

# An Open Pediatric Brain Tumor Atlas

This manuscript ([permalink](#)) was automatically generated from [AlexsLemonade/OpenPBTA-manuscript@4db896a](#) on May 25, 2022.

## Authors

---

- **Joshua A. Shapiro**

 [0000-0002-6224-0347](#) ·  [jashapiro](#) ·  [jashapiro](#)

Childhood Cancer Data Lab, Alex's Lemonade Stand Foundation, Bala Cynwyd, PA, USA · Funded by Alex's Lemonade Stand Foundation Childhood Cancer Data Lab (CCDL)

- **Candace L. Savonen**

 [0000-0001-6331-7070](#) ·  [cansavvy](#) ·  [cansavvy](#)

Childhood Cancer Data Lab, Alex's Lemonade Stand Foundation, Bala Cynwyd, PA, USA · Funded by Alex's Lemonade Stand Foundation Childhood Cancer Data Lab (CCDL)

- **Chante J. Bethell**

 [0000-0001-9653-8128](#) ·  [cbethell](#) ·  [cjbethell](#)

Childhood Cancer Data Lab, Alex's Lemonade Stand Foundation, Bala Cynwyd, PA, USA · Funded by Alex's Lemonade Stand Foundation Childhood Cancer Data Lab (CCDL)

- **Krutika S. Gaonkar**

 [0000-0003-0838-2405](#) ·  [kgaonkar6](#) ·  [aggokittu](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia; Department of Bioinformatics and Health Informatics, Children's Hospital of Philadelphia

- **Run Jin**

 [0000-0002-8958-9266](#) ·  [runjin326](#) ·  [runjin](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Yuankun Zhu**

 [0000-0002-2455-9525](#) ·  [yuankunzhu](#) ·  [zhuyuankun](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Miguel A. Brown**

 [0000-0001-6782-1442](#) ·  [migbro](#) ·  [migbro](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Nhat Duong**

 [0000-0003-2852-4263](#) ·  [fingerfen](#) ·  [asiannhat](#)

- **Komal S. Rathi**

 [0000-0001-5534-6904](#) ·  [komalsrathi](#) ·  [komalsrathi](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Department of Bioinformatics and Health Informatics, Children's Hospital of Philadelphia

- **Nighat Noureen**

 [0000-0001-7495-8201](#) ·  [NNoureen](#)

Greehey Children's Cancer Research Institute, UT Health San Antonio

- **Bo Zhang**

 [0000-0002-0743-5379](#) ·  [zhangb1](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Brian M. Ennis**

 [0000-0002-2653-5009](#) ·  [bmennis](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Stephanie J. Spielman**

 [0000-0002-9090-4788](#) ·  [sjspielman](#) ·  [stephspiel](#)

Childhood Cancer Data Lab, Alex's Lemonade Stand Foundation, Bala Cynwyd, PA, USA<sup>✉</sup>; Rowan University, Glassboro, NJ, USA · Funded by Alex's Lemonade Stand Foundation Childhood Cancer Data Lab (CCDL)

<sup>✉</sup>Current affiliation

- **Laura E. Egolf**

 [0000-0002-7103-4801](#) ·  [LauraEgolf](#) ·  [LauraEgolf](#)

Cell and Molecular Biology Graduate Group, Perelman School of Medicine at the University of Pennsylvania; Division of Oncology, Children's Hospital of Philadelphia

- **Yang Yang**

·  [yangyangclover](#)

Ben May Department for Cancer Research, University of Chicago, Chicago IL, USA

- **Bailey Farrow**

 [0000-0001-6727-6333](#) ·  [baileyckelly](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Nicolas Van Kuren**

 [0000-0002-7414-9516](#) ·  [nicholasvk](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Meen Chul Kim**

 [0000-0002-0308-783X](#) ·  [liberaliscomputing](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Tejaswi Koganti**

 [0000-0002-7733-6480](#) ·  [tkoganti](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Shrivats Kannan**

 [0000-0002-1460-920X](#) ·  [shrivatsk](#) ·  [kshrivats](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Pichai Raman**

 [0000-0001-6948-2157](#) ·  [pichairaman](#) ·  [PichaiRaman](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Department of Bioinformatics and Health Informatics, Children's Hospital of Philadelphia

- **Jennifer Mason**

·  [jenn0307](#) ·  [jenn0307](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Daniel P. Miller**

 [0000-0002-2032-4358](#) ·  [dmiller15](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Anna R. Poetsch**

 [0000-0003-3056-4360](#) ·  [arpoe](#) ·  [APoetsch](#)

Biotechnology Center, Technical University Dresden, Germany; National Center for Tumor Diseases, Dresden, Germany

- **Payal Jain**

 [0000-0002-5914-9083](#) ·  [jainpayal022](#) ·  [jainpayal022](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Adam A. Kraya**

 [0000-0002-8526-5694](#) ·  [aadamk](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Allison P. Heath**

 [0000-0002-2583-9668](#) ·  [allisonheath](#) ·  [allig8r](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia · Funded by NIH U2C HL138346-03; NCI/NIH Contract No. 75N91019D00024, Task Order No. 75N91020F00003; Australian Government, Department of Education

- **Mateusz P. Koptyra**

 [0000-0002-3857-6633](#) ·  [mkoptyra](#) ·  [koptyram](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Shannon Robbins**

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Yiran Guo**

 [0000-0002-6549-8589](#) ·  [Yiran-Guo](#) ·  [YiranGuo3](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Xiaoyan Huang**

 [0000-0001-7267-4512](#) ·  [HuangXiaoyan0106](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Jessica Wong**

 [0000-0003-1508-7631](#) ·  [wongjessica93](#) ·  [jessicawongbfx](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Mariarita Santi**

 [0000-0002-6728-3450](#)

Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia; Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine

- **Angela Viaene**

 [0000-0001-6430-8360](#)

Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia; Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine

- **Laura Scolaro**

Division of Oncology, Children's Hospital of Philadelphia

- **Angela Waanders**

 [0000-0002-0571-2889](#) ·  [awaanders](#)

Department of Oncology, Ann & Robert H. Lurie Children's Hospital of Chicago; Department of Pediatrics, Northwestern University Feinberg School of Medicine

- **Derek Hanson**

 [0000-0002-0024-5142](#)

Hackensack Meridian School of Medicine; Hackensack University Medical Center

- **Steven M. Foltz**

 [0000-0002-9526-8194](#) ·  [envest](#)

Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania; Childhood Cancer Data Lab, Alex's Lemonade Stand Foundation, Bala Cynwyd, PA, USA · Funded by Alex's Lemonade Stand Foundation GR-000002471; National Institutes of Health K12GM081259

- **Hongbo M. Xie**

 [0000-0003-2223-0029](#) ·  [xiehongbo](#) ·  [xiehb](#)

Department of Bioinformatics and Health Informatics, Children's Hospital of Philadelphia

- **Siyuan Zheng**

 [0000-0002-1031-9424](#) ·  [syzheng](#) ·  [zhengsiyuan](#)

Greehey Children's Cancer Research Institute, UT Health San Antonio

- **Cassie N. Kline**

 [0000-0001-7765-7690](#) ·  [cnkline13](#)

Division of Oncology, Children's Hospital of Philadelphia

- **Peter J. Madsen**

 [0000-0001-9266-3685](#) ·  [petermadsenmd](#)

Division of Neurosurgery, Children's Hospital of Philadelphia; Center for Data-Driven Discovery, Children's Hospital of Philadelphia

- **Jena V. Lilly**

 [0000-0003-1439-6045](#) ·  [jvlilly](#) ·  [jvlilly](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia

- **Philip B. Storm**

 [0000-0002-7964-2449](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia · Funded by Alex's Lemonade Stand Foundation (Catalyst); Children's Hospital of Philadelphia Division of Neurosurgery

- **Adam C. Resnick**

 [0000-0003-0436-4189](#) ·  [adamcresnick](#) ·  [adamcresnick](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia · Funded by Alex's Lemonade Stand Foundation (Catalyst); Children's Brain Tumor Network; NIH 3P30 CA016520-44S5, U2C HL138346-03, U24 CA220457-03; NCI/NIH Contract No. 75N91019D00024, Task Order No. 75N91020F00003; Children's Hospital of Philadelphia Division of Neurosurgery

- **Casey S. Greene**

 [0000-0001-8713-9213](#) ·  [cgreenie](#) ·  [greenescientist](#)

Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; Childhood Cancer Data Lab, Alex's Lemonade Stand Foundation, Bala Cynwyd, PA, USA; Center for Health AI, University of Colorado School of Medicine, Aurora, CO, USA; Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO, USA · Funded by Alex's Lemonade Stand Foundation Childhood Cancer Data Lab (CCDL)

- **Jo Lynne Rokita\***

 [0000-0003-2171-3627](#) ·  [jharenza](#) ·  [jolynnerokita](#)

Center for Data-Driven Discovery, Children's Hospital of Philadelphia; Division of Neurosurgery, Children's Hospital of Philadelphia; Department of Bioinformatics and Health Informatics, Children's Hospital of Philadelphia · Funded by Alex's Lemonade Stand Foundation (Young Investigator, Catalyst); NCI/NIH Contract No. 75N91019D00024, Task Order No. 75N91020F00003

- **Jaclyn N. Taroni\***

 [0000-0003-4734-4508](#) ·  [jaclyn-taroni](#) ·  [jaclyn\\_taroni](#)

Childhood Cancer Data Lab, Alex's Lemonade Stand Foundation, Bala Cynwyd, PA, USA · Funded by Alex's Lemonade Stand Foundation Childhood Cancer Data Lab (CCDL)

- **Children's Brain Tumor Network**

- **Pacific Pediatric Neurooncology Consortium**

## Contact information

---

\*Correspondence: [jaclyn.taroni@ccdatalab.org](mailto:jaclyn.taroni@ccdatalab.org), [rokita@chop.edu](mailto:rokita@chop.edu)

## In Brief

---

The OpenPBTA is a global, collaborative open-science initiative which brought together researchers and clinicians to genetically characterize over 1,000 pediatric brain tumors. Shapiro, et. al create over 40 open-source, scalable modules to perform cancer genomics analyses and provide a richly-annotated somatic dataset across 59 brain tumor histologies. The OpenPBTA framework can be used

as a model for large-scale data integration to inform basic research, therapeutic target identification, and clinical translation.

## Highlights

---

OpenPBTA collaborative analyses establish resource for 1,253 pediatric brain tumors

NGS-based WHO-aligned integrated diagnoses generated for 1,260 of 2,840 biospecimens

RNA-Seq analysis infers medulloblastoma subtypes, TP53 status, and telomerase activity

OpenPBTA will accelerate therapeutic translation of genomic insights

## Summary

---

As pediatric brain and spinal cord tumors are the leading disease-related cause of death in children, we urgently need novel, curative therapeutic strategies for these tumors. To accelerate such discoveries, the Children's Brain Tumor Network and Pacific Pediatric Neuro-Oncology Consortium created a systematic process for tumor biobanking, model generation, and sequencing while providing immediate access to harmonized data. Here, we leverage these data to create the OpenPBTA project in which we openly and collaboratively genetically characterize and molecularly subtype, to our knowledge, the largest cohort of pediatric brain tumor specimens and histologies to date. The OpenPBTA establishes over 40 scalable analysis modules to analyze these data and recapitulate key oncogenic drivers across histologies. The OpenPBTA has already emerged as a foundational analysis platform actively being applied to study other pediatric cancers and inform molecular tumor board decision-making, making it an invaluable resource to the pediatric oncology community.

## Keywords

---

pediatric brain tumors, somatic variation, open science, reproducibility, classification

## Introduction

---

Pediatric brain and spinal cord tumors are the second most common tumors in children after leukemia, and they represent the leading disease-related cause of death in children [1]. Five-year survival rates vary widely across different histologic and molecular classifications of brain tumors. For example, most high-grade and embryonal tumors carry a universally fatal prognosis, while children with pilocytic astrocytoma have an estimated 10-year survival rate of 92% [2]. Moreover, estimates from 2009 suggest that children and adolescents aged 0-19 with brain tumors in the United States have lost an average of 47,631 years of potential life [3].

The low survival rates for some pediatric tumors are clearly multifactorial, explained partly by our lack of comprehensive understanding of the ever-evolving array of brain tumor molecular subtypes, difficulty drugging these tumors, and the shortage of drugs specifically labeled for pediatric malignancies. Historically, some of the most fatal, inoperable brain tumors, such as diffuse midline gliomas, were not routinely biopsied due to perceived risks of biopsy and the paucity of therapeutic options that would require tissue. Limited access to tissue to develop patient-derived cell lines and mouse models has been a barrier to research. Furthermore, the incidence of any single brain tumor molecular subtype is relatively low due to the rarity of pediatric tumors in general. To address these

long-standing barriers, multiple national and international consortia have come together to collaboratively share specimens and data to accelerate breakthroughs and clinical translation.

Although there has been significant progress in recent years to elucidate the landscape of somatic variation responsible for pediatric brain tumor formation and progression, translation of therapeutic agents to phase II or III clinical trials and subsequent FDA approvals have not kept pace. Within the last 20 years, the FDA has approved only five drugs for the treatment of pediatric brain tumors: mTOR inhibitor, everolimus, for subependymal giant cell astrocytoma [4,5]; anti-PD-1 immunotherapy, pembrolizumab, for microsatellite instability-high or mismatch repair-deficient tumors [6]; NTRK inhibitors larotrectinib [7] and entrectinib [8] for tumors with an NTRK 1/2/3 gene fusions; MEK1/2 inhibitor, selumetinib, for neurofibromatosis type 1 (NF1) and symptomatic, inoperable plexiform neurofibromas [9]. This is, in part, due to pharmaceutical company priorities and/or concerns regarding toxicity leading, making it challenging for researchers to obtain new therapeutic agents for pediatric clinical trials. Critically, as of August 18, 2020, an amendment to the Pediatric Research Equity Act called the “Research to Accelerate Cures and Equity (RACE) for Children Act” mandates that all new adult oncology drugs also be tested in children when the molecular targets are relevant to a particular childhood cancer. This regulatory change is poised to accelerate identification of novel and effective therapeutic for pediatric diseases that have otherwise been overlooked. Here, we present a comprehensive, collaborative, open genomic analysis of nearly 2,000 tumor specimens and 38 cell lines, comprised of 59 distinct brain tumor histologies from 943 patients. These analyses can be used to support the RACE Act in rational clinical trial design to include children.

## Results

---

### Crowd-sourced Somatic Analysis to create an Open Pediatric Brain Tumor Atlas

We previously performed whole genome sequencing (WGS), whole exome sequencing (WXS), and RNA sequencing (RNA-Seq) on matched tumor and normal tissues as well as selected cell lines from 943 patient tumors from the Pediatric Brain Tumor Atlas (PBTA) [10], consisting of samples from the [Children's Brain Tumor Network \(CBTN\)](#) and the PNOC003 DMG clinical trial [11] of the Pacific Pediatric Neuro-Oncology Consortium (PNOC) (**Figure 1A**). We then harnessed the benchmarking efforts of the Kids First Data Resource Center to develop a robust and reproducible data analysis workflow within the CAVATICA platform to perform primary somatic analyses including variant calling of single nucleotide variants (SNVs), copy number variants (CNVs), structural variants (SVs), and fusions (**Figure S1**) - red boxes and **STAR Methods**). To facilitate analysis and visualization of this large, diverse cohort, we further categorized tumor broad histologies (i.e., broad 2016 WHO classifications) into smaller groupings we denote “cancer groups.” A summarized view of the number of biospecimens per phase of therapy across different broad histologies and cancer groups is shown in (**Figure 1B**). We maintained a data release folder on Amazon S3, downloadable directly from S3 or through the open-access [CAVATICA project](#), with merged files for each analysis (See **Data and code availability** section). As new analytical results (e.g., tumor mutation burden calculations) that we expected to be used across multiple analyses were produced, or issues with the data were identified, new data releases were made available in a versioned manner.

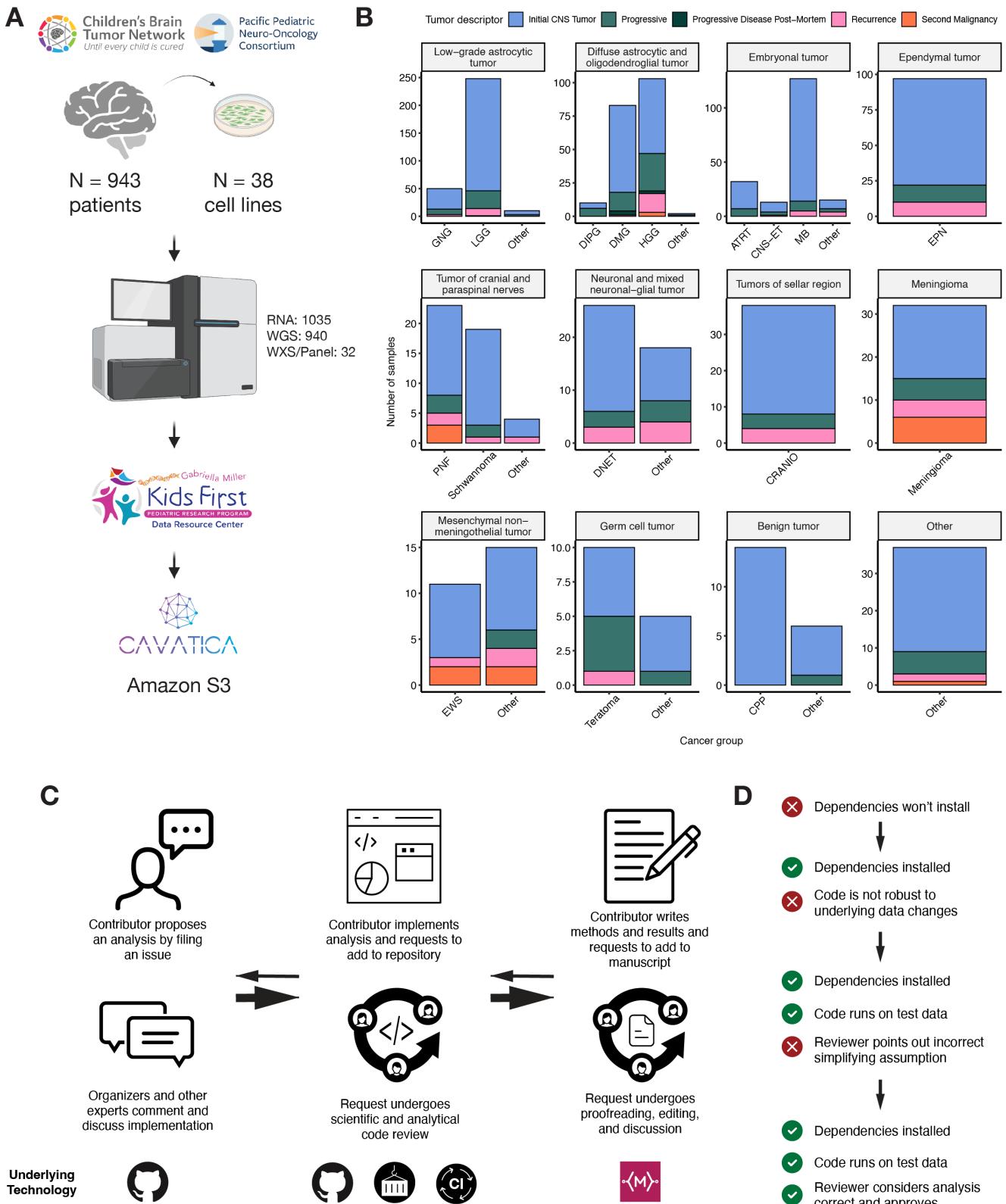
A key innovative feature of this project has been its open contribution model used for both analyses (i.e., analytical code) and scientific manuscript writing. We created a public Github analysis repository (<https://github.com/AlexsLemonade/OpenPBTA-analysis>) to hold all code associated with analyses downstream of the Kids First Data Resource Center workflows and a GitHub manuscript repository (<https://github.com/AlexsLemonade/OpenPBTA-manuscript>) with ManuBot [12] integration to enable real-time manuscript creation using Markdown within GitHub. Importantly, all analyses and

manuscript writing were conducted openly throughout the research project, allowing any researcher in the world the opportunity to contribute.

The process for analysis and manuscript contributions is outlined **Figure 1C**. First, a potential contributor would propose an analysis by filing an issue in the GitHub analysis repository. Next, organizers for the project, or other contributors with expertise, had the opportunity to provide feedback about the proposed analysis (**Figure 1C**). The contributor then forked the analysis repository (or make a copy for the purpose of adding code changes) and added their proposed analysis to their fork. The contributor would formally request to include their analytical code and results to the main OpenPBTA analysis repository by filing a pull request on GitHub. All pull requests to the analysis repository underwent peer review by organizers and/or other contributors to ensure scientific accuracy and maintainability as well as readability of code and documentation (**Figure 1C-D**).

The collaborative nature of the project required additional steps beyond peer review of analytical code (**Figure 1D**). Specifically, since multiple contributors were working on the code base and months would sometimes pass between updates to specific analysis modules, we implemented strategies to maintain a consistent software development environment that ensured code could be revised without inducing conflicts. We leveraged Docker® and the Rocker project [13] to create a monolithic image that contained all dependencies necessary for analyses in the analysis repository. To ensure that the image would build and that code would execute over time, we used the continuous integration (CI) service CircleCI® to run analytical code on a small subset of data for testing before formal code review, allowing us to detect code bugs or sensitivity to changes in the underlying data (**Figure 1D**).

We followed a similar process in our Manubot-powered [12] manuscript repository for additions to the manuscript (**Figure 1C**). Contributors forked the manuscript repository, added proposed content to their branch, and filed pull requests to the main manuscript repository with their changes. Similarly, pull requests underwent a peer review process for clarity and correctness, agreement with interpretation, and spell checking via Manubot (**Figure 1C**).



**Figure 1: Overview of the OpenPBTA Project.** A, The Children's Brain Tumor Network and the Pacific Pediatric Neuro-Oncology Consortium collected tumor samples from 943 patients. To date, 38 cell lines were created from tumor tissue, and over 2000 specimens were sequenced (N = 1035 RNA-Seq, N = 940 WGS, and N = 32 WXS or Targeted panel). Data was harmonized by the Kids First Data Resource Center using an Amazon S3 framework within CAVATICA. B, Stacked bar plot summary of the number of biospecimens per phase of therapy. Each panel denotes a broad histology and each bar denotes a cancer group. (Abbreviations: GNG = ganglioglioma, LGG = low-grade glioma, DIPG = diffuse intrinsic pontine glioma, DMG = diffuse midline glioma, EPN = ependymoma, HGG = high-grade glioma, ATRT = atypical teratoid rhabdoid tumor, CNS-ET = CNS embryonal tumor, MB = medulloblastoma, PNF = plexiform neurofibroma, DNET = dysembryoplastic neuroepithelial tumor, CRANIO = craniopharyngioma, EWS = Ewing sarcoma, CPP = choroid plexus papilloma). C, Overview of the open analysis and manuscript contribution model. In the analysis GitHub repository, a contributor would propose an analysis that other participants can comment on. Contributors would then implement the

analysis and file a request to add their changes to the analysis repository (“pull request”). Pull requests underwent review for scientific rigor and correctness of implementation. Pull requests were additionally checked to ensure that all software dependencies were included and the code was not sensitive to underlying data changes using container and continuous integration technologies. Finally, a contributor would file a pull request documenting their methods and results to the Manubot-powered manuscript repository. Pull requests in the manuscript repository were also subject to review. D, A potential path for an analytical pull request. Arrows indicate revisions to a pull request. Prior to review, a pull request was tested for dependency installation and whether or not the code would execute. Pull requests also required approval by organizers and/or other contributors, who checked for scientific correctness. Panel A created with BioRender.com.

## Molecular Subtyping of OpenPBTA CNS Tumors

Over the past two decades, experts in neurooncology have worked with the World Health Organization (WHO) to iteratively redefine the classifications of central nervous system (CNS) tumors [14,15]. More recently, in 2016 and 2021 [16,17], molecular subtypes have been integrated into these classifications. In 2011, the Children’s Brain Tumor Tissue Consortium (CBTTC), now known as the Children’s Brain Tumor Network (CBTN), opened its protocol for brain tumor and matched normal sample collection. Since the CBTN opened its protocol before molecular data were integrated into classifications, the majority of the samples within the OpenPBTA lacked molecular subtype annotations. Moreover, the OpenPBTA data does not feature methylation arrays which are now commonly used to perform molecular subtyping. Therefore, we jointly considered key genomic features of tumor entities described by the WHO in 2016, low-grade glioma (LGG) subtypes described by Ryall and colleagues [18], as well as with clinician and pathologist review to subtype 64% (1,281/2,007) of tumor biospecimens with high confidence (**Table S1**). Importantly, this collaborative molecular subtyping process allowed us to identify data entry errors (e.g., an ETMR incorrectly entered as a medulloblastoma) and mis-identified specimens (e.g., Ewing sarcoma sample labeled as a craniopharyngioma), update diagnoses using current WHO terms (e.g., tumors formerly ascribed primitive neuro-ectodermal tumor [PNET] diagnoses), and discover rarer tumor entities within the OpenPBTA (e.g., H3-mutant ependymoma, meningioma with *YAP1::FAM118B* fusion). **Table 1** lists the subtypes we defined within the OpenPBTA, comprising of LGATs (low grade astrocytic tumors; N = 569), high-grade gliomas (N = 314), embryonal tumors (N = 229), ependymomas (N = 65), tumors of sellar region (N = 51), mesenchymal non-meningothelial tumors (N = 21), glialneuronal tumors (N = 20), and chordomas (N = 12). or detailed methods, see **STAR Methods** and **Figure S1**.

