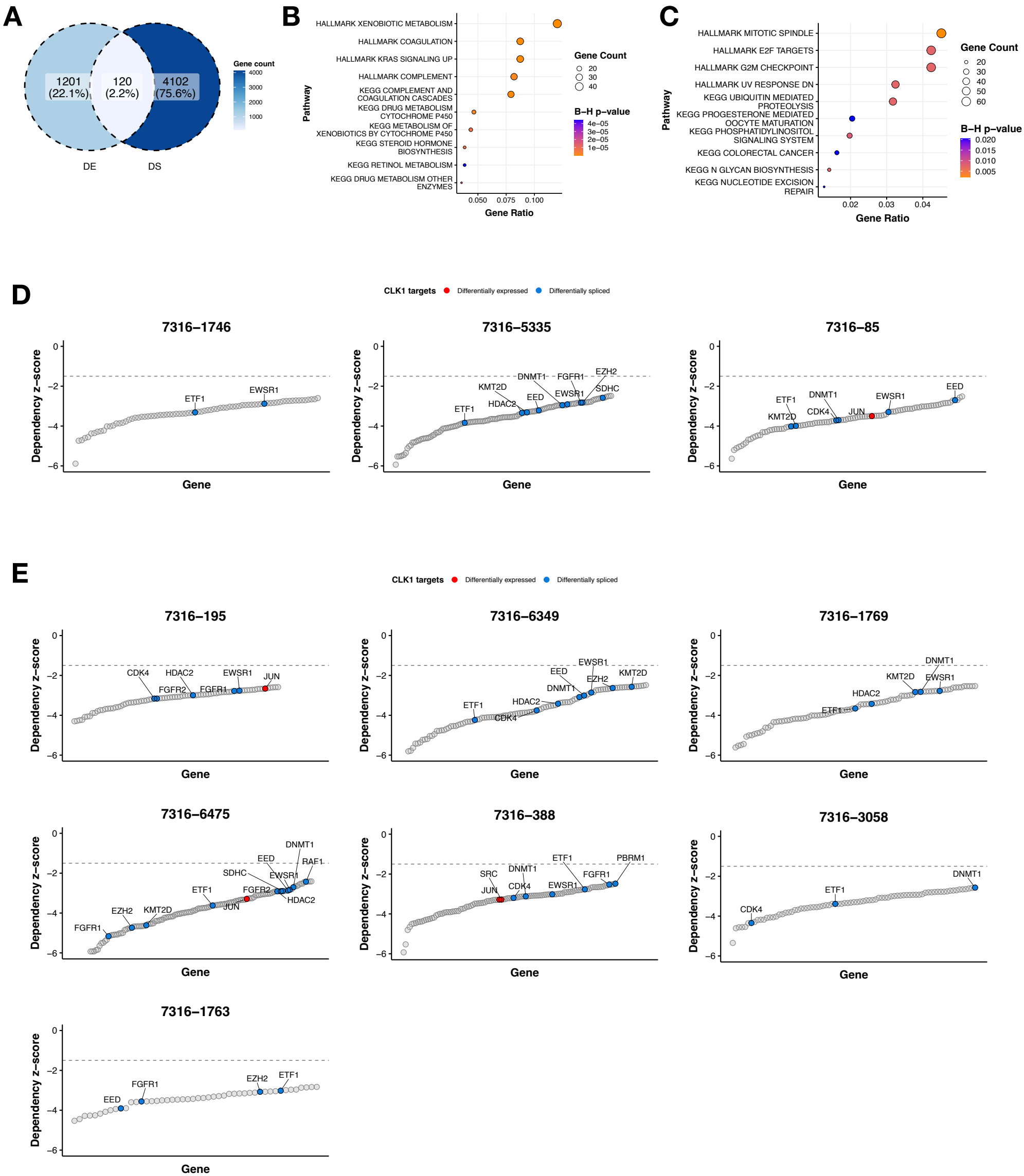

F
Treatment
CLK1 Exon 4 Morpholino
Non−targeting Morpholino
Geneset
Non−homologous End Joining
Nucleotide Excision Repair
Homologous Recombination
DNA Repair all genes
Mismatch Repair
−0.4
0.0
0.4
Mean GSVA score

Figure S5


A
1201 (22.1%)
120 (2.2%)
4102 (75.6%)
DE
DS
Gene count
4000
3000
2000
1000
B
HALLMARK XENOBIOTIC METABOLISM
HALLMARK COAGULATION
HALLMARK KRAS SIGNALING UP
HALLMARK COMPLEMENT
KEGG COMPLEMENT AND COAGULATION CASCADES
KEGG DRUG METABOLISM CYTOCHROME P450
KEGG METABOLISM OF XENOBIOTICS BY CYTOCHROME P450
KEGG STEROID HORMONE BIOSYNTHESIS
KEGG RETINOL METABOLISM
KEGG DRUG METABOLISM OTHER ENZYMES
Pathway
0.050 0.075 0.100
Gene Ratio
Gene Count
20
30
40
B–H p–value
4e–05
3e–05
2e–05
1e–05
C
HALLMARK MITOTIC SPINDLE
HALLMARK E2F TARGETS
HALLMARK G2M CHECKPOINT
HALLMARK UV RESPONSE DN
KEGG UBIQUITIN MEDIATED PROTEOLYSIS
KEGG PROGESTERONE MEDIATED OOCYTE MATURATION
KEGG PHOSPHATIDYLINOSITOL SIGNALING SYSTEM
KEGG COLORECTAL CANCER
KEGG N GLYCAN BIOSYNTHESIS
KEGG NUCLEOTIDE EXCISION REPAIR
Pathway
0.02 0.03 0.04
Gene Ratio
Gene Count
20
30
40
50
60
B–H p–value
0.020
0.015
0.010
0.005
D
CLK1 targets
Differentially expressed
Differentially spliced
7316–1746
ETF1
EWSR1
7316–5335
ETF1
HDAC2
KMT2D
EED
DNMT1
EWSR1
FGFR1
SDHC
EZH2
7316–85
ETF1
KMT2D
DNMT1
CDK4
JUN
EWSR1
EED
Dependency z–score
Gene
E
CLK1 targets
Differentially expressed
Differentially spliced
7316–195
CDK4
FGFR2
HDAC2
FGFR1
EWSR1
JUN
7316–6349
ETF1
CDK4
HDAC2
DNMT1
EED
EWSR1
EZH2
KMT2D
7316–1769
ETF1
HDAC2
KMT2D
EWSR1
DNMT1
7316–6475
FGFR1
EZH2
KMT2D
ETF1
JUN
FGFR2
SDHC
EWSR1
EED
DNMT1
RAF1
HDAC2
7316–388
SRC
JUN
CDK4
DNMT1
ETF1
EWSR1
FGFR1
PBRM1
7316–3058
CDK4
ETF1
DNMT1
7316–1763
EED
FGFR1
EZH2
ETF1
Dependency z–score
Gene