Broad histology	OpenPBTA molecular subtype	n
Chordoma	CHDM, conventional	4
Chordoma	CHDM, poorly differentiated	8
Diffuse astrocytic and oligodendroglial tumor	DMG, H3 K28	54
Diffuse astrocytic and oligodendroglial tumor	DMG, H3 K28, TP53 activated	26
Diffuse astrocytic and oligodendroglial tumor	DMG, H3 K28, TP53 loss	91
Diffuse astrocytic and oligodendroglial tumor	HGG, H3 G35	7
Diffuse astrocytic and oligodendroglial tumor	HGG, H3 G35, TP53 loss	2
Diffuse astrocytic and oligodendroglial tumor	HGG, H3 wildtype	74
Diffuse astrocytic and oligodendroglial tumor	HGG, H3 wildtype, TP53 activated	10
Diffuse astrocytic and oligodendroglial tumor	HGG, H3 wildtype, TP53 loss	45
Diffuse astrocytic and oligodendroglial tumor	HGG, IDH, TP53 activated	3
Diffuse astrocytic and oligodendroglial tumor	HGG, IDH, TP53 loss	2
Embryonal tumor	CNS Embryonal, NOS	24

<b>Broad histology</b>	<b>OpenPBTA molecular subtype</b>	<b>n</b>
Embryonal tumor	CNS HGNET-MN1	1
Embryonal tumor	CNS NB-FOXR2	5
Embryonal tumor	ETMR, C19MC-altered	8
Embryonal tumor	ETMR, NOS	1
Embryonal tumor	MB, Group3	24
Embryonal tumor	MB, Group4	91
Embryonal tumor	MB, SHH	55
Embryonal tumor	MB, WNT	20
Ependymal tumor	EPN, H3 K28	2
Ependymal tumor	EPN, PF A	6
Ependymal tumor	EPN, ST RELA	51
Ependymal tumor	EPN, ST YAP1	6
Low-grade astrocytic tumor	GNG, BRAF V600E	25
Low-grade astrocytic tumor	GNG, BRAF V600E, CDKN2A/B	2
Low-grade astrocytic tumor	GNG, FGFR	2
Low-grade astrocytic tumor	GNG, H3	2
Low-grade astrocytic tumor	GNG, IDH	4
Low-grade astrocytic tumor	GNG, KIAA1549-BRAF	10
Low-grade astrocytic tumor	GNG, MYB/MYBL1	2
Low-grade astrocytic tumor	GNG, NF1-germline	2
Low-grade astrocytic tumor	GNG, NF1-somatic, BRAF V600E	1
Low-grade astrocytic tumor	GNG, other MAPK	7
Low-grade astrocytic tumor	GNG, other MAPK, IDH	2
Low-grade astrocytic tumor	GNG, RTK	6
Low-grade astrocytic tumor	GNG, wildtype	28
Low-grade astrocytic tumor	LGG, BRAF V600E	53
Low-grade astrocytic tumor	LGG, BRAF V600E, CDKN2A/B	10
Low-grade astrocytic tumor	LGG, FGFR	16
Low-grade astrocytic tumor	LGG, IDH	6
Low-grade astrocytic tumor	LGG, KIAA1549-BRAF	222
Low-grade astrocytic tumor	LGG, KIAA1549-BRAF, other MAPK	2
Low-grade astrocytic tumor	LGG, MYB/MYBL1	4
Low-grade astrocytic tumor	LGG, NF1-germline	12
Low-grade astrocytic tumor	LGG, NF1-germline, CDKN2A/B	2
Low-grade astrocytic tumor	LGG, NF1-germline, FGFR	4
Low-grade astrocytic tumor	LGG, NF1-somatic	4
Low-grade astrocytic tumor	LGG, NF1-somatic, FGFR	2

Broad histology	OpenPBTA molecular subtype	n
Low-grade astrocytic tumor	LGG, NF1-somatic, NF1-germline, CDKN2A/B	2
Low-grade astrocytic tumor	LGG, other MAPK	23
Low-grade astrocytic tumor	LGG, RTK	22
Low-grade astrocytic tumor	LGG, RTK, CDKN2A/B	2
Low-grade astrocytic tumor	LGG, wildtype	84
Low-grade astrocytic tumor	SEGA, wildtype	6
Mesenchymal non-meningothelial tumor	EWS	21
Neuronal and mixed neuronal-glial tumor	CNC	4
Neuronal and mixed neuronal-glial tumor	EVN	2
Neuronal and mixed neuronal-glial tumor	GNT, BRAF V600E	2
Neuronal and mixed neuronal-glial tumor	GNT, KIAA1549-BRAF	4
Neuronal and mixed neuronal-glial tumor	GNT, other MAPK	2
Neuronal and mixed neuronal-glial tumor	GNT, other MAPK, FGFR	2
Neuronal and mixed neuronal-glial tumor	GNT, RTK	4
Tumors of sellar region	CRANIO, ADAM	51
	Total	1281

Table 1: Molecular subtypes determined across OpenPBTA samples.

## Somatic Mutational Landscape of Pediatric Brain Tumors

We performed a comprehensive genomic analysis of somatic SNVs, CNVs, SVs, and fusions across 1,969 tumors (N = 1,019 RNA-Seq, N = 1,719 WGS, N = 64 WXS/Panel) and 38 cell lines (N = 16 RNA-Seq, N = 22 WGS) from 943 patients. Following SNV consensus calling ([Figure S1](#) and [Figure S2A-G](#)), we observed lower expected tumor mutation burden (TMB) [Figure S2H](#) in pediatric tumors compared to adult brain tumors from The Cancer Genome Atlas (TCGA), [Figure S2I](#), with hypermutant (> 10 Mut/Mb) and ultra-hypermutant (> 100 Mut/Mb) tumors [[19](#)] only found within HGGs.

### Low-grade astrocytic tumors

[Figure 2A](#) depicts an oncprint of driver genes for 227 primary low-grade astrocytic tumors. As expected, the majority (62%, 140/227) of these tumors harbored a somatic alteration in *BRAF*, with canonical *BRAF*:*KIAA1549* fusions as the major oncogenic driver [[20](#)]. We observed additional mutations in *FGFR1* (2%), *PIK3CA* (2%), *KRAS* (2%), *TP53* (1%), and *ATRX* (1%) and fusions in *NTRK2* (2%), *RAF1* (2%), *MYB* (1%), *QKI* (1%), *ROS1* (1%), and *FGFR2* (1%), concordant with previous studies reporting the near universal upregulation of the RAS/MAPK pathway in these tumors resulting from activating mutations and/or oncogenic fusions [[18,20](#)]. Indeed, we observed significant upregulation (ANOVA p < 0.01) of the KRAS signaling pathway in LGATs ([Figure 5B](#)).

### Embryonal tumors

[Figure 2B](#) shows the mutational landscape for 128 primary embryonal tumors. The majority (N = 95) of samples were medulloblastomas that spanned the spectrum of molecular subtypes (WNT, SHH, Group3, and Group 4; see **Molecular Subtyping of CNS Tumors**), as identified by subtype-specific

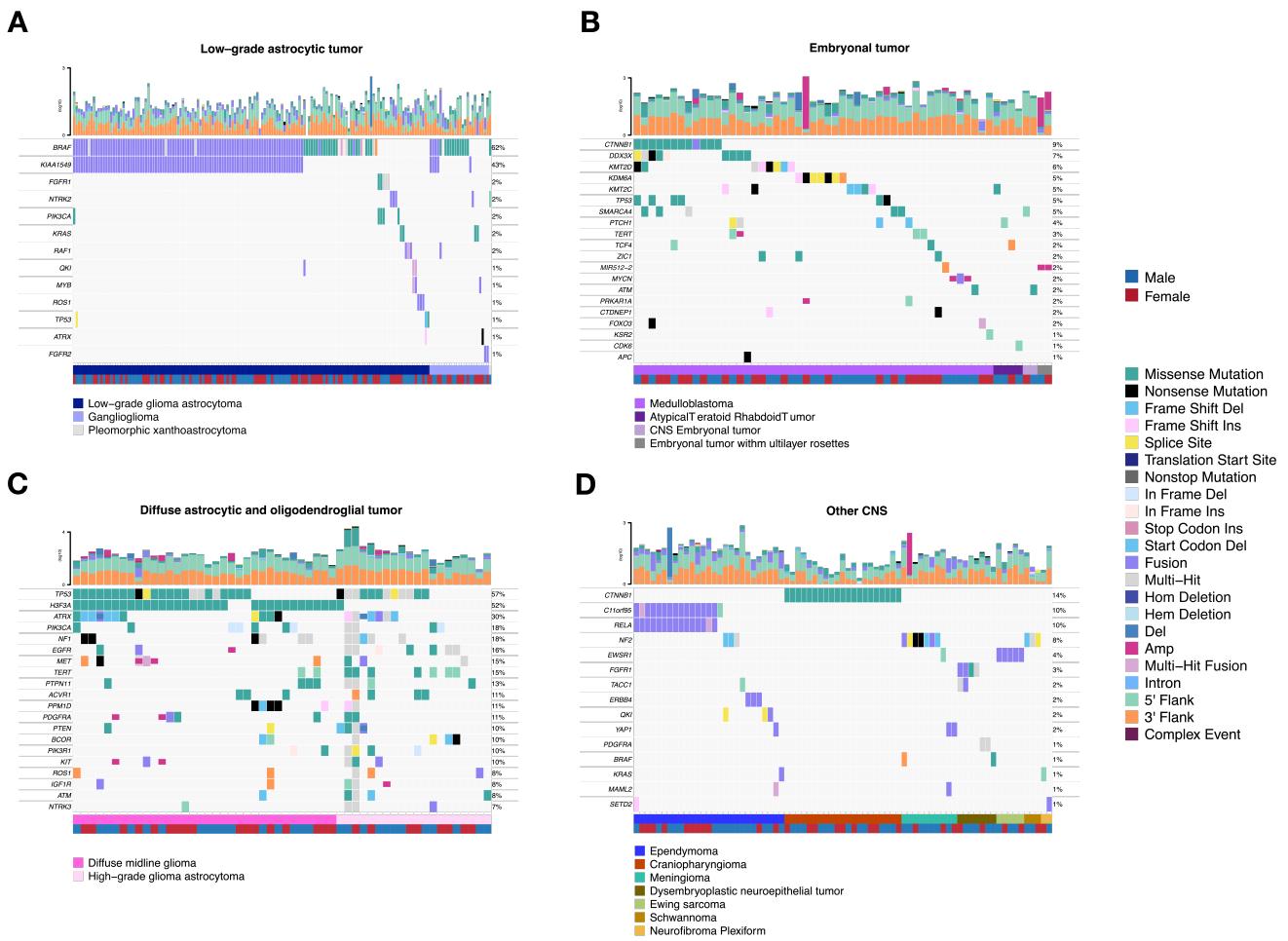
canonical mutations. We detected canonical *SMARCB1/SMARCA4* deletions or inactivating mutations in atypical teratoid rhabdoid tumors (ATRTs) and C19MC amplification in the embryonal tumors with multilayer rosettes (ETMRs) [21,22,23,24].

## Diffuse astrocytic and oligodendroglial tumors

In **Figure 2C**, we show genomic alterations in 61 HGGs biopsied at diagnosis: diffuse midline gliomas (DMGs, N = 34) and non-midline high-grade gliomas (HGGs, N = 26). The single oligodendrogloma sample (N = 1) in the OpenPBTA does not contain mutations in the genes shown and is therefore shown in this oncoprint. Across diffuse astrocytic and oligodendroglial tumors, we found that *TP53* (57%, 35/61) and *H3F3A* (52%, 32/61) were both most mutated and co-occurring genes (**Figure 2A**), followed by frequent mutations in *ATRX* (30% 18/61). We found recurrent amplifications and fusions in *EGFR*, *MET*, *PDGFRA*, and *KIT*, highlighting that these tumors utilize multiple oncogenic mechanisms to activate tyrosine kinases, as has been previously reported [11,25,26]. Gene set enrichment analysis showed upregulation (ANOVA Bonferroni-corrected p < 0.01) of DNA repair, G2M checkpoint, and MYC pathways as well as downregulation of the TP53 pathway (**Figure 5B**). The two tumors with ultra-high TMB (> 100 Mutations/Mb) were from patients with known mismatch repair deficiency syndrome [10].

## Other CNS tumors

**Figure 2D** depicts an oncoprint for the remaining primary CNS tumors in the OpenPBTA (N = 195). We observed that 25% (15/60) of ependymoma tumors were *C11orf95::RELA* (now, *ZFTA::RELA*) fusion-positive ependymomas and that 70% (21/30) of craniopharyngiomas were driven by mutations in *CTNNB1*. Multiple histologies contained somatic mutations or fusions in *NF2*: 41% (7/17) of meningiomas, 5% (3/60) of ependymomas, and 27% (3/11) schwannomas. Rare fusions in *ERBB4*, *YAP1*, *KRAS*, and *MAML2* were observed in 10% (6/60) of ependymoma tumors. DNETs harbored alterations in MAPK/PI3K pathway genes as previously reported [27], including *FGFR1* (21%, 4/19), *PDGFRA* (10%, 2/19), and *BRAF* (5%, 1/19). Frequent mutations in rarer brain tumor histologies (N < 5 samples in the OpenPBTA) are depicted in **Figure S3A**.



**Figure 2: Mutational landscape of PBTA tumors.** Shown are frequencies of canonical somatic gene mutations, CNVs, fusions, and TMB (top bar plot) for the top 20 genes mutated across primary tumors within the OpenPBTA dataset. A, Low-grade astrocytic tumors ( $N = 227$ ): low-grade glioma astrocytoma ( $N = 187$ ), ganglioglioma ( $N = 35$ ), subependymal giant cell astrocytoma ( $N = 2$ ), diffuse fibrillary astrocytoma ( $N = 1$ ), pilocytic astrocytoma ( $N = 1$ ), and pleomorphic xanthoastrocytoma ( $N = 1$ ); B, Embryonal tumors ( $N = 128$ ): medulloblastomas ( $N = 95$ ), atypical teratoid rhabdoid tumors ( $N = 24$ ), embryonal tumors with multilayer rosettes ( $N = 2$ ), other CNS embryonal tumors ( $N = 5$ ), ganglioneuroblastoma ( $N = 1$ ), and CNS neuroblastoma ( $N = 1$ ); C, Diffuse astrocytic and oligodendroglial tumors ( $N = 61$ ): diffuse midline gliomas ( $N = 34$ ) and non-midline high-grade gliomas ( $N = 26$ ), oligodendrogloma ( $N = 1$ ); D, Other CNS tumors ( $N = 195$ ): ependymomas ( $N = 60$ ), dysembryoplastic neuroepithelial tumors ( $N = 19$ ), meningiomas ( $N = 17$ ), schwannoma ( $N = 11$ ), neurofibroma plexiform ( $N = 7$ ). Other CNS tumors with  $N < 5$  are displayed in **Figure S3A**. Tumor histology and patient sex are displayed as annotations at the bottom of each plot. Only samples with mutations in the listed genes are shown. Multiple CNVs are denoted as a complex event.

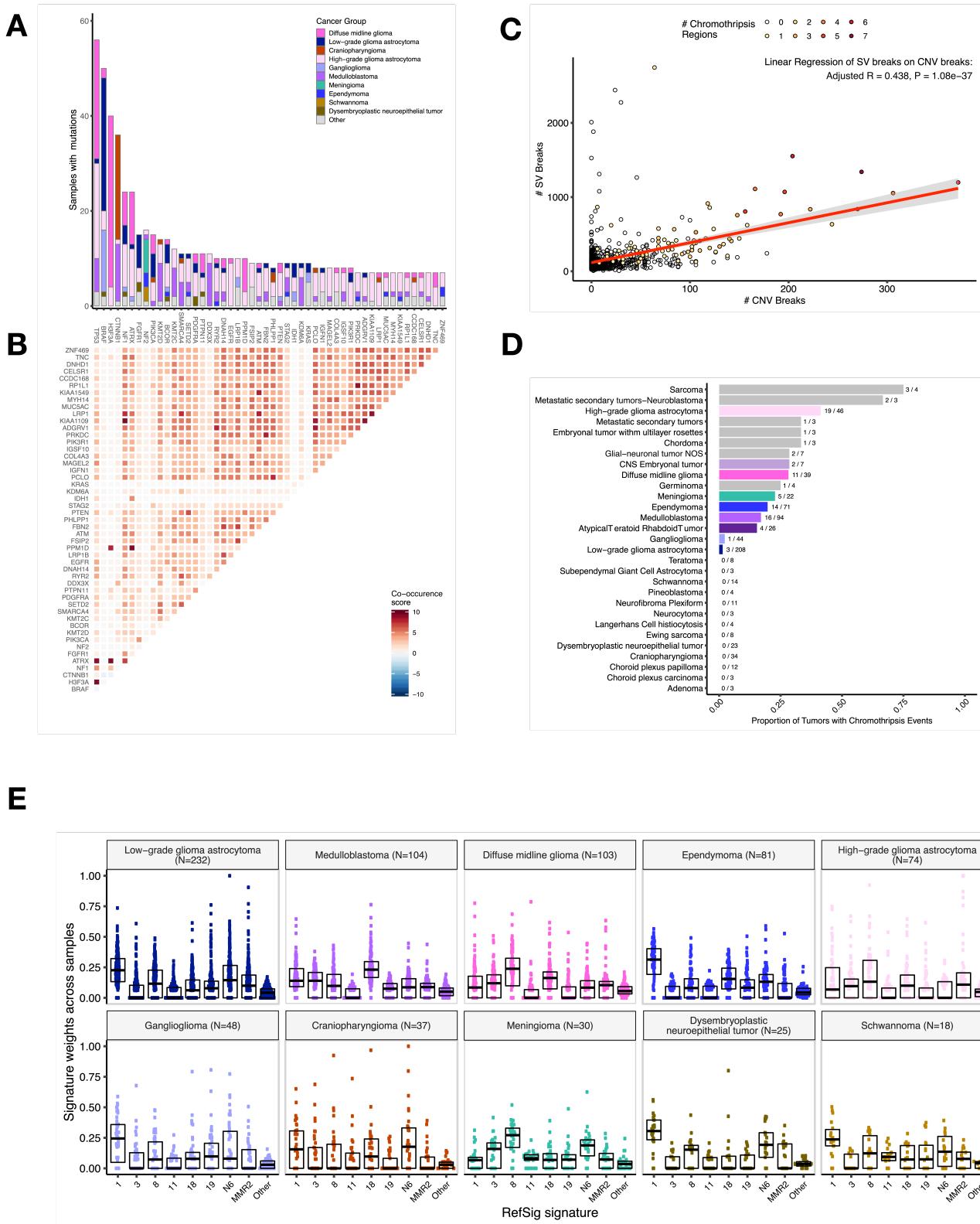
## Mutational co-occurrence, CNV and signatures highlight key oncogenic drivers

We analyzed mutational co-occurrence among OpenPBTA tumors, specifically considering only a single sequencing sample from each individual with available WGS ( $N = 666$ ). The top 50 mutated genes (see **STAR Methods** for details) in primary tumors are shown **Figure 3** by tumor type (A, bar plots), with co-occurrence scores illustrated in the heatmap (B). *TP53* was the most frequently mutated gene across OpenPBTA tumors (8.4%, 56/666), significantly co-occurring with *H3F3A* ( $OR = 32$ , 95% CI: 15.3 - 66.7,  $q = 8.46e-17$ ), *ATRX* ( $OR = 20$ , 95% CI: 8.4 - 47.7,  $q = 4.43e-8$ ), *NF1* ( $OR = 8.62$ , 95% CI: 3.7 - 20.2,  $q = 5.45e-5$ ), and *EGFR* ( $OR = 18.2$ , 95% CI: 5 - 66.5,  $q = 1.6e-4$ ). Other canonical cancer driver genes that were frequently mutated included *BRAF*, *H3F3A*, *CTNNB1*, *NF1*, *ATRX*, *FGFR1*, and *PIK3CA*. Although LGG and embryonal tumors made up the majority of tumor types within the OpenPBTA, most of the significant gene interactions stemmed from HGGs ( $N = 847/872$ ). At the broad histology level, *CTNNB1* significantly co-occurred with *TP53* ( $OR = 42.9$ , 95% CI: 7 - 261.4,  $q = 1.63e-3$ ) and *DDX3X* ( $OR = 21.1$ , 95% CI: 4.6 - 96.3,  $q = 4.46e-3$ ) in embryonal tumors, *FGFR1* and *PIK3CA* significantly co-occur in LGGs ( $OR = 76.1$ , 95% CI: 9.85 - 588.1,  $q = 3.26e-3$ ), consistent with previous

findings [28; 10.1186/s40478-020-01027-z]. We observed that 52/54 (96.3%) of tumors only have a mutation in either gene (OR = 0.188, 95% CI: 0.04 - 0.94, p = 0.0413, q = 0.0587). This trend recapitulates previous observations that *TP53* and *PPM1D* mutations tend to be mutually exclusive in HGGs [29].

We summarized broad CNV and SV and observed that HGGs and DMGs, followed by medulloblastomas, had the most unstable genomes (**Figure S3A**). By contrast, craniopharyngiomas and schwannomas generally lacked somatic CNV. Together, these CNV patterns largely aligned with our estimates of tumor mutational burden (**Figure S2H**). The number of SV and CNV breakpoints were significantly correlated across tumors (p = 1.08e-37) (**Figure 3C**) and as expected, the number of chromothripsis regions called increased as breakpoint density increased (**Figure S3B-C**). We identified chromothripsis events in 41% (N = 19/46) of non-midline high-grade gliomas (high-grade glioma astrocytomas) and in 28.2% (N = 11/39) of diffuse midline gliomas (**Figure 3D**). We also found evidence of chromothripsis in over 15% of sarcomas, metastatic secondary tumors, embryonal tumors, chordomas, glial-neuronal tumors, germinomas, meningiomas, and ependymomas, highlighting the genomic instability and complexity of these pediatric brain tumors.

We next assessed the contributions of eight previously identified adult CNS-specific mutational signatures from the RefSig database [30] across samples (**Figure 3E** and **Figure S4A**). Stage 0 and/or 1 tumors characterized by low TMBs (**Figure S2G**) such as LGGs, gangliogliomas, craniopharyngiomas, DNETs, and schwannomas were expectedly dominated by Signature 1, which results from the normal process of spontaneous deamination of 5-methylcytosine. Signature N6 is a CNS-specific signature which we observed nearly universally across samples. Drivers of Signature 18, *TP53*, *APC*, *NOTCH1* (found at <https://signal.mutationalsignatures.com/explore/referenceCancerSignature/31/drivers>), are also canonical drivers of medulloblastoma, and indeed, we observed Signature 18 as the most common signature in medulloblastoma tumors. Signatures 3, 8, 18, and MMR2 were prevalent in HGGs, including DMGs. Finally, we found that the exposure to Signature 1 was higher at diagnosis (pre-treatment) and was almost always lower in tumors at later phases of therapy (progression, recurrence, post-mortem, secondary malignancy; **Figure S4B**). This trend may have resulted from therapy-induced mutations that produced additional signatures (e.g., temozolamide treatment has been suggested to drive Signature 11 [??]), subclonal expansion, and/or acquisition of additional driver mutations during tumor progression, leading to higher overall TMBs and additional signatures.



**Figure 3: Mutational co-occurrence and signatures highlight key oncogenic drivers.** A, Bar plot of occurrence and co-occurrence of nonsynonymous mutations for the 50 most commonly mutated genes across all tumor types, which are denoted as “Other” when there are fewer than 10 samples per grouping; B, Co-occurrence and mutual exclusivity of nonsynonymous mutations between genes; The co-occurrence score is defined as  $I(-\log_{10}(P))$  where  $P$  is defined by Fisher’s exact test and  $I$  is 1 when mutations co-occur more often than expected and -1 when exclusivity is more common; C, The number of SV breaks significantly correlate with CNV breaks (Adjusted R = 0.436, p = 1.08e-37). D, Chromothripsis frequency across pediatric brain tumors for all cancer groups with N >= 3 samples. E, Sina plots of RefSig signature weights for signatures 1, 3, 8, 11, 18, 19, N6, MMR2, and Other across cancer groups. Box plot lines represent the first quartile, median, and third quartile.

## Transcriptomic Landscape of Pediatric Brain Tumors

## Prediction of *TP53* oncogenicity and telomerase activity

To understand the *TP53* phenotype in each tumor, we applied a classifier previously trained on TCGA [31] to infer *TP53* inactivation status. Using high-confidence SNVs, CNVs, SVs, and fusions in *TP53* as true positive alterations, we achieved a high accuracy (AUROC = 0.85) for rRNA-depleted, stranded samples compared to randomly shuffled *TP53* scores (Figure 4A). The classifier did not perform well on the poly-A samples (Figure S5A), potentially due to the low number of *TP53* altered (N = 29) and/or total poly-A samples in our dataset (N = 58) rather than library type, as a previous study demonstrated high accuracy of this classifier on another poly-A dataset [32]. We annotated *TP53* alterations as “activated” if samples harbored one of p.R273C or p.R248W mutations [33], “loss” if the patient had a Li Fraumeni Syndrome (LFS) predisposition diagnosis, tumor harbors a known hotspot mutation, or contains two hits (e.g.: SNV and CNV), suggesting both alleles are affected. If the *TP53* mutation did not reside within the DNA-binding domain or if an alteration was not detected in *TP53*, we annotated the tumor as “other”. While we expected that samples annotated as “loss” had higher *TP53* scores than samples annotated as “other”, we observed that those annotated as “activated” had similar *TP53* scores to those annotated as “loss” (Figure 4B, Wilcoxon p = 0.23), suggesting that the classifier detects an oncogenic, or altered, *TP53* phenotype (scores > 0.5) rather than solely *TP53* inactivation, as interpreted previously [31]. Moreover, tumors with “activating” *TP53* mutations had evidence of higher *TP53* expression than those with *TP53* “loss” mutations (Wilcoxon p = 3.5e-3, Figure 4C). To further validate the classifier’s accuracy, we assessed *TP53* scores for patients with LFS, hypothesizing that all of these tumors would have high scores. Indeed, we observed higher scores in LFS tumors (N = 8) for which we detected high-confidence *TP53* somatic alterations (Tables S1 and S3). Although we were unable to detect canonical somatic *TP53* mutations in two LFS patient tumors with low *TP53* scores (BS\_DEHJF4C7 with a score of 0.09 and BS\_ZD5HN296 with a score of 0.28, we confirmed the LFS diagnosis from pathology reports and found each to have a pathogenic germline variant in *TP53*. The tumor purity of these two LFS samples was low (16% and 37%), suggesting the classifier requires a certain level of tumor purity to achieve good performance, as we expect *TP53* to be intact in normal cells. Tumors with the highest median *TP53* scores were those known to harbor somatic *TP53* alterations: DMGs, medulloblastomas, HGGs, DNETs, ependymomas, and craniopharyngiomas (Figure 4D), while gangliogliomas, LGGs, meningiomas, and schwannomas had the lowest scores. We next used gene expression data to predict telomerase activity using EXpression-based Telomerase ENzymatic activity Detection (EXTEND) [34] as a surrogate measure of malignant potential [34; 10.1093/carcin/bgp268]. As expected, EXTEND scores significantly correlated with *TERC* ( $R = 0.619$ ,  $p < 0.01$ ) and *TERT* ( $R = 0.491$ ,  $p < 0.01$ ) expression (Figure S5B-C).

We found aggressive tumors such as CNS lymphoma, ETMR, ATRT, DMG, and HGG had high EXTEND scores (Figure 4D), while benign lesions such as teratomas, dysplasias, and hemangioblastomas had the lowest scores (Table S3). This supports previous reports of a more aggressive phenotype in tumors with higher telomerase activity [35; 10.1038/labinvest.3700710; 10.1007/s12032-016-0736-x; 10.1111/j.1750-3639.2010.00372.x].

## Hypermutant tumors share mutational signatures and have dysregulated *TP53*

We further investigated the mutational signature profiles of the hypermutant (N = 2) and ultra-hypermutant (N = 4) HGG tumors and/or derived cell lines from five patients in the OpenPBTA cohort. The relative contributions of eight RefSig mutational signatures are shown in the heatmap in Figure 4E for all tumors from patients with at least one tumor or cell line having a TMB  $\geq 10$  Mut/Mb. Signature 11, w suggested exposure to temozolomide, was indeed present in tumors with previous exposure to the drug (Table 2). We found the MMR2 signature in tumors of four patients (PT\_OSPKM4S8, PT\_3CHB9PK5, PT\_JNEV57VK, and PT\_VTM2STE3) diagnosed with Lynch syndrome (Table 2), a genetic predisposition syndrome caused by a variant in a mismatch repair gene such as *PMS2*, *MLH1*, *MSH2*, *MSH6*, or others [36]. Interestingly, the cell line derived from patient PT\_VTM2STE3’s tumor at progression was not hypermutated (TMB = 5.74) but shows exposure to only

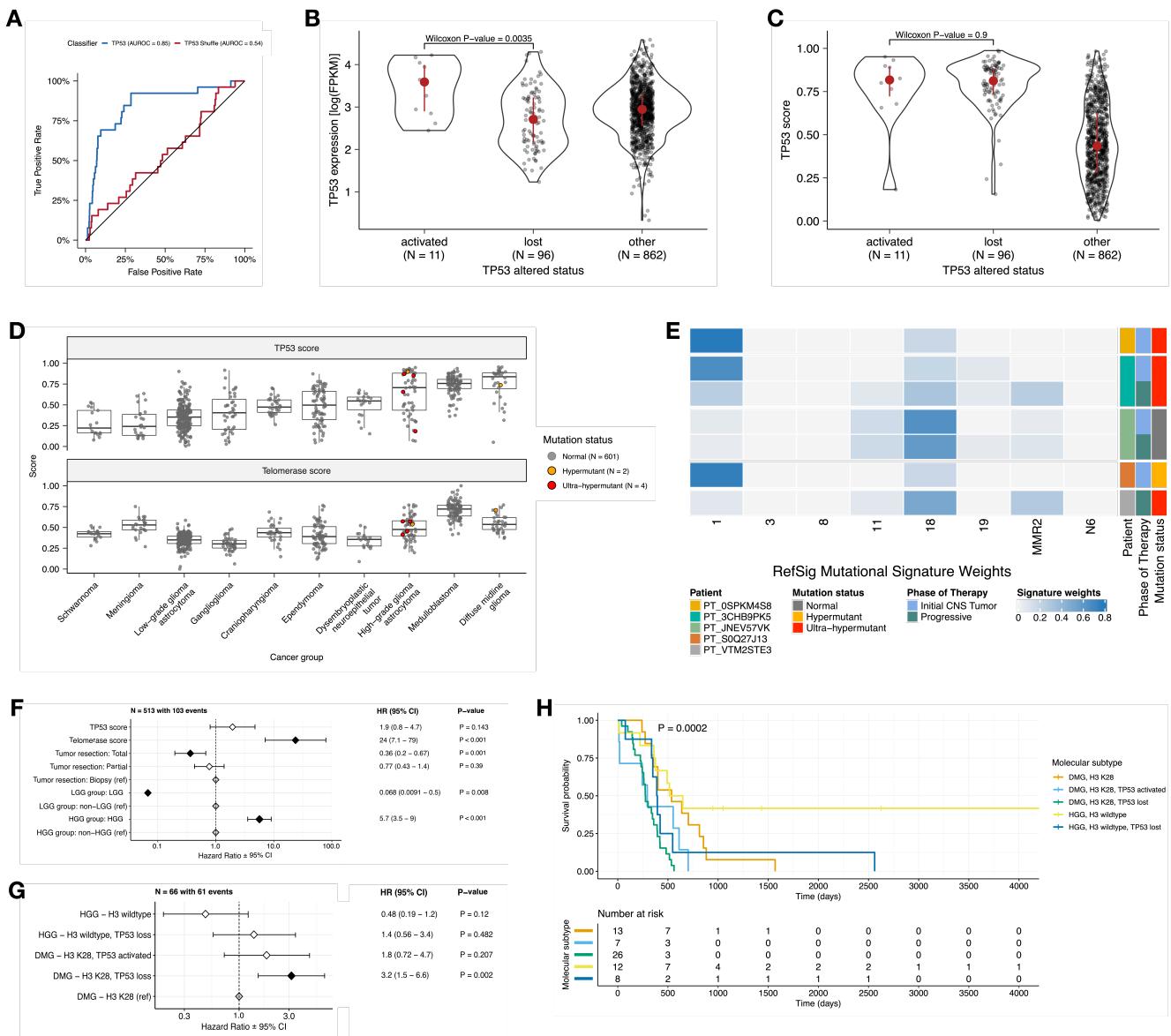
the MMR2 signature, suggesting there was selective pressure to maintain an MMR phenotype *in vitro*. From patient PT\_JNEV57VK, only one of the two cell lines derived from the progressive tumor was hypermutant (TMB of 35.9 Mut/Mb). The hypermutated cell line had a strong signature 11, while the non-hypermutated cell line from the same tumor showed a number of lesser exposures (1, 11, 18, 19, MMR2), highlighting the importance of carefully genomically characterizing and selecting models for preclinical studies based on research objectives. We observed that signature 18, which has been associated with high genomic instability and can lead to a hypermutator phenotype [30], is uniformly represented among hypermutant solid tumors. Additionally, we found that all of these tumor or cell lines have dysfunctional TP53 (**Table 2**), consistent with a previous report showing dysregulation of TP53 is a dependency in tumors with high genomic instability [30]. With one exception, hypermutant and ultra-hypermutant tumors had high TP53 scores and telomerase activity **Figure 4D**. Interestingly, none of the samples were exposed to signatures 3 (present in homologous recombination deficient tumors), signature 8 (arises from double nucleotide substitutions/unknown etiology), or signature N6 (a universal CNS tumor signature). The mutual exclusivity of signatures 3 and MMR corroborates a previous report suggesting tumors do not have both deficient homologous repair and mismatch repair [31].

Kids First Participant ID	Kids First Biospecimen ID	CB TN ID	Phase of therapy	Composition	Therapy post-biopsy	Cancer predisposition(s)	TMB	OpenPBTA molecular subtype
PT_0SPKM4S8	BS_VW4XN9Y7	73 16 - 26 40	Initial CNS Tumor	Solid Tissue	Temozolomide, CCNU, Radiation	None documented	187.387 654549 66892	HGG, H3 wildtype, TP53 activated
PT_3CHB9PK5	BS_20TBZG09	73 16 - 51 5	Initial CNS Tumor	Solid Tissue	Temozolomide, Irinotecan, Bevacizumab, Radiation	NF-1, Other inherited conditions NOS	307.049 216710 9247	HGG, H3 wildtype, TP53 loss
PT_3CHB9PK5	BS_8AY2GM4G	73 16 - 20 85	Progressive	Solid Tissue	Unknown	NF-1, Lynch syndrome, pathogenic MSH6	321.607 946782 0181	HGG, H3 wildtype, TP53 loss
PT_JNEV57VK	BS_85Q5P8GF	73 16 - 25 94	Initial CNS Tumor	Solid Tissue	Temozolomide, Radiation	Lynch syndrome	4.73158 727310 5341	DMG, H3 K28, TP53 loss
PT_JNEV57VK	BS_HM5GFJN8	73 16 - 30 58	Progressive	Derived Cell Line	Nivolumab	Lynch Syndrome, pathogenic MSH2	35.8928 691368 1093	DMG, H3 K28, TP53 loss
PT_JNEV57VK	BS_QWM9BPDY	73 16 - 30 58	Progressive	Derived Cell Line	Nivolumab	Lynch Syndrome, pathogenic MSH2	7.36335 770903 3756	DMG, H3 K28, TP53 loss

Kids First Participant ID	Kids First Biospecimen ID	CB TN ID	Phase of therapy	Composition	Therapy post-biopsy	Cancer predisposition(s )	TMB	OpenPBTA molecular subtype
PT_JNEV57 VK	BS_P0QJ1 QAH	73 16 - 30 58	Progressive	Solid Tissue	Nivolumab	Lynch Syndrome, pathogenic MSH2	6.27145 295370 1754	DMG, H3 K28, TP53 activated
PT_S0Q27J 13	BS_P3PF5 3V8	73 16 - 23 07	Initial CNS Tumor	Solid Tissue	Temozolomide, Irinotecan, Radiation	None documented	15.4826 494794 51206	HGG, H3 wildtype, TP53 activated
PT_VTM2S TE3	BS_ERFMP QN3	73 16 - 21 89	Progressive	Derived Cell Line	Unknown	Lynch Syndrome	5.73949 935495 0266	HGG, H3 wildtype, TP53 loss
PT_VTM2S TE3	BS_02YBZ SBY	73 16 - 21 89	Progressive	Solid Tissue	Unknown	Lynch Syndrome	274.488 056955 76783	HGG, H3 wildtype, TP53 activated

Table 2. Patients with hypermutant tumors.

Next, we asked whether transcriptomic classification of *TP53* dysregulation and/or telomerase activity recapitulate the known prognostic influence of these oncogenic biomarkers. We used an additive model to perform multivariate cox regression on overall survival (**STAR Methods**), controlling for extent of tumor resection and whether a tumor was low-grade (LGG group) or high-grade (HGG group). Depicted in the forest plot (**Figure 4F**), we show an expected significant overall survival benefit if the tumor was fully resected (HR = 0.36, CI = 0.2-0.67, p = 0.001) or if the tumor belongs to the LGG group (HR = 0.068, CI = 0.0091-0.5, p = 0.008). In contrast, we show an expected significant risk if the tumor belongs to the HGG group (HR = 5.7, CI = 3.5-9.0, p < 0.001). High telomerase scores were a poor prognostic indicator across brain tumor histologies (HR = 24, CI = 7.1-79, p < 0.001), demonstrating that EXTEND scores calculated from RNA-Seq can be used as a rapid surrogate measure of telomerase activity. Higher *TP53* scores, indicative of dysregulation of the *TP53* gene or pathway, while not significant across the entire OpenPBTA cohort (**Table S4**), resulted in a significant overall survival risk for both DMGs (HR = 1.77e6, CI = 1.98-1.57e12, p = 0.04) and ependymomas (HR = 1612, CI = 9.1-2.9e5, p = 0.005). Since we observed the negative prognostic effect of *TP53* for HGGs, we assessed the effect of molecular subtypes with and without *TP53* dysregulation, as determined by our *TP53* classification combined with molecular evidence. Furthermore, cox regression analysis revealed that DMG H3 K28 tumors with *TP53* loss have significantly worse prognosis (HR = 3.2, CI = 1.5-6.6, p = 0.002) than DMG H3 K28 tumors with wildtype *TP53* **Figure 4G**, which is visualized by Kaplan-Meier in **Figure 4H**.



**Figure 4: TP53 and telomerase activity** A, Receiver Operating Characteristic for TP53 classifier run on FPKM of stranded RNA-Seq samples. B, Violin and box plots of TP53 scores plotted by TP53 alteration type ( $N_{\text{activated}} = 27$ ,  $N_{\text{lost}} = 140$ ,  $N_{\text{other}} = 999$ ). C, Violin and box plots of TP53 RNA expression plotted by TP53 activation status ( $N_{\text{activated}} = 11$ ,  $N_{\text{lost}} = 96$ ,  $N_{\text{other}} = 862$ ). D, Box plots of TP53 and telomerase (EXTEND) scores grouped by cancer\_group . Mutation status is highlighted in orange (hypermutant) or red (ultra-hypermutant). E, Heatmap of RefSig mutational signatures for patients who have least one tumor or cell line with a TMB  $\geq 10$  Mut/Mb. F, Forest plot depicting the prognostic effects of TP53 and telomerase scores on overall survival, controlling for extent of tumor resection, LGG group, and HGG group. G, Forest plot depicting the effect of molecular subtype on overall survival of HGGs. For F and G, hazard ratios (HR) with 95% confidence intervals and p-values are listed. Significant p-values are denoted with black diamonds. Reference groups are denoted by grey diamonds. H, Kaplan-Meier curve of HGG tumors by molecular subtype.

## Histologic and oncogenic pathway clustering

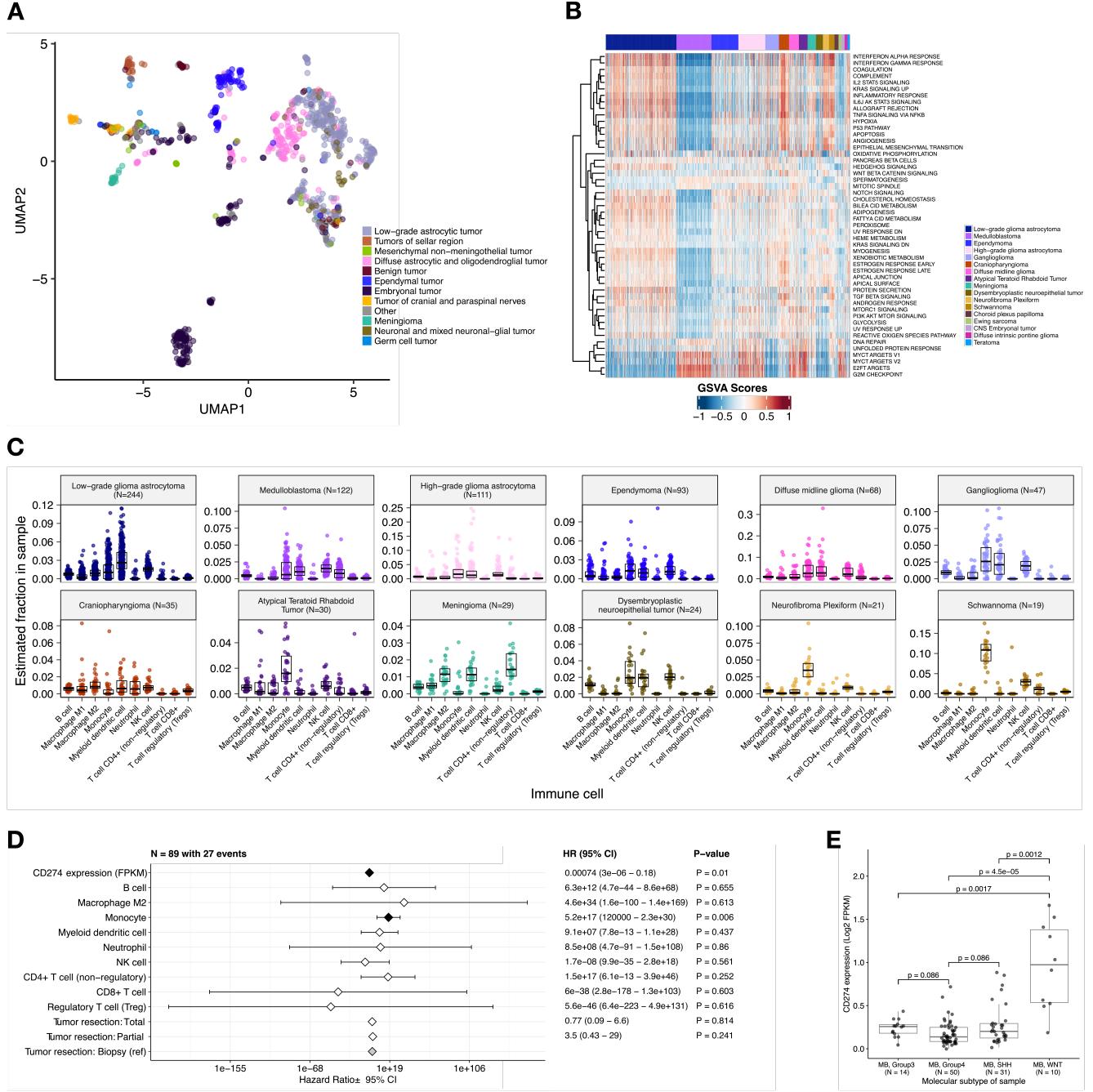
UMAP visualization of gene expression variation across brain tumors (Figure 5A) shows the expected clustering of brain tumors by histology. We observed medulloblastomas cluster by molecular subtype with WNT and SHH in distinct clusters and Groups 3 and 4 showing some overlap (Figure S6A), as expected. Of note, two samples annotated as the SHH subtype do not cluster with the MB samples and one clusters with Group 3 and 4 samples, suggesting potential subtype misclassification or different underlying biology of these tumors.

Additionally, except for three outliers, *C11orf95::RELA* (*ZFTA::RELA*) fusion positive ependymomas fall within distinct clusters (Figure S6B). *BRAF*-driven low-grade gliomas (Figure S6C) were present in three separate clusters, suggesting that there might be additional shared biology within each cluster.

Histone H3 G35-mutant HGGs generally clustered together and away from K28-mutant tumors (**Figure S6D**). Interestingly, although H3 K28-mutant tumors have different biological drivers than H3 wildtype tumors, they did not form distinct clusters. This suggests they may be driven by common transcriptional programs, have other much stronger biological drivers such as their known distinct epigenetic drivers, or our sample size is too small to detect transcriptional differences.

We next performed gene set variant analysis (GSVA) for Hallmark cancer gene sets to demonstrate activation of underlying oncogenic pathways (**Figure 5B**) and characterized the immune cell landscape across OpenPBTA tumors by running quanTlseq on RNA-Seq data (**Figure 5C** and **Figure S6E**). For example, HGG, DMG, MB, and ATRT tumors are known to upregulate *MYC* [37] which in turn activates *E2F* and S phase [38]. Indeed, we detected significant (Bonferroni  $p < 0.05$ ) upregulation of *MYC* and *E2F* targets, as well as G2M (cell cycle phase following S phase) in both embryonal and diffuse astrocytic and oligodendroglial tumors compared to several other broad histologies. In contrast, low-grade astrocytic tumors show significant downregulation (Bonferroni  $p < 0.05$ ) of these pathways. Schwannomas and neurofibromas, which have a documented inflammatory immune microenvironment of T and B lymphocytes, as well as tumor-associated macrophages (TAMs) driven by upregulation of cytokines such as IFN $\gamma$ , IL-1, and IL-6, and TNF $\alpha$  [39]. This was recapitulated in GSVA hallmark pathways (Bonferroni  $p < 0.05$ ) (**Figure 5B**) and immune cell deconvolution, showing that monocytes make up the majority of immune cells in these tumors (**Figure 5C**). We also observe significant up-regulation of pro-inflammatory cytokines IFN $\alpha$  and IFN $\gamma$  in LGG and craniopharyngiomas compared to medulloblastoma and ependymoma tumors (Bonferroni  $p < 0.05$ ), which show significant down-regulation of these cytokines (**Figure 5B**). These data support previous proteogenomic findings of lower immune infiltration in aggressive medulloblastomas and ependymomas and higher immune infiltration in *BRAF*-driven LGG and craniopharyngiomas [40]. Complete GSVA results can be found within the `gene-set-enrichment-analysis` in the [OpenPBTA analysis repository](#).

Although CD8+ T-cell infiltration across all cancer groups is quite low (**Figure 5C** and **Figure S6E**), we observed some signal in specific cancer molecular subtypes (Groups 3 and 4 medulloblastoma) as well as outlier tumors (*BRAF*-driven LGG, *BRAF*-driven and wildtype ganglioglioma, H3 wildtype HGG). Surprisingly, the classically immunologically cold HGG and DMG tumors [41; 10.1093/brain/awab155] contained higher overall fractions of immune cells, with monocytes, dendritic cells, and NK cells being the most prevalent (**Figure 5C**). Thus, we suspect that quanTlseq might actually be capturing microglia within these immune cell fractions. While we did not detect notable prognostic effects of immune cell infiltration in HGG or DMG, we did find that high levels of macrophage M1 and monocytes were associated with poorer overall survival (monocyte HR = 5.2e17, CI = 1.2e5-2.3e30,  $p = 0.006$ ) in medulloblastoma tumors (**Figure 5D**). Medulloblastomas typically have low expression of *CD274* (PD-L1) [42] a result we reproduced here (**Figure 5E**). However, we also found that higher expression of *CD274* is significantly associated with better overall prognosis (HR = 0.00074, CI = 3.0e-6 to 0.18,  $p = 0.01$ ). Consistent with a previous study, we found higher expression of *CD274* in the WNT subgroup, which has the best prognosis of all medulloblastoma subgroups [43] (**Figure 5E**). Finally, we asked whether any molecular subtypes might show an immunologically hot phenotype, as roughly defined by a greater proportion of CD8+ to CD4+ T cells and while we do see the highest ratios in Group 3 and Group 4 medulloblastomas (**Figure S6F**), very few tumors had ratios  $> 1$ , highlighting an urgent need to identify novel therapeutics for these immunologically cold pediatric brain tumors with poor prognosis.



**Figure 5: Transcriptomic and immune landscape of pediatric brain tumors** A, First two dimensions from UMAP of sample transcriptome data. Points are colored by `broad_histology` of the samples they represent. B, Heatmap of GSVA scores for Hallmark gene sets with significant differences, with samples ordered by `cancer_group`. C, Box plots of quanTlseq estimates of immune cell proportions in histologies with N > 15. D, Forest plot depicting the effect of immune cell proportion or *CD274* expression on overall survival of medulloblastomas. Hazard ratio (HR) with 95% confidence intervals and p-values are listed and significant p-values are denoted with colored diamonds. Of note, the Macrophage M1 HR was 0 (coefficient = -8.95e04) with infinite upper and lower CIs, thus it was not plotted. E, Box plot of *CD274* expression ( $\log_2$  FPKM) for medulloblastoma samples grouped by molecular subtype. Adjusted p-values are shown.

## Discussion

We created the OpenPBTA to provide an open, reproducible analysis framework within GitHub and bring together researchers and clinicians from across the globe to genetically characterize 1,253 pediatric brain tumors across 59 distinct histologies. We provide robust code and data resources to the pediatric oncology community, as well as to entice interdisciplinary scientists to collaborate on

new analyses in order to accelerate therapeutic translation for children with cancer, goals we are seeing play out in real-time. To our knowledge, this initiative represents the first large-scale, collaborative, open analysis of genomic data coupled with open manuscript writing, in which we comprehensively analyzed the largest cohort of pediatric brain tumors to date. We used available WGS, WXS, and RNA-Seq data to generate high-confidence consensus SNV and CNV calls, prioritize putative oncogenic fusions, and establish over 40 scalable modules to perform common downstream cancer genomics analyses, all of which have undergone rigorous scientific and analytical code review. We detected and showed expected frequencies of genomic lesions, mutational signatures, and aberrantly regulated signaling pathways across multiple pediatric brain tumor histologies. Molecular subtyping information from pathology reports was largely not collected nor available for this cohort, and if available (e.g.: sparse medulloblastoma subtypes), it had to be manually curated from pathology reports and/or free text clinical data fields. Furthermore, in the absence of DNA methylation data, rapid classification to derive molecular subtypes could not be performed. Thus, to enable biological interrogation of specific tumor subtypes, we created RNA- and DNA-based subtyping modules aligned with WHO molecularly-defined diagnoses. We worked closely with pathologists and clinicians to build modules from which we determined a research-grade integrated diagnosis for 44% (1260/2840) of biospecimens as well as discovered incorrectly diagnosed or mis-identified samples in the OpenPBTA cohort.

We harnessed RNA expression data for a number of analyses, yielding important biological insights across multiple brain tumor histologies. For example, from 43/122 (35.2%) medulloblastoma tumors whose pathology reports contained subtype information, we accurately predicted subtypes using MM2S (90.7%) or medulloPackage (95.4%) [44,45]. We used an expression classifier to determine whether tumors have dysfunctional *TP53* [31] and the EXTEND algorithm to determine their degree of telomerase activity using a 13-gene signature [34]. We found that hypermutant tumors universally displayed dysregulation of *TP53* and that EXTEND scores are a robust surrogate measure for telomerase activity, as tumors with high scores had a significant adverse risk for overall survival. Additionally, high *TP53* scores were a significant negative prognostic indicator for overall survival for patients with certain tumor types, such as H3 K28-altered DMGs and ependymomas. By assessing *TP53* and telomerase activity prospectively from expression data, information usually only attainable with DNA sequencing and/or qPCR, we can quickly incorporate oncogenic biomarker and prognostic knowledge and expand our biological understanding of these tumors. We identified enrichment of hallmark cancer pathways and characterized the immune cell landscape across pediatric brain tumors, demonstrating tumors in some histologies, such as schwannomas, craniopharyngiomas, and low-grade gliomas, may have an inflammatory tumor microenvironment. Of note, we observed upregulation of IFN $\gamma$ , IL-1, and IL-6, and TNF $\alpha$  in craniopharyngiomas, tumors difficult to resect due to their anatomical location and critical surrounding structures. Interferon alpha immunotherapy has been administered to reduce cystic craniopharyngioma tumor size and/or delay progression, neurotoxic side effects have been reported [46,47]. Thus, additional immune vulnerabilities, such as IL-6 inhibition and immune checkpoint blockade, have recently been proposed as therapies for cystic adamantinomatous craniopharyngiomas [48,49,50,51,52] and our results noted above support this approach. Finally, our study reproduced the overall known poor infiltration of CD8+ T cells and general low expression of *CD274*(PD-L1) in pediatric brain tumors, further highlighting the urgent need to identify novel therapeutic strategies for these immunologically cold tumors.

OpenPBTA has rapidly become a foundational layer for a number of discovery research and translational projects which will continue to add other genomic modalities and analyses, such as germline, methylation, single cell, epigenomic, mRNA splicing, imaging, and model drug response data. For example, the RNA fusion filtering module created within OpenPBTA set the stage for development of the R package *annoFuse* [53] and an R Shiny application *shinyFuse*. Using medulloblastoma subtyping and immune deconvolution analyses performed herein, Dang and colleagues showed enrichment of monocyte and microglia-derived macrophages within the SHH subgroup which they suggest may accumulate following radiation therapy [54]. Expression and copy

number analyses were used to demonstrate that *GPC2* is a highly expressed and copy number gained immunotherapeutic target in ETMRs, medulloblastomas, choroid plexus carcinomas, H3 wildtype high-grade gliomas, as well as DMGs [55]. This led Foster and colleagues to subsequently develop a chimeric antigen receptor (CAR) directed against *GPC2*, for which they show preclinical efficacy in mouse models [55]. Moreover, the OpenPBTA has enabled a framework to support real-time integration of subjects as each was enrolled on the PNOC008 high-grade glioma clinical trial [56], allowing researchers and clinicians to link tumor biology to translational impact through clinical decision support during tumor board discussions. Finally, the OpenPBTA project was recently expanded into a pan-pediatric cancer effort to support the RACE Act (<https://github.com/PediatricOpenTargets/OpenPedCan-analysis>) as part of the NCI's Childhood Cancer Data Initiative. The OpenPBTA project has paved the way for collaborative, open, reproducible, and scalable analyses and we anticipate this foundational work will have limitless impact within the pediatric oncology community, ultimately leading to improved outcomes for children with cancer.

All code and processed data are openly available through GitHub, CAVATICA, and PedcBioPortal (see **STAR METHODS**).

## Acknowledgments

---

We graciously thank the patients and families who have donated their tumors to the Children's Brain Tumor Network and/or the Pacific Pediatric Neuro-oncology Consortium, without which, this research would not be possible. This work was funded through the Alex's Lemonade Stand Foundation (ALSF) Childhood Cancer Data Lab (JNT, CSG, JAS, CLS, CJB, SJS), ALSF Young Investigator Award (JLR), ALSF Catalyst Award (JLR, ACR, PBS), ALSF Catalyst Award (SJS), ALSF CCDL Postdoctoral Training Grant (SMF), Children's Hospital of Philadelphia Division of Neurosurgery (PBS and ACR), the Australian Government, Department of Education (APH), and NIH Grants 3P30 CA016520-44S5 (ACR), U2C HL138346-03 (ACR, APH), U24 CA220457-03 (ACR), and K12GM081259 (SMF). This project has been funded in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. 75N91019D00024, Task Order No. 75N91020F00003 (JLR, ACR, APH). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the U.S. Government.

The authors would like to thank the following collaborators who contributed or supervised analyses present in the analysis repository that were not included in the manuscript: William Amadio, Holly Beale, Ellen Kephart, Alfred Lyle, and Olena Vaske. Finally, we would like to thank Yuanchao Zhang for adding to the project codebase and Jessica B. Foster for helpful discussions while drafting the manuscript.

## Author Contributions

---

Author	Contributions
Joshua A. Shapiro	Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original draft, Writing - Review and editing, Visualization, Supervision
Candace L. Savonen	Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original draft, Visualization
Chante J. Bethell	Methodology, Validation, Formal analysis, Investigation, Writing - Original draft, Visualization

<b>Author</b>	<b>Contributions</b>
Krutika S. Gaonkar	Data curation, Formal Analysis, Investigation, Methodology, Software, Writing – original draft
Run Jin	Data curation, Formal Analysis, Visualization, Writing – original draft
Yuankun Zhu	Data curation, Formal Analysis, Investigation, Methodology, Supervision
Miguel A. Brown	Data curation, Methodology
Nhat Duong	Formal Analysis, Investigation, Methodology
Komal S. Rathi	Formal Analysis, Investigation, Methodology, Writing – original draft
Nighat Noureen	Formal analysis, Visualization, Writing - Original draft
Bo Zhang	Data curation, Formal Analysis
Brian M. Ennis	Data curation, Formal Analysis
Stephanie J. Spielman	Validation, Formal analysis, Writing - Review and editing, Visualization, Supervision
Laura E. Egolf	Formal analysis, Writing - Original draft
Yang Yang	Formal analysis, Software
Bailey Farrow	Data curation, Software
Nicolas Van Kuren	Data curation, Software
Meen Chul Kim	Data curation
Tejaswi Koganti	Formal Analysis, Investigation
Shrivats Kannan	Formal Analysis, Methodology, Writing – original draft
Pichai Raman	Conceptualization, Formal Analysis, Methodology
Jennifer Mason	Supervision
Daniel P. Miller	Formal Analysis
Anna R. Poetsch	Formal Analysis
Payal Jain	Data curation, Investigation, Validation
Adam A. Kraya	Methodology
Allison P. Heath	Project administration
Mateusz P. Koptyra	Formal Analysis, Writing – original draft
Shannon Robbins	Data curation
Yiran Guo	Formal Analysis
Xiaoyan Huang	Formal Analysis
Jessica Wong	Writing – original draft
Mariarita Santi	Investigation, Validation
Angela Viaene	Investigation, Validation
Laura Scolaro	Data Curation
Angela Waanders	Supervision
Derek Hanson	Validation

<b>Author</b>	<b>Contributions</b>
Steven M. Foltz	Validation
Hongbo M. Xie	Methodology, Supervision
Siyuan Zheng	Formal analysis, Visualization, Writing - Original draft, Supervision
Cassie N. Kline	Supervision
Peter J. Madsen	Writing – review & editing
Jena V. Lilly	Conceptualization, Funding acquisition, Project administration
Philip B. Storm	Conceptualization, Funding acquisition, Resources
Adam C. Resnick	Conceptualization, Funding acquisition, Resources, Supervision
Casey S. Greene	Conceptualization, Funding acquisition, Methodology, Project administration, Software, Supervision, Writing – review & editing
Jo Lynne Rokita*	Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Software, Supervision, Writing – original draft
Jaclyn N. Taroni*	Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - Review and editing, Visualization, Supervision, Project administration
Children's Brain Tumor Network	Conceptualization
Pacific Pediatric Neurooncology Consortium	Conceptualization

## Declarations of Interest

---

CSG's spouse was an employee of Alex's Lemonade Stand Foundation, which was a sponsor of this research. JAS, CLS, CJB, SJS, and JNT are or were employees of Alex's Lemonade Stand Foundation, a sponsor of this research.

## Figure Titles and Legends

---

**Figure 1. Overview of the OpenPBTA Project.** A, The Children's Brain Tumor Network and the Pacific Pediatric Neuro-Oncology Consortium collected tumor samples from 943 patients. To date, 38 cell lines were created from tumor tissue, and over 2000 specimens were sequenced (N = 1035 RNA-Seq, N = 940 WGS, and N = 32 WXS or Targeted panel). Data was harmonized by the Kids First Data Resource Center using an Amazon S3 framework within CAVATICA. B, Stacked bar plot summary of the number of biospecimens per phase of therapy per broad histology (GNG = ganglioglioma, LGG = low-grade glioma, DIPG = diffuse intrinsic pontine glioma, DMG = diffuse midline glioma, EPN = ependymoma, HGG = high-grade glioma, ATRT = atypical teratoid rhabdoid tumor, CNS-ET = CNS embryonal tumor, MB = medulloblastoma, PNF = plexiform neurofibroma, DNET = dysembryoplastic neuroepithelial tumor, CRANIO = craniopharyngioma, EWS = Ewing sarcoma, CPP = choroid plexus papilloma). C, Overview of the open analysis and manuscript contribution model. In the analysis GitHub repository, a contributor would propose an analysis that other participants can comment on. Contributors would then implement the analysis and file a request to add their changes to the analysis repository ("pull request"). Pull requests underwent review for scientific rigor and correctness of implementation. Pull requests were additionally checked to ensure that all software dependencies

were included and the code was not sensitive to underlying data changes using container and continuous integration technologies. Finally, a contributor would file a pull request documenting their methods and results to the Manubot-powered manuscript repository. Pull requests in the manuscript repository were also subject to review. D, A potential path for an analytical pull request. Arrows indicate revisions to a pull request. Prior to review, a pull request was tested for dependency installation and whether or not the code would execute. Pull requests also required approval by organizers and/or other contributors, who checked for scientific correctness. Panel A created with [BioRender.com](https://biorender.com).

**Figure 2. Mutational landscape of PBTA tumors.** Shown are frequencies of canonical somatic gene mutations, CNVs, fusions, and TMB (top bar plot) for the top 20 genes mutated across primary tumors within the OpenPBTA dataset. A, Low-grade astrocytic tumors ( $N = 227$ ): low-grade glioma astrocytoma ( $N = 187$ ), ganglioglioma ( $N = 35$ ), subependymal giant cell astrocytoma ( $N = 2$ ), diffuse fibrillary astrocytoma ( $N = 1$ ), pilocytic astrocytoma ( $N = 1$ ), and pleomorphic xanthoastrocytoma ( $N = 1$ ); B, Embryonal tumors ( $N = 128$ ): medulloblastomas ( $N = 95$ ), atypical teratoid rhabdoid tumors ( $N = 24$ ), embryonal tumors with multilayer rosettes ( $N = 2$ ), other CNS embryonal tumors ( $N = 5$ ), ganglioneuroblastoma ( $N = 1$ ), and CNS neuroblastoma ( $N = 1$ ); C, Diffuse astrocytic and oligodendroglial tumors ( $N = 61$ ): diffuse midline gliomas ( $N = 34$ ) and non-midline high-grade gliomas ( $N = 26$ ), oligodendrogloma ( $N = 1$ ); D, Other CNS tumors ( $N = 195$ ): ependymomas ( $N = 60$ ), dysembryoplastic neuroepithelial tumors ( $N = 19$ ), meningiomas ( $N = 17$ ), schwannoma ( $N = 11$ ), neurofibroma plexiform ( $N = 7$ ). Other CNS tumors with  $N < 5$  are displayed in **Figure S3A**. Patient sex (`germline_sex_estimate`) and tumor histology (`cancer_group`) are displayed as annotations at the bottom of each plot. Only samples with mutations in the listed genes are shown. Multiple CNVs are denoted as a complex event.

**Figure 3. Mutational co-occurrence and signatures highlight key oncogenic drivers.** A, Bar plot of occurrence and co-occurrence of nonsynonymous mutations for the 50 most commonly mutated genes across all tumor types (annotated from `cancer_group` if  $N \geq 10$  or `Other` if  $N < 10$ ); B, Co-occurrence and mutual exclusivity of nonsynonymous mutations between genes; The co-occurrence score is defined as  $I(-\log_{10}(P))$  where  $P$  is defined by Fisher's exact test and  $I$  is 1 when mutations co-occur more often than expected and -1 when exclusivity is more common; C, The number of SV breaks significantly correlates with the number of CNV breaks (Adjusted  $R = 0.436$ ,  $p = 1.08e-37$ ). D, Chromothriplis frequency across pediatric brain tumors shown by `cancer_group` with  $N \geq 3$ . E, Sina plots of RefSig signature weights for signatures 1, 3, 8, 11, 18, 19, N6, MMR2, and Other across cancer groups. Box plot lines represent the first quartile, median, and third quartile.

**Figure 4. TP53 and telomerase activity** A, Receiver Operating Characteristic for *TP53* classifier run on FPKM of stranded RNA-Seq samples. B, Violin and box plots of *TP53* scores plotted by *TP53* alteration type ( $N_{\text{activated}} = 27$ ,  $N_{\text{lost}} = 140$ ,  $N_{\text{other}} = 999$ ). C, Violin and box plots of *TP53* RNA expression plotted by *TP53* activation status ( $N_{\text{activated}} = 11$ ,  $N_{\text{lost}} = 96$ ,  $N_{\text{other}} = 862$ ). D, Box plots of *TP53* and telomerase (EXTEND) scores grouped by `cancer_group`. Mutation status is highlighted in orange (hypermutant) or red (ultra-hypermutant). E, Heatmap of RefSig mutational signatures for patients who have least one tumor or cell line with a TMB  $\geq 10$  Mut/Mb. F, Forest plot depicting the prognostic effects of *TP53* and telomerase scores on overall survival, controlling for extent of tumor resection, LGG group, and HGG group. G, Forest plot depicting the effect of molecular subtype on overall survival of HGGs. For F and G, hazard ratios (HR) with 95% confidence intervals and p-values are listed. Significant p-values are denoted with black diamonds. Reference groups are denoted by grey diamonds. H, Kaplan-Meier curve of HGG tumors by molecular subtype.

**Figure 5. Transcriptomic and immune landscape of pediatric brain tumors** A, First two dimensions from UMAP of sample transcriptome data. Points are colored by `broad_histology` of the samples they represent. B, Heatmap of GSVA scores for Hallmark gene sets with significant differences, with samples ordered by `cancer_group`. C, Box plots of quanTlseq estimates of

immune cell fractions in histologies with N >= 15. D, Forest plot depicting the effect of immune cell fraction or *CD274* expression on overall survival of medulloblastomas. Hazard ratio (HR) with 95% confidence intervals and p-values are listed and significant p-values are denoted with colored diamonds. Of note, the upper CI for macrophage M1 HR was too large to plot. E, Box plot of *CD274* expression ( $\log_2$  FPKM) for medulloblastoma samples grouped by molecular subtype. Adjusted p-values are shown.

## Table Titles and Legends

---

**Table 1. Molecular subtypes generated through the OpenPBTA project.** Listed are broad tumor histologies, molecular subtypes generated, and number of specimens subtyped within the OpenPBTA project.

**Table 2. Patients with hypermutant tumors.** Listed are patients with at least one hypermutant or ultra-hypermutant tumor or cell line. Coding region TMB, phase of therapy, therapeutic interventions, cancer predispositions, and molecular subtypes are included.

## STAR METHODS

---

### RESOURCE AVAILABILITY

#### Lead contact

Requests for access to OpenPBTA 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

Raw and harmonized WGS, WXS, and RNA-Seq data derived from human samples are available within the KidsFirst Portal [57] upon access request to the CBTN (<https://cbtn.org/>) as of the date of the publication. In addition, merged summary files are openly accessible at <https://cavatica.sbggenomics.com/u/cavatica/openpbta> or via download script in the <https://github.com/AlexsLemonade/OpenPBTA-analysis> repository. Summary data are visible within PedcBioPortal at <https://pedcbioportal.kidsfirstdrc.org/study/summary?id=openpbta>. Associated DOIs are listed in the **Key Resources Table**.

All original code was developed within the following repositories and is publicly available as follows. Primary data analyses can be found at <https://github.com/d3b-center/OpenPBTA-workflows>. Downstream data analyses can be found at <https://github.com/AlexsLemonade/OpenPBTA-analysis>. Manuscript code can be found at <https://github.com/AlexsLemonade/OpenPBTA-manuscript>. Associated DOIs are listed in the **Key Resources Table**. Software versions are documented in **Table S5** as an appendix to the **Key Resources Table**.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## METHOD DETAILS

### Biospecimen Collection

The Pediatric Brain Tumor Atlas specimens are comprised of samples from Children's Brain Tumor Network (CBTN) and the Pediatric Pacific Neuro-Oncology Consortium (PNOC). The [CBTN](#) is a collaborative, multi-institutional (26 institutions worldwide) research program dedicated to the study of childhood brain tumors. [PNOC](#) is an international consortium dedicated to bringing new therapies to children and young adults with brain tumors. We also include blood and tumor biospecimens from newly-diagnosed diffuse intrinsic pontine glioma (DIPG) patients as part of the PNOC003 clinical trial [PNOC003/NCT02274987](#) [11].

The CBTN-generated cell lines were derived from either fresh tumor tissue directly obtained from surgery performed at Children's Hospital of Philadelphia (CHOP) or from prospectively collected tumor specimens stored in Recover Cell Culture Freezing medium (cat# 12648010, Gibco). We dissociated tumor tissue using enzymatic method with papain as described [58]. Briefly, we washed tissue with HBSS (cat# 14175095, Gibco), and we minced and incubated the tissue with activated papain solution (cat# LS003124, SciQuest) for up to 45 minutes. We used ovomucoid solution (cat# 542000, SciQuest) to inactivate the papain, briefly treated tissue with DNase (cat# 10104159001, Roche), passed it through the 100 $\mu$ m cell strainer (cat# 542000, Greiner Bio-One). We initiated two cell culture conditions based on the number of cells available. For cultures utilizing the fetal bovine serum (FBS), we plated a minimum density of  $3 \times 10^5$  cells/mL in DMEM/F-12 medium (cat# D8062, Sigma) supplemented with 20% FBS (cat# SH30910.03, Hyclone), 1% GlutaMAX (cat# 35050061, Gibco), Penicillin/Streptomycin-Amphotericin B Mixture (cat# 17-745E, Lonza), and 0.2% Normocin (cat# ant-nr-2, Invivogen). For serum-free media conditions, we plated cells at minimum density of  $1 \times 10^6$  cells/mL in DMEM/F12 medium supplemented with 1% GlutaMAX, 1X B-27 supplement minus vitamin A (cat# 12587-010, Gibco), 1x N-2 supplement (cat# 17502001, Gibco), 20 ng/ml epidermal growth factor (cat# PHG0311L, Gibco), 20 ng/mL basic fibroblast growth factor (cat# 100-18B, PeproTech), 2.5 $\mu$ g/mL heparin (cat# H3149, Sigma), Penicillin/Streptomycin-Amphotericin B Mixture, and 0.2% Normocin.

### Nucleic acids extraction and library preparation

#### PNOC samples

The Translational Genomic Research Institute (TGEN; Phoenix, AZ) performed DNA and RNA extractions on tumor biopsies using a DNA/RNA AllPrep Kit (Qiagen, #80204). All RNA used for library prep had a minimum RIN of seven, but no QC thresholds were implemented for the DNA. For library preparation, 500 ng of nucleic acids were used as input for RNA-Seq, WXS, and targeted DNA panel (panel) sequencing. RNA library preparation was performed using the TruSeq RNA Sample Prep Kit (Illumina, #FC-122-1001) and the exome prep was performed using KAPA Library Preparation Kit (Roche, #KK8201) using Agilent's SureSelect Human All Exon V5 backbone with custom probes. The targeted DNA panel developed by Ashion Analytics (formerly known as the GEM Cancer panel) consisted of exonic probes against 541 cancer genes. Both panel and WXS assays contained 44,000 probes across evenly spaced genomic loci used for genome-wide copy number analysis. For the panel, additional probes tiled across intronic regions of 22 known tumor suppressor genes and 22 genes involved in common cancer translocations for structural analysis. All extractions and library preparations were performed according to manufacturer's instructions.

#### CBTN samples

Blood, tissue, and cell line DNA/RNA extractions were performed at the Biorepository Core at CHOP. Briefly, 10-20 mg frozen tissue, 0.4-1ml of blood, or 2e6 cells pellet was used for extractions. Tissues were lysed using a Qiagen TissueLyser II (Qiagen) with 2x30 sec at 18Hz settings using 5 mm steel beads (cat# 69989, Qiagen). Both tissue and cell pellets processes included a CHCl<sub>3</sub> extraction and were run on the QIAcube automated platform (Qiagen) using the AllPrep DNA/RNA/miRNA Universal kit (cat# 80224, Qiagen). Blood was thawed and treated with RNase A (cat#, 19101, Qiagen); 0.4-1ml was processed using the Qiagen QIAasympathy automated platform (Qiagen) using the QIAasympathy DSP DNA Midi Kit (cat# 937255, Qiagen). DNA and RNA quantity and quality was assessed by PerkinElmer DropletQuant UV-VIS spectrophotometer (PerkinElmer) and an Agilent 4200 TapeStation (Agilent, USA) for RIN and DIN (RNA Integrity Number and DNA Integrity Number, respectively). The NantHealth Sequencing Center, BGI at CHOP, or the Genomic Clinical Core at Sidra Medical and Research Center performed library preparation and sequencing. Briefly, DNA sequencing libraries were prepared for tumor and matched-normal DNA using the KAPA HyperPrep kit (cat# 08098107702, Roche), and tumor RNA-Seq libraries were prepared using KAPA Stranded RNA-Seq with RiboErase kit (cat# 07962304001, Roche).

## Data generation

NantHealth and Sidra performed 2x150 bp WGS on paired tumor (~60X) and constitutive DNA (~30X) samples on an Illumina X/400. BGI at CHOP performed 2x100 bp WGS sequenced at 60X depth for both tumor and normal samples. NantHealth performed ribosomal-depleted whole transcriptome stranded RNA-Seq to an average depth of 200M. BGI at CHOP performed poly-A or ribosomal-depleted whole transcriptome stranded RNA-Seq to an average depth of 100M. The Translational Genomic Research Institute (TGEN; Phoenix, AZ) performed paired tumor (~200X) and constitutive whole exome sequencing (WXS) or targeted DNA panel (panel) and poly-A selected RNA-Seq (~200M reads) for PNOC tumor samples. The panel tumor sample was sequenced to 470X, and the normal panel sample was sequenced to 308X. PNOC 2x100 bp WXS and RNA-Seq libraries were sequenced on an Illumina HiSeq 2500.

## DNA WGS Alignment

We used BWA-MEM [59] to align paired-end DNA-seq reads to the version 38 patch release 12 of the *Homo sapiens* genome reference, obtained as a FASTA file from UCSC (see **Key Resources Table**). Next, we used the [Broad Institute's Best Practices](#) to process Binary Alignment/Map files (BAMs) in preparation for variant discovery. We marked duplicates using SAMBLASTER [60], and we merged and sorted BAMs using Sambamba [61]. We used the Broad's Genome Analysis Tool Kit [GATK](#) v4.0.3.0, specifically the BaseRecalibrator submodule, to process BAMs. Lastly, for normal/germline input, we ran the GATK HaplotypeCaller [62] submodule on the recalibrated BAM to generate a genomic variant call format (GVCF) file. This file is used as the basis for germline calling, described in the **SNV calling for B-allele Frequency (BAF) generation** section.

We obtained references from the [Broad Genome References on AWS](#) bucket with a general description of references at <https://s3.amazonaws.com/broad-references/broad-references-readme.html>.

## Quality Control of Sequencing Data

To confirm sample matches and remove mis-matched samples from the dataset, we performed NGSCheckMate [63] on matched tumor/normal CRAM files. Briefly, we processed CRAMs using BCFtools to filter and call 20k common single nucleotide polymorphisms (SNPs) using default parameters. We used the resulting VCFs to run NGSCheckMate following the `bcf_call.cwl` workflow found in the [D3b GitHub repository](#). Per NGSCheckMate author recommendations, we used

$\leq 0.61$  as a correlation coefficient cutoff at sequencing depths  $> 10$  to predict mis-matched samples. For RNA-Seq, we determined read strandedness by running the [infer\\_experiment.py script](#) on the first 200k mapped reads. We removed any samples whose calculated strandedness did not match strandedness information provided by the sequencing center. We required that at least 60% of RNA-Seq reads mapped to the human reference for samples to be included in analysis.

## Germline Variant Calling

### SNP calling for B-allele Frequency (BAF) generation

We performed germline haplotype calls using the [GATK Joint Genotyping Workflow](#) on individual GVCFs from the normal sample alignment workflow. Using only SNPs, we applied the [GATK generic hard filter suggestions](#) to the VCF, with an additional requirement of 10 reads minimum depth per SNP. We used the filtered VCF as input to Control-FREEC and CNVkit (below) to generate B-allele frequency (BAF) files. Note that GATK v4.0.12.0 was used for all of these steps except VariantFiltration, which used v3.8.0 because this tool was beta in GATK v4.0.12.0 and known to be unreliable for this analysis. This single-sample workflow is available in the [D3b GitHub repository](#). References can be obtained from the [Broad Genome References on AWS](#) bucket, and a general description of references can be found at <https://s3.amazonaws.com/broad-references/broad-references-readme.html>.

## Somatic Mutation Calling

### SNV and indel calling

For PBTA samples, we used four variant callers to call SNVs and indels from panel, WXS, and WGS data: Strelka2 [64], Mutect2 [65], Lancet [66], and VarDict [67]. VarDict-only calls were not retained since  $\sim 39M$  calls with low VAF were uniquely called and may be potential false positives. We used only Strelka2, Mutect2 and Lancet to analyze WXS samples from TCGA. TCGA samples were captured using various WXS target capture kits and we downloaded the BED files from the [GDC portal](#). The manufacturers provided the input interval BED files for both panel and WXS data for PBTA samples. We padded all panel and WXS BED files by 100 bp on each side for Strelka2, Mutect2, and VarDict runs and by 400 bp for the Lancet run.

For WGS calling, we utilized the non-padded BROAD Institute interval calling list [wgs\\_calling\\_regions.hg38.interval\\_list](#), comprised of the full genome minus N bases, unless otherwise noted below. We ran Strelka2 [64] using default parameters for canonical chromosomes (chr1-22, X,Y,M), as recommended by the authors, and we filtered the final Strelka2 VCF for PASS variants. We ran Mutect2 from GATK v4.1.1.0 according to Broad best practices outlined from their [Workflow Description Language \(WDL\)](#), and we filtered the final Mutect2 VCF for PASS variants. To manage memory issues, we ran VarDictJava [67] using 20 Kb interval chunks of the input BED, padded by 100 bp on each side, such that if an indel occurred in between intervals, it would be captured. Parameters and filtering followed [BCBIO standards](#) except that variants with a variant allele frequency (VAF)  $\geq 0.05$  (instead of  $\geq 0.10$ ) were retained. The 0.05 VAF increased the true positive rate for indels and decreased the false positive rate for SNVs when using VarDict in consensus calling. We filtered the final VarDict VCF for PASS variants with `TYPE=StronglySomatic`. We ran Lancet using default parameters, except for those noted below. For input intervals to Lancet WGS, we created a reference BED from only the UTR, exome, and start/stop codon features of the GENCODE 31 reference, augmented as recommended with PASS variant calls from Strelka2 and Mutect2 [68]. We then padded these intervals by 300 bp on each side during Lancet variant calling. Per recommendations by the New York Genome Center [68] for WGS samples, we augmented the Lancet input intervals described above with PASS variant calls from Strelka2 and Mutect2 as validation.

### VCF annotation and MAF creation

We normalized INDELS with `bcftools norm` on all PASS VCFs using the `kfdrc_annot_vcf_sub_wf.cwl` subworkflow, release v3 (See **Table S5**). The ENSEMBL Variant Effect Predictor [69], reference release 93, was used to annotate variants and bcftools was used to add population allele frequency (AF) from gnomAD [70]. We annotated SNV and INDEL hotspots from v2 of Memorial Sloan Kettering Cancer Center's (MSKCC) database (See **Key Resources Table**) as well as the TERT promoter mutations C228T and C250T [71]. We annotated SNVs by matching amino acid position (`Protein_position` column in MAF file) with SNVs in the MSKCC database, we matched splice sites to `HGVSp_Short` values in the MSKCC database, and we matched INDELS based on amino acid present within the range of INDEL hotspots values in the MSKCC database. We removed non-hotspot annotated variants with a normal depth less than or equal to 7 and/or gnomAD allele frequency (AF) greater than 0.001 as potential germline variants. We matched TERT promoter mutations using hg38 coordinates as indicated in ref. [71]: C228T occurs at 5:1295113 is annotated as existing variant `s1242535815`, `COSM1716563`, or `COSM1716558`, and is 66 bp away from the TSS; C250T occurs at Chr5:1295135, is annotated as existing variant `COSM1716559`, and is 88 bp away from the TSS. We retained variants annotated as PASS or HotSpotAllele=1 in the final set, and we created MAFs using MSKCC's `vcf2maf` tool.

## Gather SNV and INDEL Hotspots

We retained all variant calls from Strelka2, Mutect2, or Lancet that overlapped with an SNV or INDEL hotspot in a hotspot-specific MAF file, which we then used for select analyses as described below.

## Consensus SNV Calling

Our SNV calling process led to separate sets of predicted mutations for each caller. We considered mutations to describe the same change if they were identical for the following MAF fields: `Chromosome`, `Start_Position`, `Reference_Allele`, `Allele`, and `Tumor_Sample_Barcode`. Strelka2 does not call multinucleotide variants (MNV), but instead calls each component SNV as a separate mutation, so we separated MNV calls from Mutect2 and Lancet into consecutive SNVs before comparing them to Strelka2 calls. We examined VAFs produced by each caller and compared their overlap with each other (**Figure S2**). VarDict calls included many variants that were not identified by other callers (**Figure S2C**), while the other callers produced results that were relatively consistent with one another. Many of these VarDict-specific calls were variants with low allele frequency (**Figure S2B**). We therefore derived consensus mutation calls as those shared among the other three callers (Strelka2, Mutect2, and Lancet), and we did not further consider VarDict calls due to concerns it called a large number of false positives. This decision had minimal impact on results because VarDict also identified nearly every mutation that the other three callers identified, in addition to many unique mutations.

## Somatic Copy Number Variant Calling (WGS samples only)

We used Control-FREEC [72,73] and CNVkit [74] for copy number variant calls. For both algorithms, the `germline_sex_estimate` (described below) was used as input for sample sex and germline variant calls (above) were used as input for BAF estimation. Control-FREEC was run on human genome reference hg38 using the optional parameters of a 0.05 coefficient of variation, ploidy choice of 2-4, and BAF adjustment for tumor-normal pairs. Theta2 [75] used VarDict germline and somatic calls, filtered on PASS and strongly somatic, to infer tumor purity. Theta2 purity was added as an optional parameter to CNVkit to adjust copy number calls. CNVkit was run on human genome reference hg38 using the optional parameters of Theta2 purity and BAF adjustment for tumor-normal pairs. We used GISTIC [76] on the CNVkit and the consensus CNV segmentation files to generate gene-level copy number abundance (Log R Ratio) as well as chromosomal arm copy number alterations using the parameters specified in the (`run-gistic` analysis module in the OpenPBTA Analysis repository).

## Consensus CNV Calling

For each caller and sample, we called CNVs based on consensus among Control-FREEC [72,73], CNVkit [74], and Manta [77]. We specifically included CNVs called significant by Control-FREEC ( $p\text{-value} < 0.01$ ) and Manta calls that passed all filters in consensus calling. We removed sample and consensus caller files with more than 2,500 CNVs because we expected these to be noisy and derive poor quality samples based on cutoffs used in GISTIC [76]. For each sample, we included the regions in the final consensus set: 1) regions with reciprocal overlap of 50% or more between at least two of the callers; 2) smaller CNV regions in which more than 90% of regions are covered by another caller. We did not include any copy number alteration called by a single algorithm in the consensus file. We defined copy number as `NA` for any regions that had a neutral call for the samples included in the consensus file. We merged CNV regions within 10,000 bp of each other with the same direction of gain or loss into single region. We filtered out any CNVs that overlapped 50% or more with immunoglobulin, telomeric, centromeric, segment duplicated regions, or that were shorter than 3000 bp.

## Somatic Structural Variant Calling (WGS samples only)

We used Manta [77] for structural variant (SV) calls, and we limited to regions used in Strelka2. The hg38 reference for SV calling used was limited to canonical chromosome regions. The somatic DNA workflows for SNV, indel, copy number, and SV calling can be found in the [OpenPBTA workflows Github repository](#). We used [AnnotSV v2.1](#) [78] to annotate Manta output. This associated workflow is available in the [workflows GitHub repository](#).

## Gene Expression

### Abundance Estimation

We used STAR [79] to align paired-end RNA-seq reads, and we used the associated alignment for all subsequent RNA analysis. We used Ensembl GENCODE 27 “Comprehensive gene annotation” (see [Key Resources Table](#)) as a reference. We used RSEM [80] for both FPKM and TPM transcript- and gene-level quantification.

### Gene Expression Matrices with Unique HUGO Symbols

To enable downstream analyses, we next identified gene symbols that map to multiple Ensembl gene identifiers (in GENCODE v27, 212 gene symbols map to 1866 Ensembl gene identifiers), known as multi-mapped gene symbols, and ensured unique mappings ( `collapse-rnaseq` analysis module in the OpenPBTA Analysis repository). To this end, we first removed genes with no expression from the RSEM abundance data by requiring an  $\text{FPKM} > 0$  in at least 1 sample across the PBTA cohort. We computed the mean FPKM across all samples per gene. For each multi-mapped gene symbol, we chose the Ensembl identifier corresponding to the maximum mean FPKM, using the assumption that the gene identifier with the highest expression best represented the expression of the gene. After collapsing gene identifiers, 46,400 uniquely-expressed genes remained in the poly-A dataset, and 53,011 uniquely-expressed genes remained in the stranded dataset.

## Gene fusion detection

We set up Arriba [81] and STAR-Fusion [82] fusion detection tools using CWL on CAVATICA. For both of these tools, we used aligned BAM and chimeric SAM files from STAR as inputs and `GRCh38_gencode_v27` GTF for gene annotation. We ran STAR-Fusion with default parameters and annotated all fusion calls with the `GRCh38_v27_CTAT_lib_Feb092018.plug-n-play.tar.gz` file from the STAR-Fusion release. For Arriba, we used a blacklist file `blacklist_hg38_GRCh38_2018-`

`11-04.tsv.gz` from the Arriba release to remove recurrent fusion artifacts and transcripts present in healthy tissue. We provided Arriba with strandedness information for stranded samples, or we set it to auto-detection for poly-A samples. We used [FusionAnnotator](#) on Arriba fusion calls to harmonize annotations with those of STAR-Fusion. The RNA expression and fusion workflows can be found in the [D3b GitHub repository](#). The FusionAnnotator workflow can also be found in the [D3b GitHub repository](#).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Recurrently mutated genes and co-occurrence of gene mutations (`interaction-plots` analysis module)

Using the consensus SNV calls, we identified genes that were recurrently mutated in the OpenPBTA cohort, including nonsynonymous mutations with a VAF > 5% among the set of independent samples. We used the ENSEMBL Variant Effect Predictor [69] annotations, including “High” and “Moderate” consequence types as defined in the R package Maftools [83], to determine the set of nonsynonymous mutations. For each gene, we then tallied the number of samples that had at least one nonsynonymous mutation.

For genes that contained nonsynonymous mutations in multiple samples, we calculated pairwise mutation co-occurrence scores. This score was defined as  $I(-\log_{10}(P))$  where  $I$  is 1 when the odds ratio is > 1 (indicating co-occurrence), and -1 when the odds ratio is < 1 (indicating mutual exclusivity), with  $P$  defined by Fisher’s Exact Test.

### Focal Copy Number Calling (`focal-cn-file-preparation` analysis module)

We added the ploidy inferred via Control-FREEC to the consensus CNV segmentation file and used the ploidy and copy number values to define gain and loss values broadly at the chromosome level. We used `bedtools coverage` [84] to add cytoband status using the UCSC cytoband file [85] (See **Key Resources Table**). The output status call fractions, which are values of the loss, gain, and callable fractions of each cytoband region, were used to define dominant status at the cytoband-level. We calculated the weighted means of each status call fraction using band length. We used the weighted means to define the dominant status at the chromosome arm-level.

A status was considered dominant if more than half of the region was callable and the status call fraction was greater than 0.9 for that region. We adopted this 0.9 threshold to ensure that the dominant status fraction call was greater than the remaining status fraction calls in a region.

We aimed to define focal copy number units to avoid calling adjacent genes in the same cytoband or arm as copy number losses or gains where it would be more appropriate to call the broader region a loss or gain. To determine the most focal units, we first considered the dominant status calls at the chromosome arm-level. If the chromosome arm dominant status was callable but not clearly defined as a gain or loss, we instead included the cytoband-level status call. Similarly, if a cytoband dominant status call was callable but not clearly defined as a gain or loss, we instead included gene-level status call. To obtain the gene-level data, we used the `mergeByOverlaps` function from the IRanges package in R [86] to find overlaps between the segments in the consensus CNV file and the exons in the GENCODE v27 annotation file (See **Key Resources Table**) . If the copy number value was 0, we set the status to “deep deletion”. For autosomes only, we set the status to “amplification” when the copy number value was greater than two times the ploidy value.

### Chromothripsis Analysis (WGS samples only; `chromothripsis` analysis module)

Considering only chromosomes 1-22 and X, we identified candidate chromothripsis regions in the set of independent tumor WGS samples with ShatterSeek [87], using Manta SV calls that passed all filters and consensus CNV calls. We modified the consensus CNV data to fit ShatterSeek input requirements as follows: CNV-neutral or excluded regions (both annotated as NA in the consensus data) were filled in with the respective sample's ploidy value from Control-FREEC, and consecutive segments with the same copy number value were merged. We classified candidate chromothripsis regions as high- or low-confidence using the statistical criteria described by the ShatterSeek authors.

## Immune Profiling and Deconvolution (`immune-deconv` analysis module)

We used the R package `immunedeconv` [88] with the method `quantiSeq` [89] to deconvolute various immune cell types across tumors from the PBTA cohort in the stranded and poly-A collapsed FPKM RNA-seq datasets (`immune-deconv` analysis module). The `quantiSeq` deconvolution method directly estimates absolute fractions of 10 immune cell types that represent inferred proportions of the cell types in the mixture. Therefore, we utilized `quantiSeq` for inter-sample, intra-sample, and inter-histology score comparisons.

## Gene Set Variation Analysis (`gene-set-enrichment-analysis` analysis module)

We performed Gene Set Variation Analysis (GSVA) [90] on collapsed, log2-transformed RSEM FPKM data using the GSVA Bioconductor package [91]. We specified the parameter `mx.diff=TRUE` to obtain Gaussian-distributed scores for each of the MSigDB hallmark gene sets [92]. We compared GSVA scores among histology groups using ANOVA and subsequent Tukey tests; p-values were Bonferroni-corrected for multiple hypothesis testing.

## Transcriptomic Dimension Reduction (`transcriptomic-dimension-reduction` analysis module)

We applied Uniform Manifold Approximation and Projection (UMAP) [93] to log2-transformed FPKM data using the `umap` R package (See **Key Resources Table**). We set the number of neighbors to 15.

## Fusion prioritization (`fusion_filtering` analysis module)

We performed artifact filtering and additional annotation on fusion calls to prioritize putative oncogenic fusions. Briefly, we considered all in-frame and frameshift fusion calls with at least one junction read and at least one gene partner expressed (TPM > 1) to be true calls. If a fusion call had a large number of spanning fragment reads compared to junction reads (spanning fragment minus junction read greater than ten), we removed these calls as potential false positives. We prioritized a union of fusion calls as true calls if the fused genes were detected by both callers, the same fusion was recurrent within a broad histology grouping (> 2 samples), or the fusion was specific to the given broad histology. If either 5' or 3' genes fused to more than five different genes within a sample, we removed these calls as potential false positives. We annotated putative driver fusions and prioritized fusions based on partners containing known [kinases](#), [oncogenes](#), [tumor suppressors](#), curated transcription factors [94], [COSMIC genes](#), and/or known [TCGA fusions](#) from curated references. Based on pediatric cancer literature review, we added *MYBL1* [95], *SNCAIP* [96], *FOXR2* [97], *TTYH1* [98], and *TERT* [99,100,101,102] to the oncogene list, and we added *BCOR* [97] and *QKI* [103] to the tumor suppressor gene list.

## Oncoprint figure generation (`oncoprint-landscape` analysis module)

We used Maftools [83] to generate oncoprints depicting the frequencies of canonical somatic gene mutations, CNVs, and fusions for the top 20 genes mutated across primary tumors within broad

histologies of the OpenPBTA dataset. We collated canonical genes from the literature for low-grade astrocytic tumors [20], embryonal tumors [21,23,24,104,105], diffuse astrocytic and oligodendroglial tumors [11,17,25,26], and other tumors: ependymal tumors, craniopharyngiomas, neuronal-glial mixed tumors, histiocytic tumors, chordoma, meningioma, and choroid plexus tumors [???,106,107,108,109,110,111,112,113,114].

## Mutational Signatures (mutational-signatures analysis module)

We obtained weights (i.e., exposures) for signature sets using the `deconstructSigs` R package function `whichSignatures()` [115] from consensus SNVs with the `BSgenome.Hsapiens.UCSC.hg38` annotations (see **Key Resources Table**). Specifically, we estimated signature weights across samples for eight signatures previously identified in the Signal reference set of signatures ("RefSig") as associated with adult central nervous system (CNS) tumors [30]. These eight RefSig signatures are 1, 3, 8, 11, 18, 19, N6, and MMR2. Weights for signatures fall in the range zero to one inclusive. `deconstructSigs` estimates the weights for each signature across samples and allows for a proportion of unassigned weights referred to as "Other" in the text. These results do not include signatures with small contributions; `deconstructSigs` drops signature weights that are less than 6% [115].

## Tumor Mutation Burden (snv-callers analysis module)

We consider tumor mutation burden (TMB) to be the number of consensus SNVs per effectively surveyed base of the genome. We considered base pairs to be effectively surveyed if they were in the intersection of the genomic ranges considered by the callers used to generate the consensus and where appropriate, regions of interest, such as coding sequences. We calculated TMB as,

$$\text{TMB} = \frac{\text{\# of coding sequence SNVs}}{\text{Size in Mb of \{ effectively surveyed\} genome}}$$

We used the total number coding sequence consensus SNVs for the numerator and the size of the intersection of the regions considered by Strelka2 and Mutect2 with coding regions (CDS from GENCODE v27 annotation, see **Key Resources Table**) as the denominator.

## Clinical Data Harmonization

### WHO Classification of Disease Types

**Table S1** contains a README, along with sample technical, clinical, and additional metadata used for this study.

### Molecular Subtyping

We performed molecular subtyping on tumors in the OpenPBTA to the extent possible. The `molecular_subtype` field in `pbta-histologies.tsv` contains molecular subtypes for tumor types selected from `pathology_diagnosis` and `pathology_free_text_diagnosis` fields as described below, following World Health Organization 2016 classification criteria [16].

Medulloblastoma (MB) subtypes SHH, WNT, Group 3, and Group 4 were predicted using the consensus of two RNA expression classifiers: Medulloblastoma Classifier [45] and MM2S Classifier [44] on the RSEM FPKM data (`molecular-subtyping-MB` analysis module).

High-grade glioma (HGG) subtypes were derived (molecular-subtyping-HGG analysis module) using the following criteria:

1. If any sample contained an *H3F3A* p.K28M, *HIST1H3B* p.K28M, *HIST1H3C* p.K28M, or *HIST2H3C* p.K28M mutation and no *BRAF* p.V600E mutation, it was subtyped as DMG, H3K28.
2. If any sample contained an *HIST1H3B* p.K28M, *HIST1H3C* p.K28M, or *HIST2H3C* p.K28M mutation and a *BRAF* p.V600E mutation, it was subtyped as DMG, H3 K28, BRAF V600E.
3. If any sample contained an *H3F3A* p.G35V or p.G35R mutation, it was subtyped as HGG, H3 G35.
4. If any high-grade glioma sample contained an *IDH1* p.R132 mutation, it was subtyped as HGG, IDH.
5. If a sample was initially classified as HGG, had no defining histone mutations, and a *BRAF* p.V600E mutation, it was subtyped as BRAF V600E.
6. All other high-grade glioma samples that did not meet any of these criteria were subtyped as HGG, H3 wildtype.

Embryonal tumors were included in non-MB and non-ATRT embryonal tumor subtyping (molecular-subtyping-embryonal analysis module) if they met any of the following criteria: 1. A *TTYH1* (5' partner) fusion was detected. 2. A *MN1* (5' partner) fusion was detected, with the exception of *MN1--PATZ1* since it is an entity separate of CNS HGNET-MN1 tumors [116]. 3. Pathology diagnoses included "Supratentorial or Spinal Cord PNET" or "Embryonal Tumor with Multilayered Rosettes". 4. A pathology diagnosis of "Neuroblastoma", where the tumor was not indicated to be peripheral or metastatic and was located in the CNS. 5. Any sample with "embryonal tumor with multilayer rosettes, ros (who grade iv)", "embryonal tumor, nos, congenital type", "ependymoblastoma" or "medulloepithelioma" in pathology free text.

Non-MB and non-ATRT embryonal tumors identified with the above criteria were further subtyped (molecular-subtyping-embryonal analysis module) using the criteria below [117,118,119,120].

1. Any RNA-seq biospecimen with *LIN28A* overexpression, plus a *TTYH1* fusion (5' partner) with a gene adjacent or within the C19MC miRNA cluster and/or copy number amplification of the C19MC region was subtyped as ETMR, C19MC-altered (Embryonal tumor with multilayer rosettes, chromosome 19 miRNA cluster altered) [121; 10.1038/ng.2849].
2. Any RNA-seq biospecimen with *LIN28A* overexpression, a *TTYH1* fusion (5' partner) with a gene adjacent or within the C19MC miRNA cluster but no evidence of copy number amplification of the C19MC region was subtyped as ETMR, NOS (Embryonal tumor with multilayer rosettes, not otherwise specified) [98,121].
3. Any RNA-seq biospecimen with a fusion having a 5' *MN1* and 3' *BEND2* or *CXXC5* partner were subtyped as CNS HGNET-MN1 [Central nervous system (CNS) high-grade neuroepithelial tumor with *MN1* alteration].
4. Non-MB and non-ATRT embryonal tumors with internal tandem duplication (as defined in [122]) of *BCOR* were subtyped as CNS HGNET-BCOR (CNS high-grade neuroepithelial tumor with *BCOR* alteration).
5. Non-MB and non-ATRT embryonal tumors with over-expression and/or gene fusions in *FOXR2* were subtyped as CNS NB-FOXR2 (CNS neuroblastoma with *FOXR2* activation).
6. Non-MB and non-ATRT embryonal tumors with *CIC-NUTM1* or other *CIC* fusions, were subtyped as CNS EFT-CIC (CNS Ewing sarcoma family tumor with *CIC* alteration) [97]
7. Non-MB and non-ATRT embryonal tumors that did not fit any of the above categories were subtyped as CNS Embryonal, NOS (CNS Embryonal tumor, not otherwise specified).

Neurocytoma subtypes central neurocytoma (CNC) and extraventricular neurocytoma (EVN) were assigned (molecular-subtyping-neurocytoma analysis module) based on the primary site of the

tumor [123]. If the tumor's primary site was "ventricles," we assigned the subtype as CNC; otherwise, we assigned the subtype as EVN.

Craniopharyngiomas (CRANIO) were subtyped (molecular-subtyping-CRANIO analysis module) into adamantinomatous (CRANIO, ADAM), papillary (CRANIO, PAP) or undetermined (CRANIO, To be classified) based on the following criteria [124,125]: 1. Craniopharyngiomas from patients over 40 years old with a BRAF p.V600E mutation were subtyped as CRANIO, PAP. 2. Craniopharyngiomas from patients younger than 40 years old with mutations in exon 3 of CTNNB1 were subtyped as CRANIO, ADAM. 3. Craniopharyngiomas that did not fall into the above two categories were subtyped as CRANIO, To be classified.

A molecular subtype of EWS was assigned to any tumor with a EWSR1 fusion or with a pathology\_diagnosis of Ewings Sarcoma (molecular-subtyping-EWS analysis module).

Low-grade gliomas (LGG) or glialneuronal tumors (GNT) were subtyped (molecular-subtyping-LGAT analysis module), based on SNV, fusion and CNV status based on 18, and as described below. 1. If a sample contained a NF1 somatic mutation, either nonsense or missense, it was subtyped as LGG, NF1-somatic. 2. If a sample contained NF1 germline mutation, as indicated by a patient having the neurofibromatosis cancer predisposition, it was subtyped as LGG, NF1-germline. 3. If a sample contained the IDH p.R132 mutation, it was subtyped as LGG, IDH. 4. If a sample contained a histone p.K28M mutation in either H3F3A, H3F3B, HIST1H3B, HIST1H3C, or HIST2H3C, or if it contained a p.G35R or p.G35V mutation in H3F3A, it was subtyped as LGG, H3. 5. If a sample contained BRAF p.V600E or any other non-canonical BRAF mutations in the kinase (PK\_Tyr\_Ser-Thr) domain PF07714 (see **Key Resources Table**), it was subtyped as LGG, BRAF V600E. 6. If a sample contained KIAA1549--BRAF fusion, it was subtyped as LGG, KIAA1549-BRAF. 7. If a sample contained SNV or indel in either KRAS, NRAS, HRAS, MAP2K1, MAP2K2, MAP2K1, ARAF, RAF1, or non-kinase domain of BRAF, or if it contained RAF1 fusion, or BRAF fusion that was not KIAA1549--BRAF, it was subtyped as LGG, other MAPK. 8. If a sample contained SNV in either MET, KIT or PDGFRA, or if it contained fusion in ALK, ROS1, NTRK1, NTRK2, NTRK3 or PDGFRA, it was subtyped as LGG, RTK. 9. If a sample contained FGFR1 p.N546K, p.K656E, p.N577, or p. K687 hotspot mutations, or tyrosine kinase domain tandem duplication (See **Key Resources Table**), or FGFR1 or FGFR2 fusions, it was subtyped as LGG, FGFR. 10. If a sample contained MYB or MYBL1 fusion, it was subtyped as LGG, MYB/MYBL1. 11. If a sample contained focal CDKN2A and/or CDKN2B deletion, it was subtyped as LGG, CDKN2A/B.

For LGG tumors that did not have any of the above molecular alterations, if both RNA and DNA samples were available, it was subtyped as LGG, wildtype. Otherwise, if either RNA or DNA sample was unavailable, it was subtyped as LGG, To be classified.

If pathology diagnosis was Subependymal Giant Cell Astrocytoma (SEGA), the LGG portion of molecular subtype was recoded to SEGA.

Lastly, for all LGG- and GNT- subtyped samples, if the tumors were glialneuronal in origin, based on pathology\_free\_text\_diagnosis entries of desmoplastic infantile, desmoplastic infantile ganglioglioma, desmoplastic infantile astrocytoma or glioneuronal, each was recoded as follows: If pathology diagnosis is Low-grade glioma/astrocytoma (WHO grade I/II) or Ganglioglioma, the LGG portion of the molecular subtype was recoded to GNT.

Ependymomas (EPN) were subtyped (molecular-subtyping-EPN analysis module) into EPN, ST RELA, EPN, ST YAP1, EPN, PF A and EPN, PF B based on evidence for these molecular subgroups as described in Pajtler et al. [107]. Briefly, fusion, CNV and gene expression data were used to subtype EPN as follows:

1. Any tumor with fusions containing *RELA* as fusion partner, e.g., *C11orf95--RELA*, *LTBP3--RELA*, was subtyped as *EPN, ST RELA*.
2. Any tumor with fusions containing *YAP1* as fusion partner, such as *C11orf95--YAP1*, *YAP1--MAMLD1* and *YAP1--FAM118B*, was subtyped as *EPN, ST YAP1*.
3. Any tumor with the following molecular characterization would be subtyped as *EPN, PF A*:
  - *CXorf67* expression z-score of over 3
  - *TKTL1* expression z-score of over 3 and 1q gain
4. Any tumor with the following molecular characterization would be subtyped as *EPN, PF B*:
  - *GPBP17* expression z-score of over 3 and loss of 6q or 6p
  - *IFT46* expression z-score of over 3 and loss of 6q or 6p

Any tumor with the above molecular characteristics would be exclusively subtyped to the designated group.

For all other remaining EPN tumors without above molecular characteristics, they would be subtyped to *EPN, ST RELA* and *EPN, ST YAP1* in a non-exclusive way (e.g., a tumor could have both *EPN, ST RELA* and *EPN, ST YAP1* subtypes) if any of the following alterations were present.

1. Any tumor with the following alterations was assigned *EPN, ST RELA*:
  - *PTEN--TAS2R1* fusion
  - chromosome 9 arm (9p or 9q) loss
  - *RELA* expression z-score of over 3
  - *L1CAM* expression z-score of over 3
2. Any tumor with the following alterations was assigned *EPN, ST YAP1*:
  - *C11orf95--MAML2* fusion
  - chromosome 11 short arm (11p) loss
  - chromosome 11 long arm (11q) gain
  - *ARL4D* expression z-score of over 3
  - *CLDN1* expression z-score of over 3

After all relevant tumor samples were subtyped by the above molecular subtyping modules, the results from these modules, along with other clinical information (such as pathology diagnosis free text), were compiled in the *molecular-subtyping-pathology* module and integrated into the OpenPBTA data in the *molecular-subtyping-integrate* module.

## **TP53 Alteration Annotation (*tp53\_nf1\_score* analysis module)**

We annotated *TP53* altered HGG samples as either *TP53 lost* or *TP53 activated* and integrated this within the molecular subtype. To this end, we applied a *TP53* inactivation classifier originally trained on TCGA pan-cancer data [31] to the matched RNA expression data for each sample. Along with the *TP53* classifier scores, we collectively used consensus SNV and CNV, SV, and reference databases that list *TP53* hotspot mutations [126,127] and functional domains [128] to determine *TP53* alteration status for each sample. We adopted the following rules for calling either *TP53 lost* or *TP53 activated*:

1. If a sample had either of the two well-characterized *TP53* gain-of-function mutations, p.R273C or p.R248W [33], we assigned *TP53 activated* status.

2. Samples were annotated as `TP53 lost` if they contained i) a *TP53* hotspot mutation as defined by IARC *TP53* database or the MSKCC cancer hotspots database [126,127] (see also, **Key Resources Table**), ii) two *TP53* alterations, including SNV, CNV or SV, indicative of probable bi-allelic alterations; iii) one *TP53* somatic alteration, including SNV, CNV, or SV or a germline *TP53* mutation indicated by the diagnosis of Li-Fraumeni syndrome [129], or iv) one germline *TP53* mutation indicated by Li-Fraumeni syndrome and the *TP53* classifier score for matched RNA-Seq was greater than 0.5.

## Prediction of participants' genetic sex

Participant metadata included a reported gender. We used WGS germline data, in concert with the reported gender, to predict participant genetic sex so that we could identify sexually dimorphic outcomes. This analysis may also indicate samples that may have been contaminated. We used the `idxstats` utility from SAMtools [130] to calculate read lengths, the number of mapped reads, and the corresponding chromosomal location for reads to the X and Y chromosomes. We used the fraction of total normalized X and Y chromosome reads that were attributed to the Y chromosome as a summary statistic. We manually reviewed this statistic in the context of reported gender and determined that a threshold of less than 0.2 clearly delineated female samples. We marked fractions greater than 0.4 as predicted males, and we marked samples with values in the inclusive range 0.2-0.4 as unknown. We performed this analysis through [CWL](#) on CAVATICA. We added resulting calls to the histologies file under the column header `germline_sex_estimate`.

## Selection of independent samples (`independent-samples` analysis module)

Certain analyses required that we select only a single representative specimen for each individual. In these cases, we identified a single specimen by prioritizing primary tumors and those with whole-genome sequencing available. If this filtering still resulted in multiple specimens, we randomly selected a single specimen from the remaining set.

## Quantification of Telomerase Activity using Gene Expression Data (`telomerase-activity-prediction` analysis module)

We predicted telomerase activity of tumor samples using the recently developed EXTEND method [131]. Briefly, EXTEND estimates telomerase activity based on the expression of a 13-gene signature. We derived this signature by comparing telomerase-positive tumors and tumors with activated alternative lengthening of telomeres pathway, a group presumably negative of telomerase activity.

## Survival models (`survival-analysis` analysis module)

We calculated overall survival (OS) as days since initial diagnosis and performed several survival analyses on the OpenPBTA cohort using the `survival R package`. We performed survival analysis for patients by HGG subtype using the Kaplan-Meier estimator [132] and a log-rank test (Mantel-Cox test) [133] on the different HGG subtypes. Next, we used multivariate cox (proportional hazards) regression analysis [134] to model the following: a) `tp53 scores + telomerase scores + extent of tumor resection + LGG group + HGG group`, in which `tp53 scores` and `telomerase scores` are numeric, `extent of tumor resection` is categorical, and `LGG group` and `HGG group` are binary variables indicating whether the sample is in either broad histology grouping, b) `tp53 scores + telomerase scores + extent of tumor resection` for each `cancer_group` with an  $N \geq 3$  deceased patients (DIPG, DMG, HGG, MB, and EPN), and c) `quantiseq cell type fractions + CD274 expression + extent of tumor resection` for each

cancer\_group with an N>=3 deceased patients (DIPG, DMG, HGG, MB, and EPN), in which quantiseq cell type fractions and CD274 expression are numeric.

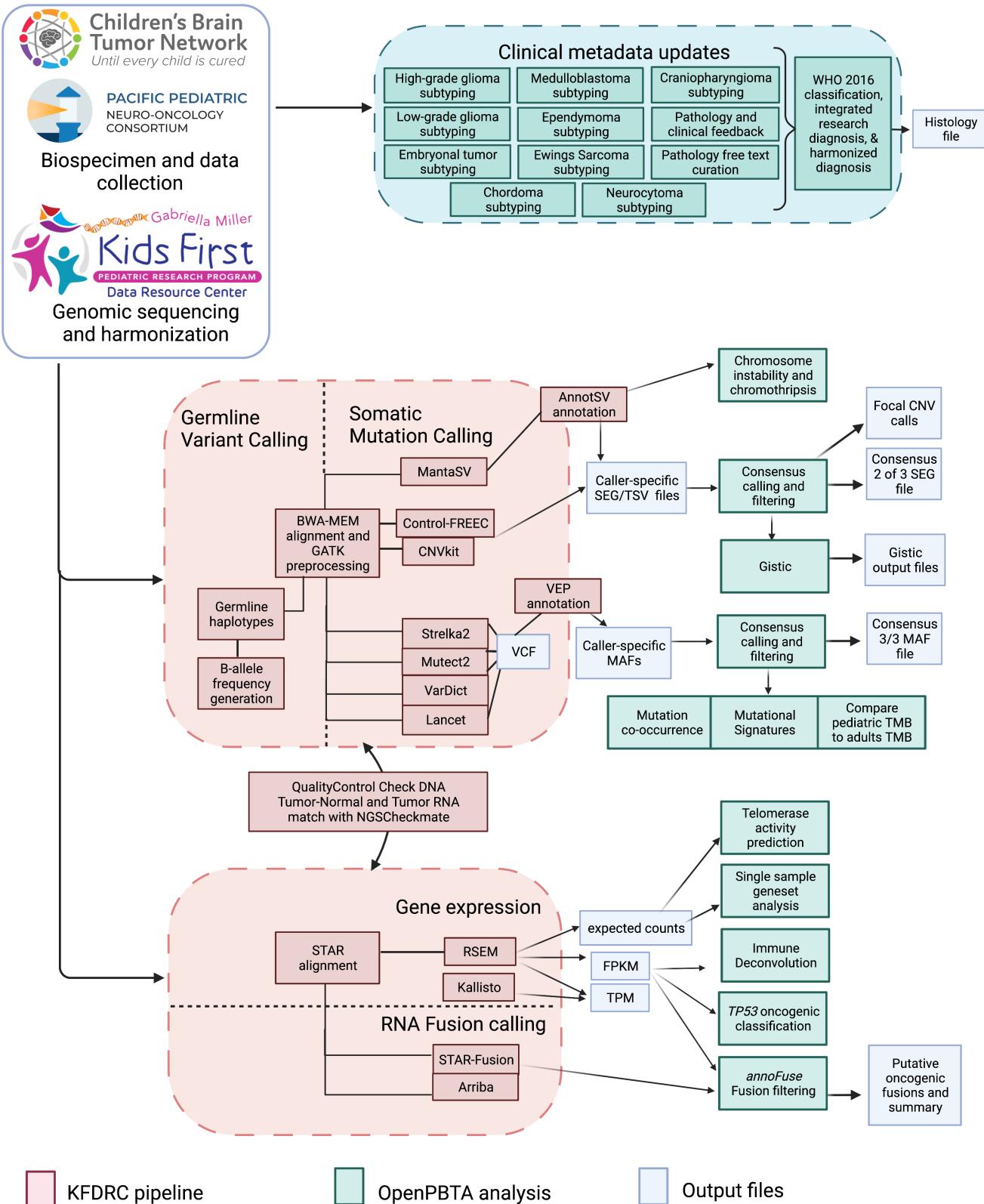
## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
Recover Cell Culture Freezing media	Gibco	12648010
Hank's Balanced Salt Solution (HBSS)	Gibco	14175095
Papain	SciQuest	LS003124
Ovomucoid	SciQuest	542000
DNase	Roche	10104159001
100µm cell strainer	Greiner Bio-One	542000
DMEM/F-12 medium	Sigma	D8062
Fetal Bovine Serum (FBS)	Hyclone	SH30910.03
GlutaMAX	Gibco	35050061
Penicillin/Streptomycin-Amphotericin B	Lonza	17-745E
Normocin	Invivogen	ant-nr-2
B-27 supplement minus vitamin A	Gibco	12587-010
N-2 supplement	Gibco	17502001
Epidermal growth factor	Gibco	PHG0311L
Basic fibroblast growth factor	PeproTech	100-18B
Heparin	Sigma	H3149

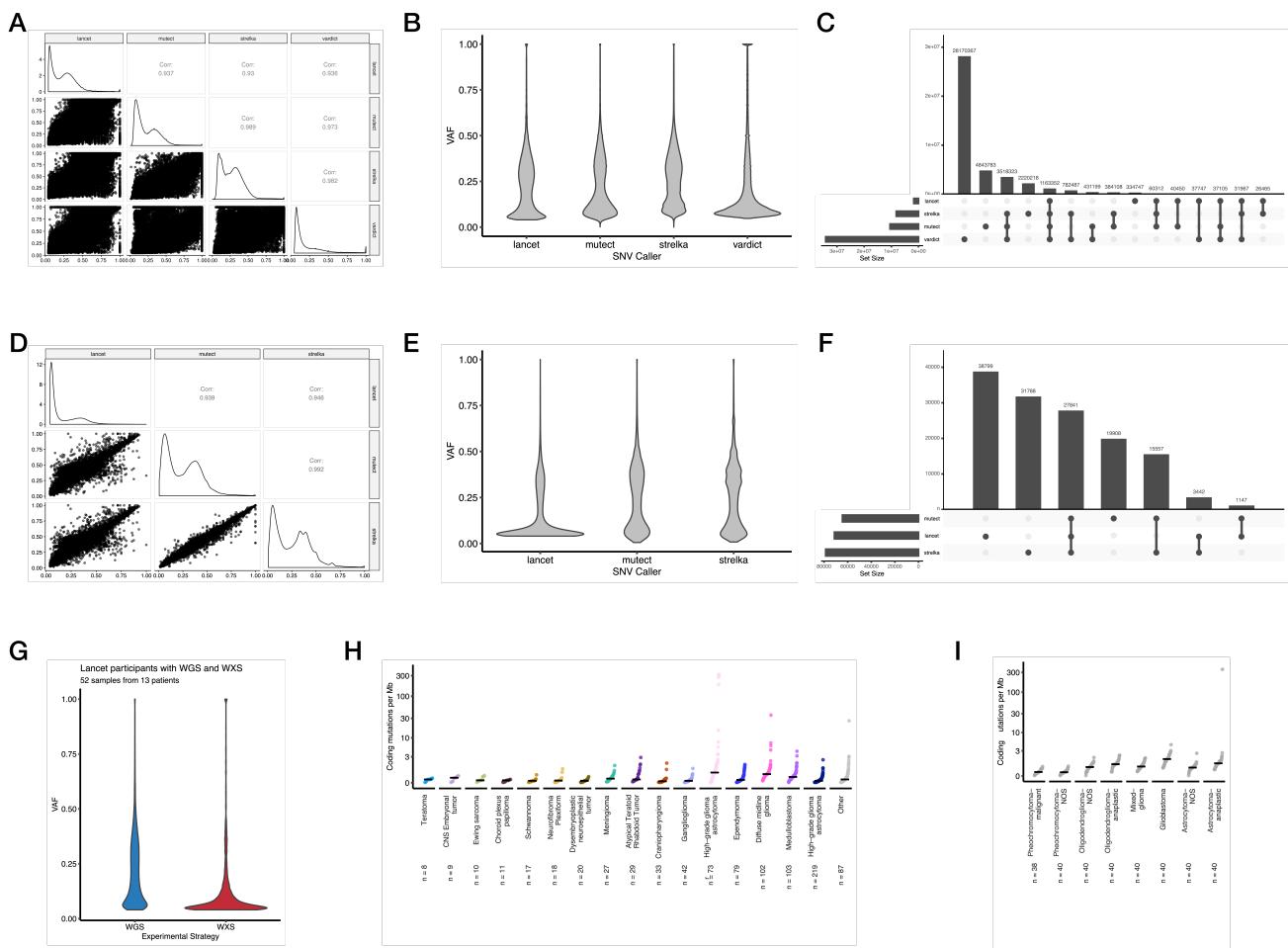
REAGENT or RESOURCE	SOURCE	IDENTIFIER
DNA/RNA AllPrep Kit	Qiagen	80204
TruSeq RNA Sample Prep Kit	Illumina	FC-122-1001
KAPA Library Preparation Kit	Roche	KK8201
AllPrep DNA/RNA/miRNA Universal kit	Qiagen	80224
RNase A	Qiagen	19101
QIAAsymphony DSP DNA Midi Kit	Qiagen	937255
KAPA HyperPrep kit	Roche	08098107702
RiboErase kit	Roche	07962304001
Deposited data		
Raw and harmonized WGS, WXS, Panel, RNA-Seq	KidsFirst Data Resource Center, this project	[57]
Merged summary files	this project	<a href="https://cavatica.sbggenomics.com/u/cavatica/openpbta">https://cavatica.sbggenomics.com/u/cavatica/openpbta</a>
Merged summary files and downstream analyses	this project	<a href="https://github.com/AlexsLemonade/OpenPBTA-analysis/">https://github.com/AlexsLemonade/OpenPBTA-analysis/</a>
Processed data	this project	<a href="https://pedcbioportal.kidsfirstdrc.org/study/summary?id=openpbta">https://pedcbioportal.kidsfirstdrc.org/study/summary?id=openpbta</a>
Experimental models: Cell lines		
CBTN pediatric brain tumor-derived cell lines	[10]	See <b>Table S1</b> for identifiers

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Data processing and analysis software	Multiple	See <b>Table S5</b> for identifiers
OpenPBTA workflows repository	this project	[135]
OpenPBTA analysis repository	this project	
OpenPBTA manuscript repository	this project	
Other		
TCGA WXS dataset	National Institutes of Health The Cancer Genome Atlas (TCGA)	
dbGAP phs000178. v11.p8		
Cancer hotspots	MSKCC	<a href="https://www.cancerhotspots.org/#/download (v2)">https://www.cancerhotspots.org/#/download (v2)</a>
Reference genomes	Broad	<a href="https://s3.console.aws.amazon.com/s3/buckets/broad-references/hg38/v0/">https://s3.console.aws.amazon.com/s3/buckets/broad-references/hg38/v0/</a>
Reference genome hg38, patch release 12	UCSC	<a href="http://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/">http://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/</a>
Human Cytoband file	UCSC	<a href="http://hgdownload.cse.ucsc.edu/goldenpath/hg38/database/cytoBand.txt.gz">http://hgdownload.cse.ucsc.edu/goldenpath/hg38/database/cytoBand.txt.gz</a>
CDS from GENCODE v27 annotation	GENCODE	<a href="https://www.gencodegenes.org/human/release_27.html">https://www.gencodegenes.org/human/release_27.html</a>
PFAM domains and locations	UCSC	<a href="http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/pfamDesc.txt.gz">http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/pfamDesc.txt.gz</a> ; <a href="https://pfam.xfam.org/family/PF07714">https://pfam.xfam.org/family/PF07714</a>
BSgenome.Hsapiens.UCSC.hg38 annotations	Bioconductor	<a href="https://bioconductor.org/packages/release/data/annotation/html/BSgenome.Hsapiens.UCSC.hg38.html">https://bioconductor.org/packages/release/data/annotation/html/BSgenome.Hsapiens.UCSC.hg38.html</a>

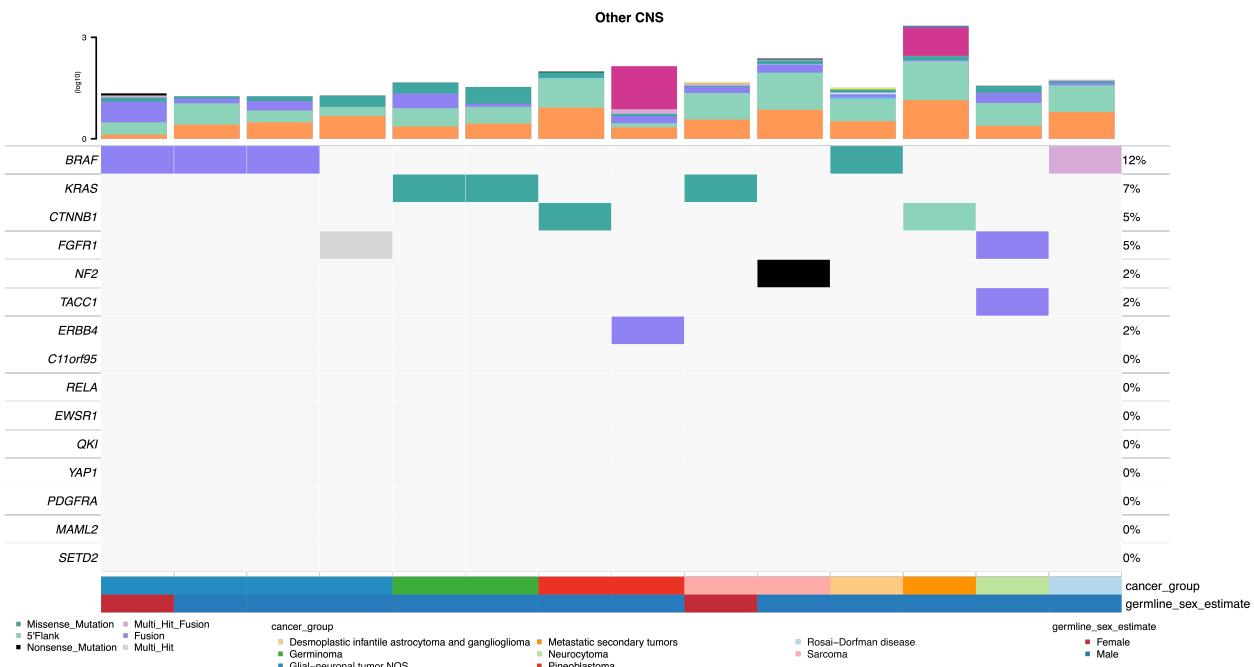
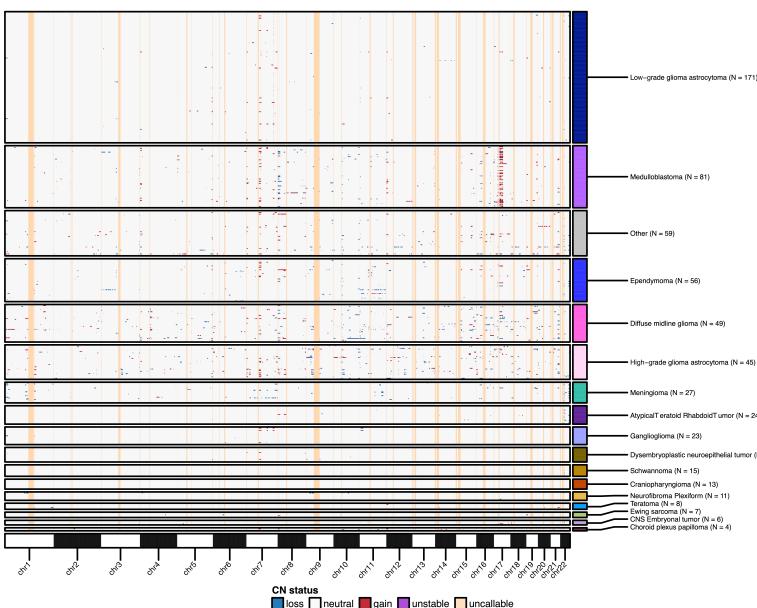
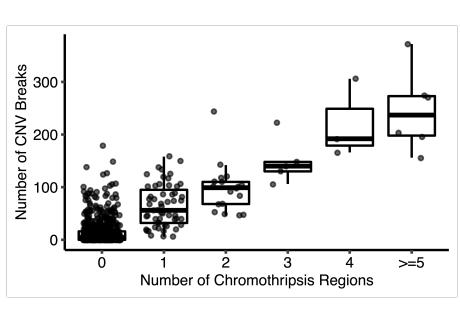
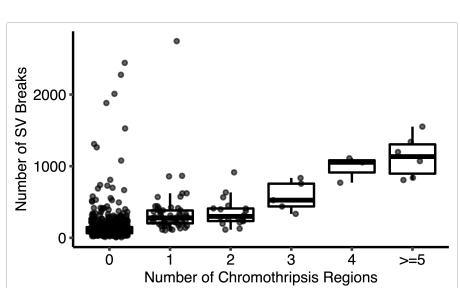
## Supplemental Information Titles and Legends



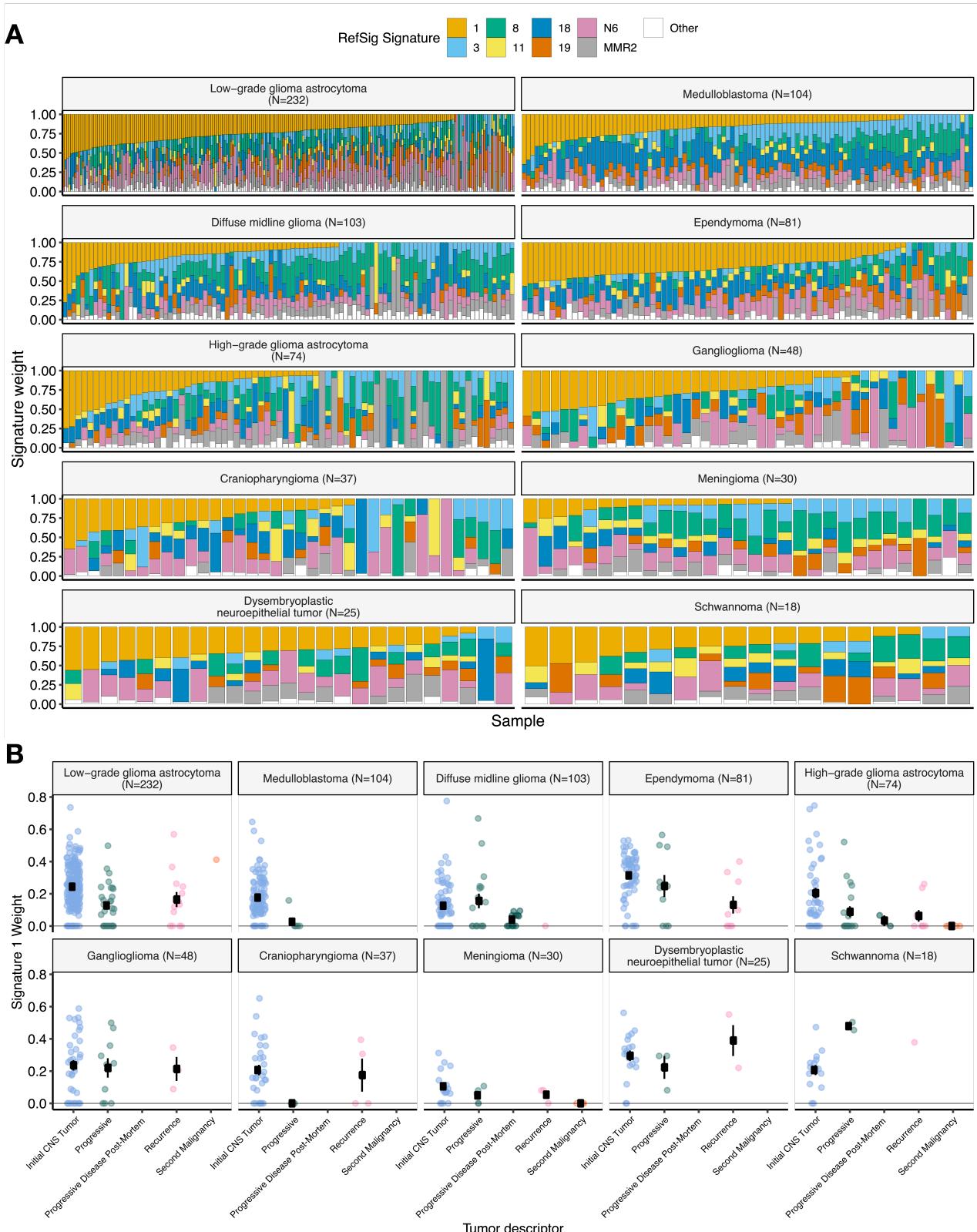
**Figure S1: OpenPBTA Project Workflow, Related to Figure 1.** Biospecimens and data were collected by CBTN and PNOC. Genomic sequencing and harmonization (orange boxes) were performed by the Kids First Data Resource Center (KFDRC). Analyses in the green boxes were performed by contributors of the OpenPBTA project. Output files are denoted in blue. Figure created with [BioRender.com](#).



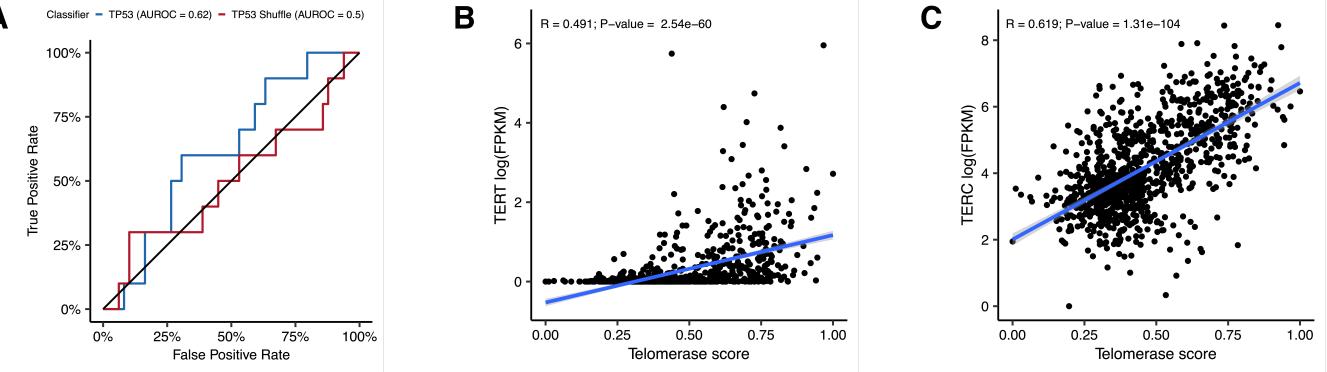
**Figure S2: Validation of Consensus SNV calls and Tumor Mutation Burden, Related to Figures 2 and 3.** Correlation (A) and violin (B) plots of mutation variant allele frequencies (VAFs) comparing the variant callers (Lancet, Strelka2, Mutect2, and VarDict) used for PBTA samples. Upset plot (C) showing overlap of variant calls. Correlation (D) and violin (E) plots of mutation variant allele frequencies (VAFs) comparing the variant callers (Lancet, Strelka2, and Mutect2) used for TCGA samples. Upset plot (F) showing overlap of variant calls. Violin plots (G) showing VAFs for Lancet calls performed on WGS and WXS from the same tumor (N = 52 samples from 13 patients). Cumulative distribution TMB plots for PBTA (H) and TCGA (I) tumors using consensus SNV calls.

**A****B****C****D**

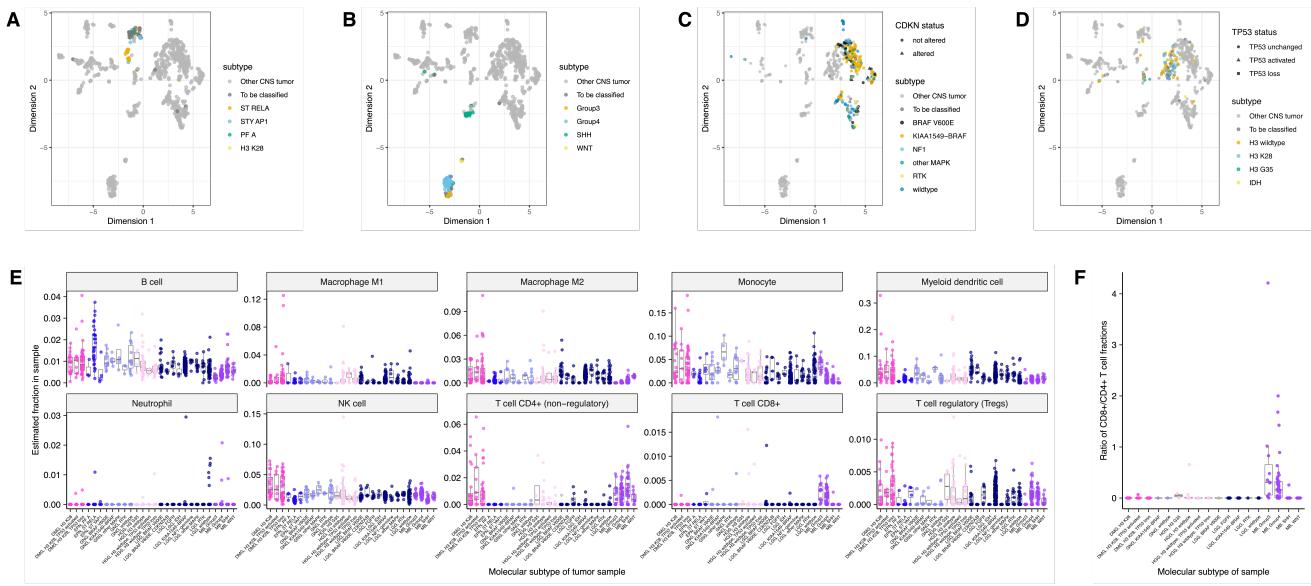
**Figure S3: Genomic instability of pediatric brain tumors, Related to Figures 2 and 3.** (A) Oncoprint of canonical somatic gene mutations, CNVs, fusions, and TMB (top bar plot) for the top 20 genes mutated across rare CNS tumors (N < 5 each): desmoplastic infantile astrocytoma and ganglioglioma (N = 1), germinoma (N = 4), glial-neuronal NOS (N = 4), metastatic secondary tumors (N = 3), neurocytoma (N = 2), and pineoblastoma (N = 3). Patient sex (germline\_sex\_estimate) and tumor histology (cancer\_group) are displayed as annotations at the bottom of each plot. Only primary tumors with mutations in the listed genes are shown. Multiple CNVs are denoted as a complex event. (B) Genome-wide plot of CNV alterations by broad histology. Each row represents one sample. Box and whisker plots of number of CNV breaks (C) or SV breaks (D) by number of chromothripsis regions.



**Figure S4: Mutational signatures in pediatric brain tumors, Related to Figure 3.** (A) Sample-specific RefSig signature weights across cancer groups ordered by decreasing Signature 1 exposure. (B) Proportion of Signature 1 plotted by phase of therapy for each cancer group.



**Figure S5: Quality control metrics for TP53 and EXTEND scores, Related to Figure 4.** (A) Receiver Operating Characteristic for TP53 classifier run on FPKM of poly-A RNA-Seq samples. Correlation plots for telomerase scores (EXTEND) with RNA expression of TERT(B) and TER(C).



**Figure S6: Subtype-specific clustering and immune cell fractions, Related to Figure 5.** First two dimensions from UMAP of sample transcriptome data with points colored by `molecular_subtype` for medulloblastoma (A), ependymoma (B), low-grade glioma (C), and high-grade diffuse astrocytic tumors (D). (E) Box plots of quanTlseq estimates of immune cell fractions in histologies with more than one molecular subtype with  $N \geq 3$ . (F) Box plots of the ratio of immune cell fractions of CD8+ to CD4+ T cells in histologies with more than one molecular subtype with  $N \geq 3$ .

**Table S1. Related to Figure 1.** Table of specimens and associated metadata, clinical data, and histological data utilized in the OpenPBTA project.

**Table S2. Related to Figures 2 and 3.** Excel file with three sheets representing tables of TMB, eight CNS mutational signatures, and chromothripsis events per sample, respectively.

**Table S3. Related to Figures 4 and 5.** Excel file with three sheets representing tables of TP53 scores, telomerase EXTEND scores, and quanTlseq immune scores, respectively.

**Table S4. Related to Figures 4 and 5.** Excel file with six sheets representing the survival analyses performed for this manuscript. See **Star Methods** for details.

**Table S5. Related to Figure 1.** Excel file with four sheets representing of all software and their respective versions used for the OpenPBTA project, including the R packages in the OpenPBTA Docker image, Python packages in the OpenPBTA Docker image, other command line tools in the OpenPBTA Docker image, and all software used in the OpenPBTA workflows, respectively. Note that all software

in the OpenPBTA Docker image was utilized within the analysis repository, but not all software was used for the final manuscript.

## Consortia

---

The past and present members of the Children's Brain Tumor Network who contributed to the generation of specimens and data are Adam C. Resnick, Alexa Plisiewicz, Allison M. Morgan, Allison P. Heath, Alyssa Paul, Amanda Saratsis, Amy Smith, Ana Aguilar, Ana Guerreiro Stucklin, Anastasia Arynchyna, Andrea Franson, Angela J. Waanders, Angela N. Viaene, Anita Nirenberg, Anna Maria Buccoliero, Anna Yaffe, Anny Shai, Anthony Bet, Antoinette Price, Ashley Plant, Augustine Eze, Bailey C. Farrow, Baoli Hu, Beth Frenkel, Bo Zhang, Bobby Moulder, Bonnie Cole, Brian M. Ennis, Brian R. Rood, Brittany Lebert, Carina A. Leonard, Carl Koschmann, Caroline Caudill, Caroline Drinkwater, Cassie N. Kline, Catherine Sullivan, Chanel Keoni, Chiara Caporalini, Christine Bobick-Butcher, Christopher Mason, Chunde Li, Claire Carter, Claudia MaduroCoronado, Clayton Wiley, Cynthia Wong, David E. Kram, David Haussler, David Kram, David Pisapia, David Ziegler, Denise Morinigo, Derek Hanson, Donald W. Parsons, Elizabeth Appert, Emily Drake, Emily Golbeck, Ena Agbodza, Eric H. Raabe, Eric M. Jackson, Eric Raabe, Esteban Uceda, Eugene Hwang, Fausto Rodriquez, Gabrielle S. Stone, Gary Kohanbash, George Rafidi, Gerald Grant, Gerri Trooskin, Gilad Evrony, Graham Keyes, Hagop Boyajian, Holly B. Lindsay, Holly Beale, Ian F. Pollack, James Johnston, James Palmer, Jane Minturn, Jared Pisapia, Jason E. Cain, Javad Nazarian, Jeanette Haugh, Jeff Greenfield, Jeff Stevens, Jeffrey P. Greenfield, Jeffrey Rubens, Jena V. Lilly, Jennifer L. Mason, Jessica B. Foster, Jim Olson, Jo Lynne Rokita, Joanna J. Phillips, Jonathan Waller, Josh Rubin, Judy E. Palma, Justine Rizzo, Kaitlin Lehmann, Kamnaa Arya, Karlene Hall, Katherine Pehlivan, Kenneth Seidl, Kimberly Diamond, Kristen Harnett, Kristina Cole, Krutika S. Gaonkar, Lamiya Tauhid, Lina Lopez, Lionel Chow, Mahdi Sarmady, Margaret Nevins, Mari Groves, Mariarita Santi-Vicini, Marilyn M. Li, Marion Mateos, Mateusz Koptyra, Matija Snuderl, Matthew Miller, Matthew Sklar, Matthew Wood, Meghan Connors, Melissa Williams, Meredith Egan, Michael Fisher, Michael Koldobskiy, Michelle Monje, Migdalia Martinez, Miguel A. Brown, Mike Prados, Miriam Bornhorst, Mirko Scagnet, Mohamed AbdelBaki, Monique Carrero-Tagle, Nadia Dahmane, Nalin Gupta, Nathan Young, Nicholas A. Vitanza, Nicholas Tassone, Nicolas Gerber, Nithin D. Adappa, Nitin Wadhwan, Obi Obayashi, Olena M. Vaske, Olivier Elemento, Oren Becher, Philbert Oliveros, Phillip B. Storm, Pichai Raman, Rintaro Hashizume, Robert Keating, Robert M. Lober, Ron Firestein, Sabine Mueller, Sameer Agnihotri, Samuel G. Winebrake, Samuel Rivero-Hinojosa, Sarah Diane Black, Sarah Leary, Schuyler Stoller, Shannon Robins, Sharon Gardner, Sherri Mayans, Sherry Tutson, Shida Zhu, Sofie Salama, Sonia Partap, Sonika Dahiya, Sriram Venneti, Stacie Stapleton, Stephani Campion, Stephanie Stefankiewicz, Swetha Thambireddy, Tatiana Patton, Teresa Hidalgo, Theo Nicolaides, Thinh Q. Nguyen, Tiffany Walker, Valeria Lopez-Gil, Valerie Baubet, Xiao-Nan Li, Ximena Cuellar, Yiran Guo, Yuankun Zhu, and Zeinab Helil.

The past and present members of the Pacific Pediatric Neuro-Oncology Consortium who contributed to the generation of specimens and data are Adam C. Resnick, Alicia Lenzen, Alyssa Reddy, Amar Gajjar, Ana Guerreiro Stucklin, Anat Epstein, Andrea Franson, Angela Waanders, Anne Bendel, Anu Banerjee, Ashley Margol, Ashley Plant, Brian Rood, Carl Koschmann, Carol Bruggers, Caroline Hastings, Cassie N. Kline, Christina Coleman Abadi, Christopher Tinkle, Corey Raffel, Dan Runco, Daniel Landi, Daphne Adele Haas-Kogan, David Ashley, David Ziegler, Derek Hanson, Dong Anh Khuong Quang, Duane Mitchell, Elias Sayour, Eric Jackson, Eric Raabe, Eugene Hwang, Fatema Malbari, Geoffrey McCowage, Girish Dhall, Gregory Friedman, Hideho Okada, Ibrahim Qaddoumi, Iris Fried, Jae Cho, Jane Minturn, Jason Blatt, Javad Nazarian, Jeffrey Rubens, Jena Lilly, Jennifer Elster, Jennifer Mason, Jessica Schulte, Jonathan Schoenfeld, Josh Rubin, Karen Gauvain, Karen Wright, Katharine Offer, Katie Metrock, Kellie Haworth, Ken Cohen, Kristina A Cole, Lance Governale, Linda Stork, Lindsay Kilburn, Lissa Baird, Maggie Skrypek, Marcia Leonard, Margaret Shatara, Margot Lazow, Mariella Filbin, Maryam Fouladi, Matthew Miller, Megan Paul, Michael Fisher, Michael Koldobskiy, Michal Yalon Oren, Mimi Bandopadhyay, Miriam Bornhorst, Mohamed AbdelBaki, Nalin Gupta, Nathan Robison, Nicholas Whipple, Nick Gottardo, Nick Vitanza, Nicolas Gerber, Patricia Robertson,

Payal Jain, Peter Sun, Prados, Michael, Priya Chan, Richard S Lemons, Robert Wechsler-Reya, Roger Packer, Russ Geyer, Sabine Mueller, Sahaja Acharya, Sam Cheshier, Sarah Leary, Scott Coven, Sharon Gardner, Sri Gururangan, Stewart Goldman, Susan Chi, Tab Cooney, Tatiana S. Patton, Theodore Nicolaides, and Tom Belle Davidson.

# References

---

1. **CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012–2016**  
Quinn T Ostrom, Gino Cioffi, Haley Gittleman, Nirav Patil, Kristin Waite, Carol Kruchko, Jill S Barnholtz-Sloan  
*Neuro-Oncology* (2019-11-01) <https://doi.org/gg4d4k>  
DOI: [10.1093/neuonc/noz150](https://doi.org/10.1093/neuonc/noz150) · PMID: [31675094](#) · PMCID: [PMC6823730](#)
2. **CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2009–2013**  
Quinn T. Ostrom, Haley Gittleman, Jordan Xu, Courtney Kromer, Yingli Wolinsky, Carol Kruchko, Jill S. Barnholtz-Sloan  
*Neuro-Oncology* (2016-10) <https://doi.org/ggzh6m>  
DOI: [10.1093/neuonc/nov207](https://doi.org/10.1093/neuonc/nov207) · PMID: [28475809](#) · PMCID: [PMC8483569](#)
3. **Years of life lived with disease and years of potential life lost in children who die of cancer in the United States, 2009**  
Peter M. Blank, Quinn T. Ostrom, Chaturia Rouse, Yingli Wolinsky, Carol Kruchko, Joanne Salcido, Jill S. Barnholtz-Sloan  
*Cancer Medicine* (2015-04) <https://doi.org/f69p34>  
DOI: [10.1002/cam4.410](https://doi.org/10.1002/cam4.410) · PMID: [25627000](#) · PMCID: [PMC4402075](#)
4. **FDA Approves Pediatric Formulation of Afinitor for Children with Subependymal Giant Cell Astrocytoma:**  
&Na;  
*Oncology Times* (2012-09) <https://doi.org/gp4qzf>  
DOI: [10.1097/01.cot.0000421359.88140.a8](https://doi.org/10.1097/01.cot.0000421359.88140.a8)
5. **FDA approves everolimus for tuberous sclerosis complex-associated partial-onset seizures**  
Center for Drug Evaluation and Research  
*FDA* (2018-11-03) <https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-everolimus-tuberous-sclerosis-complex-associated-partial-onset-seizures>
6. **FDA approves pembrolizumab for adults and children with TMB-H solid tumors**  
Center for Drug Evaluation and Research  
*FDA* (2020-06-17) <https://www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-pembrolizumab-adults-and-children-tmb-h-solid-tumors>
7. **FDA approves larotrectinib for solid tumors with NTRK gene fusions**  
Center for Drug Evaluation and Research  
*FDA* (2019-12-20) <https://www.fda.gov/drugs/fda-approves-larotrectinib-solid-tumors-ntrk-gene-fusions>
8. **FDA approves entrectinib for NTRK solid tumors and ROS-1 NSCLC**  
Center for Drug Evaluation and Research  
*FDA* (2019-12-20) <https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-entrectinib-ntrk-solid-tumors-and-ros-1-nsclc>
9. **FDA approves selumetinib for neurofibromatosis type 1 with symptomatic, inoperable plexiform neurofibromas**

Center for Drug Evaluation and Research

FDA (2020-04-13) <https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-selumetinib-neurofibromatosis-type-1-symptomatic-inoperable-plexiform-neurofibromas>

**10. Pediatric high-grade glioma resources from the Children's Brain Tumor Tissue Consortium**

Heba Ijaz, Mateusz Koptyra, Krutika S Gaonkar, Jo Lynne Rokita, Valerie P Baubet, Lamiya Tauhid, Yuankun Zhu, Miguel Brown, Gonzalo Lopez, Bo Zhang, ... Kristina A Cole

*Neuro-Oncology* (2020-01-11) <https://doi.org/gm3hpz>

DOI: [10.1093/neuonc/noz192](https://doi.org/10.1093/neuonc/noz192) · PMID: [31613963](https://pubmed.ncbi.nlm.nih.gov/31613963/) · PMCID: [PMC6954395](https://pubmed.ncbi.nlm.nih.gov/PMC6954395/)

**11. A pilot precision medicine trial for children with diffuse intrinsic pontine glioma—PNOC003: A report from the Pacific Pediatric Neuro-Oncology Consortium**

Sabine Mueller, Payal Jain, Winnie S. Liang, Lindsay Kilburn, Cassie Kline, Nalin Gupta, Eshini Panditharatna, Suresh N. Magge, Bo Zhang, Yuankun Zhu, ... Adam C. Resnick

*International Journal of Cancer* (2019-04-03) <https://doi.org/gf6pfb>

DOI: [10.1002/ijc.32258](https://doi.org/10.1002/ijc.32258) · PMID: [30861105](https://pubmed.ncbi.nlm.nih.gov/30861105/)

**12. Open collaborative writing with Manubot**

Daniel S. Himmelstein, Vincent Rubinetti, David R. Slochower, Dongbo Hu, Venkat S. Malladi, Casey S. Greene, Anthony Gitter

*PLOS Computational Biology* (2019-06-24) <https://doi.org/c7np>

DOI: [10.1371/journal.pcbi.1007128](https://doi.org/10.1371/journal.pcbi.1007128) · PMID: [31233491](https://pubmed.ncbi.nlm.nih.gov/31233491/) · PMCID: [PMC6611653](https://pubmed.ncbi.nlm.nih.gov/PMC6611653/)

**13. An Introduction to Rocker: Docker Containers for R**

Carl Boettiger, Dirk Eddelbuettel

*arXiv* (2017-10-11) <https://arxiv.org/abs/1710.03675>

**14. The WHO Classification of Tumors of the Nervous System**

Paul Kleihues, David N. Louis, Bernd W. Scheithauer, Lucy B. Rorke, Guido Reifenberger, Peter C. Burger, Webster K. Cavenee

*Journal of Neuropathology & Experimental Neurology* (2002-03) <https://doi.org/gg9fd2>

DOI: [10.1093/jnen/61.3.215](https://doi.org/10.1093/jnen/61.3.215) · PMID: [11895036](https://pubmed.ncbi.nlm.nih.gov/11895036/)

**15. The 2007 WHO Classification of Tumours of the Central Nervous System**

David N. Louis, Hiroko Ohgaki, Otmar D. Wiestler, Webster K. Cavenee, Peter C. Burger, Anne Jouvet, Bernd W. Scheithauer, Paul Kleihues

*Acta Neuropathologica* (2007-08) <https://doi.org/cm372n>

DOI: [10.1007/s00401-007-0243-4](https://doi.org/10.1007/s00401-007-0243-4) · PMID: [17618441](https://pubmed.ncbi.nlm.nih.gov/17618441/) · PMCID: [PMC1929165](https://pubmed.ncbi.nlm.nih.gov/PMC1929165/)

**16. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary**

David N. Louis, Arie Perry, Guido Reifenberger, Andreas von Deimling, Dominique Figarella-Branger, Webster K. Cavenee, Hiroko Ohgaki, Otmar D. Wiestler, Paul Kleihues, David W. Ellison

*Acta Neuropathologica* (2016-06) <https://doi.org/f8mspx>

DOI: [10.1007/s00401-016-1545-1](https://doi.org/10.1007/s00401-016-1545-1) · PMID: [27157931](https://pubmed.ncbi.nlm.nih.gov/27157931/)

**17. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary**

David N Louis, Arie Perry, Pieter Wesseling, Daniel J Brat, Ian A Cree, Dominique Figarella-Branger, Cynthia Hawkins, HK Ng, Stefan M Pfister, Guido Reifenberger, ... David W Ellison

*Neuro-Oncology* (2021-08-02) <https://doi.org/gmghbf>

DOI: [10.1093/neuonc/noab106](https://doi.org/10.1093/neuonc/noab106) · PMID: [34185076](https://pubmed.ncbi.nlm.nih.gov/34185076/) · PMCID: [PMC8328013](https://pubmed.ncbi.nlm.nih.gov/PMC8328013/)

## **18. Integrated Molecular and Clinical Analysis of 1,000 Pediatric Low-Grade Gliomas**

Scott Ryall, Michal Zapotocky, Kohei Fukuoka, Liana Nobre, Ana Guerreiro Stucklin, Julie Bennett, Robert Siddaway, Christopher Li, Sanja Pajovic, Anthony Arnoldo, ... Cynthia Hawkins  
*Cancer Cell* (2020-04) <https://doi.org/ggsmx2>  
DOI: [10.1016/j.ccr.2020.03.011](https://doi.org/10.1016/j.ccr.2020.03.011) · PMID: [32289278](https://pubmed.ncbi.nlm.nih.gov/32289278/) · PMCID: [PMC7169997](https://pubmed.ncbi.nlm.nih.gov/PMC7169997/)

## **19. Comprehensive Analysis of Hypermutation in Human Cancer**

Brittany B. Campbell, Nicholas Light, David Fabrizio, Matthew Zatzman, Fabio Fuligni, Richard de Borja, Scott Davidson, Melissa Edwards, Julia A. Elvin, Karl P. Hodel, ... Adam Shlien  
*Cell* (2017-11) <https://doi.org/gcjk7b>  
DOI: [10.1016/j.cell.2017.09.048](https://doi.org/10.1016/j.cell.2017.09.048) · PMID: [29056344](https://pubmed.ncbi.nlm.nih.gov/29056344/) · PMCID: [PMC5849393](https://pubmed.ncbi.nlm.nih.gov/PMC5849393/)

## **20. Pediatric low-grade glioma in the era of molecular diagnostics**

Scott Ryall, Uri Tabori, Cynthia Hawkins  
*Acta Neuropathologica Communications* (2020-12) <https://doi.org/gm3kst>  
DOI: [10.1186/s40478-020-00902-z](https://doi.org/10.1186/s40478-020-00902-z) · PMID: [32164789](https://pubmed.ncbi.nlm.nih.gov/32164789/) · PMCID: [PMC7066826](https://pubmed.ncbi.nlm.nih.gov/PMC7066826/)

## **21. ETMR: a tumor entity in its infancy**

Sander Lambo, Katja von Hoff, Andrey Korshunov, Stefan M. Pfister, Marcel Kool  
*Acta Neuropathologica* (2020-09) <https://doi.org/gm3q25>  
DOI: [10.1007/s00401-020-02182-2](https://doi.org/10.1007/s00401-020-02182-2) · PMID: [32601913](https://pubmed.ncbi.nlm.nih.gov/32601913/) · PMCID: [PMC7423804](https://pubmed.ncbi.nlm.nih.gov/PMC7423804/)

## **22. Emergence and maintenance of actionable genetic drivers at medulloblastoma relapse**

Stacey Richardson, Rebecca M Hill, Christopher Kui, Janet C Lindsey, Yura Grabovksa, Claire Keeling, Louise Pease, Matthew Bashton, Stephen Crosier, Maria Vinci, ... Steven C Clifford  
*Neuro-Oncology* (2022-01-05) <https://doi.org/gm3q3n>  
DOI: [10.1093/neuonc/noab178](https://doi.org/10.1093/neuonc/noab178) · PMID: [34272868](https://pubmed.ncbi.nlm.nih.gov/34272868/) · PMCID: [PMC8730763](https://pubmed.ncbi.nlm.nih.gov/PMC8730763/)

## **23. Molecular identification of CNS NB-FOXR2, CNS EFT-CIC, CNS HGNET-MN1 and CNS HGNET-BCOR pediatric brain tumors using tumor-specific signature genes**

Maria Łastowska, Joanna Trubicka, Anna Sobocińska, Bartosz Wojtas, Magdalena Niemira, Anna Szałkowska, Adam Krętowski, Agnieszka Karkucińska-Więckowska, Magdalena Kaleta, Maria Ejmont, ... Ewa Matyja  
*Acta Neuropathologica Communications* (2020-12) <https://doi.org/gm3q3p>  
DOI: [10.1186/s40478-020-00984-9](https://doi.org/10.1186/s40478-020-00984-9) · PMID: [32650833](https://pubmed.ncbi.nlm.nih.gov/32650833/) · PMCID: [PMC7350623](https://pubmed.ncbi.nlm.nih.gov/PMC7350623/)

## **24. The whole-genome landscape of medulloblastoma subtypes**

Paul A. Northcott, Ivo Buchhalter, A. Sorana Morrissy, Volker Hovestadt, Joachim Weischenfeldt, Tobias Ehrenberger, Susanne Gröbner, Maia Segura-Wang, Thomas Zichner, Vasilisa A. Rudneva, ... Peter Lichter  
*Nature* (2017-07) <https://doi.org/gbnzc9>  
DOI: [10.1038/nature22973](https://doi.org/10.1038/nature22973) · PMID: [28726821](https://pubmed.ncbi.nlm.nih.gov/28726821/) · PMCID: [PMC5905700](https://pubmed.ncbi.nlm.nih.gov/PMC5905700/)

## **25. Integrated Molecular Meta-Analysis of 1,000 Pediatric High-Grade and Diffuse Intrinsic Pontine Glioma**

Alan Mackay, Anna Burford, Diana Carvalho, Elisa Izquierdo, Janat Fazal-Salom, Kathryn R. Taylor, Lynn Bjerke, Matthew Clarke, Mara Vinci, Meera Nandhabalan, ... Chris Jones  
*Cancer Cell* (2017-10) <https://doi.org/gb4kpn>  
DOI: [10.1016/j.ccr.2017.08.017](https://doi.org/10.1016/j.ccr.2017.08.017) · PMID: [28966033](https://pubmed.ncbi.nlm.nih.gov/28966033/) · PMCID: [PMC5637314](https://pubmed.ncbi.nlm.nih.gov/PMC5637314/)

## **26. Diffuse intrinsic pontine glioma-like tumor with EZHIP expression and molecular features of PFA ependymoma**

Drew Pratt, Martha Quezado, Zied Abdullaev, Debra Hawes, Fusheng Yang, Hugh J. L. Garton,

Alexander R. Judkins, Rajen Mody, Arul Chinnaiyan, Kenneth Aldape, ... Sriram Venneti  
*Acta Neuropathologica Communications* (2020-12) <https://doi.org/gm3ksv>  
DOI: [10.1186/s40478-020-00905-w](https://doi.org/10.1186/s40478-020-00905-w) · PMID: [32197665](https://pubmed.ncbi.nlm.nih.gov/32197665/) · PMCID: [PMC7083001](https://pubmed.ncbi.nlm.nih.gov/PMC7083001/)

**27. Genomic Analysis of Dysembryoplastic Neuroepithelial Tumor Spectrum Reveals a Diversity of Molecular Alterations Dysregulating the MAPK and PI3K/mTOR Pathways**

Lea F Surrey, Payal Jain, Bo Zhang, Joshua Straka, Xiaonan Zhao, Brian N Harding, Adam C Resnick, Phillip B Storm, Anna Maria Buccoliero, Lorenzo Genitori, ... Mariarita Santi  
*Journal of Neuropathology & Experimental Neurology* (2019-12-01) <https://doi.org/gpnmp3p>  
DOI: [10.1093/jnen/nlz101](https://doi.org/10.1093/jnen/nlz101) · PMID: [31617914](https://pubmed.ncbi.nlm.nih.gov/31617914/)

**28. TP53 Mutation Is Frequently Associated With CTNNB1 Mutation or MYCN Amplification and Is Compatible With Long-Term Survival in Medulloblastoma**

Elke Pfaff, Marc Remke, Dominik Sturm, Axel Benner, Hendrik Witt, Till Milde, André O. von Bueren, Andrea Wittmann, Anna Schöttler, Norbert Jorch, ... Stefan M. Pfister  
*Journal of Clinical Oncology* (2010-12-10) <https://doi.org/dpd6hh>  
DOI: [10.1200/jco.2010.31.1670](https://doi.org/10.1200/jco.2010.31.1670) · PMID: [21060032](https://pubmed.ncbi.nlm.nih.gov/21060032/)

**29. The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma**

the St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project  
*Nature Genetics* (2014-05) <https://doi.org/f526f8>  
DOI: [10.1038/ng.2938](https://doi.org/10.1038/ng.2938) · PMID: [24705251](https://pubmed.ncbi.nlm.nih.gov/24705251/) · PMCID: [PMC4056452](https://pubmed.ncbi.nlm.nih.gov/PMC4056452/)

**30. A practical framework and online tool for mutational signature analyses show intertissue variation and driver dependencies**

Andrea Degasperi, Tauanne Dias Amarante, Jan Czarnecki, Scott Shooter, Xueqing Zou, Dominik Glodzik, Sandro Morganella, Arjun S. Nanda, Cherif Badja, Gene Koh, ... Serena Nik-Zainal  
*Nature Cancer* (2020-02-14) <https://doi.org/gjddc6>  
DOI: [10.1038/s43018-020-0027-5](https://doi.org/10.1038/s43018-020-0027-5) · PMID: [32118208](https://pubmed.ncbi.nlm.nih.gov/32118208/) · PMCID: [PMC7048622](https://pubmed.ncbi.nlm.nih.gov/PMC7048622/)

**31. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas**

Theo A. Knijnenburg, Linghua Wang, Michael T. Zimmermann, Nyasha Chambwe, Galen F. Gao, Andrew D. Cherniack, Huihui Fan, Hui Shen, Gregory P. Way, Casey S. Greene, ... Armaz Mariamidze  
*Cell Reports* (2018-04) <https://doi.org/gfspsc>  
DOI: [10.1016/j.celrep.2018.03.076](https://doi.org/10.1016/j.celrep.2018.03.076) · PMID: [29617664](https://pubmed.ncbi.nlm.nih.gov/29617664/) · PMCID: [PMC5961503](https://pubmed.ncbi.nlm.nih.gov/PMC5961503/)

**32. Genomic Profiling of Childhood Tumor Patient-Derived Xenograft Models to Enable Rational Clinical Trial Design**

Jo Lynne Rokita, Komal S. Rathi, Maria F. Cardenas, Kristen A. Upton, Joy Jayaseelan, Katherine L. Cross, Jacob Pfeil, Laura E. Egolf, Gregory P. Way, Alvin Farrel, ... John M. Maris  
*Cell Reports* (2019-11) <https://doi.org/gg596n>  
DOI: [10.1016/j.celrep.2019.09.071](https://doi.org/10.1016/j.celrep.2019.09.071) · PMID: [31693904](https://pubmed.ncbi.nlm.nih.gov/31693904/) · PMCID: [PMC6880934](https://pubmed.ncbi.nlm.nih.gov/PMC6880934/)

**33. Gain of function mutations in p53**

Dirk Dittmer, Sibani Pati, Gerard Zambetti, Shelley Chu, Angelika K. Teresky, Mary Moore, Cathy Finlay, Arnold J. Levine  
*Nature Genetics* (1993-05) <https://doi.org/crqst7>  
DOI: [10.1038/ng0593-42](https://doi.org/10.1038/ng0593-42) · PMID: [8099841](https://pubmed.ncbi.nlm.nih.gov/8099841/)

**34. Integrated analysis of telomerase enzymatic activity unravels an association with cancer stemness and proliferation**

Nighat Noureen, Shaofang Wu, Yingli Lv, Juechen Yang, W. K. Alfred Yung, Jonathan Gelfond, Xiaojing Wang, Dimpy Koul, Andrew Ludlow, Siyuan Zheng  
*Nature Communications* (2021-12) <https://doi.org/gmfxd>  
DOI: [10.1038/s41467-020-20474-9](https://doi.org/10.1038/s41467-020-20474-9) · PMID: [33420056](https://pubmed.ncbi.nlm.nih.gov/33420056/) · PMCID: [PMC7794223](https://pubmed.ncbi.nlm.nih.gov/PMC7794223/)

**35. Shorter telomeres and high telomerase activity correlate with a highly aggressive phenotype in breast cancer cell lines**

Hugo A. Ceja-Rangel, Patricia Sánchez-Suárez, Emilio Castellanos-Juárez, Rubicelia Peñaroja-Flores, Diego J. Arenas-Aranda, Patricio Gariglio, Luis Benítez-Bribiesca  
*Tumor Biology* (2016-09) <https://doi.org/f8987w>  
DOI: [10.1007/s13277-016-5045-7](https://doi.org/10.1007/s13277-016-5045-7) · PMID: [27072825](https://pubmed.ncbi.nlm.nih.gov/27072825/)

**36. Lynch Syndrome: An Updated Review**

Rishabh Sehgal, Kieran Sheahan, Patrick O'Connell, Ann Hanly, Sean Martin, Desmond Winter  
*Genes* (2014-06-27) <https://doi.org/gcfjbd>  
DOI: [10.3390/genes5030497](https://doi.org/10.3390/genes5030497) · PMID: [24978665](https://pubmed.ncbi.nlm.nih.gov/24978665/) · PMCID: [PMC4198913](https://pubmed.ncbi.nlm.nih.gov/PMC4198913/)

**37. Modeling and Targeting MYC Genes in Childhood Brain Tumors**

Sonja Hutter, Sara Bolin, Holger Weishaupl, Fredrik Swartling  
*Genes* (2017-03-23) <https://doi.org/gifbg8>  
DOI: [10.3390/genes8040107](https://doi.org/10.3390/genes8040107) · PMID: [28333115](https://pubmed.ncbi.nlm.nih.gov/28333115/) · PMCID: [PMC5406854](https://pubmed.ncbi.nlm.nih.gov/PMC5406854/)

**38. Myc requires distinct E2F activities to induce S phase and apoptosis.**

G Leone, R Sears, E Huang, R Rempel, F Nuckolls, CH Park, P Giangrande, L Wu, HI Saavedra, SJ Field, ... JR Nevins  
*Molecular cell* (2001-07) <https://www.ncbi.nlm.nih.gov/pubmed/11511364>  
DOI: [10.1016/s1097-2765\(01\)00275-1](https://doi.org/10.1016/s1097-2765(01)00275-1) · PMID: [11511364](https://pubmed.ncbi.nlm.nih.gov/11511364/)

**39. The inflammatory microenvironment in vestibular schwannoma**

Cathal John Hannan, Daniel Lewis, Claire O'Leary, Carmine A Donofrio, Dafydd Gareth Evans, Federico Roncaroli, David Brough, Andrew Thomas King, David Cope, Omar Nathan Pathmanaban  
*Neuro-Oncology Advances* (2020-01-01) <https://doi.org/gpwh65>  
DOI: [10.1093/noajnl/vdaa023](https://doi.org/10.1093/noajnl/vdaa023) · PMID: [32642684](https://pubmed.ncbi.nlm.nih.gov/32642684/) · PMCID: [PMC7212860](https://pubmed.ncbi.nlm.nih.gov/PMC7212860/)

**40. Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer**

Francesca Petralia, Nicole Tignor, Boris Reva, Mateusz Koptyra, Shrabanti Chowdhury, Dmitry Rykunov, Azra Krek, Weiping Ma, Yuankun Zhu, Jiayi Ji, ... William E. Bocik  
*Cell* (2020-12) <https://doi.org/ghqjkz>  
DOI: [10.1016/j.cell.2020.10.044](https://doi.org/10.1016/j.cell.2020.10.044) · PMID: [33242424](https://pubmed.ncbi.nlm.nih.gov/33242424/) · PMCID: [PMC8143193](https://pubmed.ncbi.nlm.nih.gov/PMC8143193/)

**41. Non-inflammatory tumor microenvironment of diffuse intrinsic pontine glioma**

Grant L. Lin, Surya Nagaraja, Mariella G. Filbin, Mario L. Suvà, Hannes Vogel, Michelle Monje  
*Acta Neuropathologica Communications* (2018-12) <https://doi.org/gp2nq6>  
DOI: [10.1186/s40478-018-0553-x](https://doi.org/10.1186/s40478-018-0553-x) · PMID: [29954445](https://pubmed.ncbi.nlm.nih.gov/29954445/) · PMCID: [PMC6022714](https://pubmed.ncbi.nlm.nih.gov/PMC6022714/)

**42. PD-L1 expression in medulloblastoma: an evaluation by subgroup**

Allison M. Martin, Christopher J. Nirschl, Magda J. Polanczyk, W Robert Bell, Thomas R. Nirschl, Sarah Harris-Bookman, Jillian Phallen, Jessica Hicks, Daniel Martinez, Aleksandra Ogurtsova, ... Michael Lim  
*Oncotarget* (2018-04-10) <https://doi.org/gdhjdd>  
DOI: [10.18632/oncotarget.24951](https://doi.org/10.18632/oncotarget.24951) · PMID: [29721192](https://pubmed.ncbi.nlm.nih.gov/29721192/) · PMCID: [PMC5922386](https://pubmed.ncbi.nlm.nih.gov/PMC5922386/)

#### **43. Subgroup-specific immune and stromal microenvironment in medulloblastoma**

Michael Bockmayr, Malte Mohme, Frederick Klauschen, Beate Winkler, Jan Budczies, Stefan Rutkowski, Ulrich Schüller

*OncolImmunology* (2018-09-02) <https://doi.org/gd9q8n>

DOI: [10.1080/2162402x.2018.1462430](https://doi.org/10.1080/2162402x.2018.1462430) · PMID: [30228931](#) · PMCID: [PMC6140816](#)

#### **44. MM2S: personalized diagnosis of medulloblastoma patients and model systems**

Deena M. A. Gendoo, Benjamin Haibe-Kains

*Source Code for Biology and Medicine* (2016-12) <https://doi.org/ghcqf2>

DOI: [10.1186/s13029-016-0053-y](https://doi.org/10.1186/s13029-016-0053-y) · PMID: [27069505](#) · PMCID: [PMC4827218](#)

#### **45. A transcriptome-based classifier to determine molecular subtypes in medulloblastoma**

Komal S. Rathi, Sherjeel Arif, Mateusz Koptyra, Ammar S. Naqvi, Deanne M. Taylor, Phillip B. Storm, Adam C. Resnick, Jo Lynne Rokita, Pichai Raman

*PLOS Computational Biology* (2020-10-29) <https://doi.org/gm84kg>

DOI: [10.1371/journal.pcbi.1008263](https://doi.org/10.1371/journal.pcbi.1008263) · PMID: [33119584](#) · PMCID: [PMC7654754](#)

#### **46. Intracystic interferon- $\alpha$ treatment leads to neurotoxicity in craniopharyngioma: case report**

Julia Sharma, Christopher M. Bonfield, Ash Singhal, Juliette Hukin, Paul Steinbok

*Journal of Neurosurgery: Pediatrics* (2015-09) <https://doi.org/f7n4gk>

DOI: [10.3171/2015.2.peds14656](https://doi.org/10.3171/2015.2.peds14656) · PMID: [26023721](#)

#### **47. Unexpected brain atrophy following administration of intratumoral interferon alpha-2b for cystic craniopharyngioma: A case report**

Khadiga Elfadil Ahmed Mohammed, Kellie R. Alleyne Mike, Jeannette Parkes

*International Journal of Case Reports and Images (IJCRI)* (2014-02-08)

<http://www.ijcasereportsandimages.com/archive/2013/012-2013-ijcri/013-12-2013-mohammed/index.php>

DOI: [10.5348/ijcri-2013-12-419-cr-13](https://doi.org/10.5348/ijcri-2013-12-419-cr-13)

#### **48. Multiplexed immunofluorescence reveals potential PD-1/PD-L1 pathway vulnerabilities in craniopharyngioma**

Shannon Coy, Rumana Rashid, Jia-Ren Lin, Ziming Du, Andrew M Donson, Todd C Hankinson, Nicholas K Foreman, Peter E Manley, Mark W Kieran, David A Reardon, ... Sandro Santagata

*Neuro-Oncology* (2018-07-05) <https://doi.org/gdw7v7>

DOI: [10.1093/neuonc/noy035](https://doi.org/10.1093/neuonc/noy035) · PMID: [29509940](#) · PMCID: [PMC6280314](#)

#### **49. A Novel Immune Classification for Predicting Immunotherapy Responsiveness in Patients With Adamantinomatous Craniopharyngioma.**

Feng Yuan, Xiangming Cai, Junhao Zhu, Lei Yuan, Yingshuai Wang, Chao Tang, Zixiang Cong, Chiyuan Ma

*Frontiers in neurology* (2021-12-13) <https://www.ncbi.nlm.nih.gov/pubmed/34966342>

DOI: [10.3389/fneur.2021.704130](https://doi.org/10.3389/fneur.2021.704130) · PMID: [34966342](#) · PMCID: [PMC8710480](#)

#### **50. The Inflammatory Milieu of Adamantinomatous Craniopharyngioma and Its Implications for Treatment.**

Ros Whelan, Eric Prince, Ahmed Gilani, Todd Hankinson

*Journal of clinical medicine* (2020-02-14) <https://www.ncbi.nlm.nih.gov/pubmed/32075140>

DOI: [10.3390/jcm9020519](https://doi.org/10.3390/jcm9020519) · PMID: [32075140](#) · PMCID: [PMC7074265](#)

#### **51. Tumour compartment transcriptomics demonstrates the activation of inflammatory and odontogenic programmes in human adamantinomatous craniopharyngioma and identifies the MAPK/ERK pathway as a novel therapeutic target**

John R. Apps, Gabriela Carreno, Jose Mario Gonzalez-Meljem, Scott Haston, Romain Guiho, Julie E. Cooper, Saba Manshaei, Nital Jani, Annett Hölsken, Benedetta Pettorini, ... Juan Pedro Martinez-Barbera

*Acta Neuropathologica* (2018-05) <https://doi.org/gdfjjc>

DOI: [10.1007/s00401-018-1830-2](https://doi.org/s00401-018-1830-2) · PMID: [29541918](https://pubmed.ncbi.nlm.nih.gov/29541918/) · PMCID: [PMC5904225](https://pubmed.ncbi.nlm.nih.gov/PMC5904225/)

## 52. Targeting IL-6 Is a Potential Treatment for Primary Cystic Craniopharyngioma

Sydney Grob, David M. Mirsky, Andrew M. Donson, Nathan Dahl, Nicholas K. Foreman, Lindsey M. Hoffman, Todd C. Hankinson, Jean M. Mulcahy Levy

*Frontiers in Oncology* (2019-08-21) <https://doi.org/gp3s4d>

DOI: [10.3389/fonc.2019.00791](https://doi.org/10.3389/fonc.2019.00791) · PMID: [31497533](https://pubmed.ncbi.nlm.nih.gov/31497533/) · PMCID: [PMC6712354](https://pubmed.ncbi.nlm.nih.gov/PMC6712354/)

## 53. annoFuse: an R Package to annotate, prioritize, and interactively explore putative oncogenic RNA fusions

Krutika S. Gaonkar, Federico Marini, Komal S. Rathi, Payal Jain, Yuankun Zhu, Nicholas A. Chimicles, Miguel A. Brown, Ammar S. Naqvi, Bo Zhang, Phillip B. Storm, ... Jo Lynne Rokita

*BMC Bioinformatics* (2020-12) <https://doi.org/gm84mh>

DOI: [10.1186/s12859-020-03922-7](https://doi.org/10.1186/s12859-020-03922-7) · PMID: [33317447](https://pubmed.ncbi.nlm.nih.gov/33317447/) · PMCID: [PMC7737294](https://pubmed.ncbi.nlm.nih.gov/PMC7737294/)

## 54. Macrophages in SHH subgroup medulloblastoma display dynamic heterogeneity that varies with treatment modality

Mai T. Dang, Michael V. Gonzalez, Krutika S. Gaonkar, Komal S. Rathi, Patricia Young, Sherjeel Arif, Li Zhai, Zahidul Alam, Samir Devalaraja, Tsun Ki Jerrick To, ... Malay Haldar

*Cell Reports* (2021-03) <https://doi.org/gjmxjr>

DOI: [10.1016/j.celrep.2021.108917](https://doi.org/10.1016/j.celrep.2021.108917) · PMID: [33789113](https://pubmed.ncbi.nlm.nih.gov/33789113/)

## 55. Development of GPC2-directed chimeric antigen receptors using mRNA for pediatric brain tumors

Jessica B. Foster, Crystal Griffin, Jo Lynne Rokita, Allison Stern, Cameron Brimley, Komal Rathi, Maria V. Lane, Samantha N. Buongervino, Tiffany Smith, Peter J. Madsen, ... Kristopher R. Bosse

*Cancer Biology* (2021-07-07) <https://doi.org/gpvsz8>

DOI: [10.1101/2021.07.06.451385](https://doi.org/10.1101/2021.07.06.451385)

## 56. PNOC008: A Pilot Trial Testing the Clinical Benefit of Using Molecular Profiling to Determine an Individualized Treatment Plan in Children and Young Adults with High Grade Glioma (Excluding Diffuse Intrinsic Pontine Glioma) – Pacific Pediatric Neuro-Oncology Consortium <https://pnoc.us/clinical-trial/pnoc008/>

## 57. Open Pediatric Brain Tumor Atlas

Children's Brain Tumor Network, Pediatric Neuro Oncology Consortium Open Pediatric Brain Tumor Atlas

*Kids First Data Resource Center* (2022) <https://doi.org/gpp5dv>

DOI: [10.24370/openpbta](https://doi.org/10.24370/openpbta)

## 58. Pediatric High Grade Glioma Resources From the Children's Brain Tumor Tissue Consortium (CBTTC) and Pediatric Brain Tumor Atlas (PBTA)

Heba Ijaz, Mateusz Koptyra, Krutika S. Gaonkar, Jo Lynne Rokita, Valerie P. Baubet, Lamiya Tauhid, Yankun Zhu, Miguel Brown, Gonzalo Lopez, Bo Zhang, ... Kristina A. Cole

*Cancer Biology* (2019-05-31) <https://doi.org/gf66qt>

DOI: [10.1101/656587](https://doi.org/10.1101/656587)

## 59. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM

Heng Li

60. **SAMBLASTER: fast duplicate marking and structural variant read extraction**

G. G. Faust, I. M. Hall

*Bioinformatics* (2014-09-01) <https://doi.org/f6kft3>

DOI: [10.1109/bioinformatics.btu314](https://doi.org/btu314) · PMID: [24812344](https://pubmed.ncbi.nlm.nih.gov/24812344/) · PMCID: [PMC4147885](https://pubmed.ncbi.nlm.nih.gov/PMC4147885/)

61. **Sambamba: fast processing of NGS alignment formats**

Artem Tarasov, Albert J. Vilella, Edwin Cuppen, Isaac J. Nijman, Pjotr Prins

*Bioinformatics* (2015-06-15) <https://doi.org/gfzsfw>

DOI: [10.1109/bioinformatics.btv098](https://doi.org/btv098) · PMID: [25697820](https://pubmed.ncbi.nlm.nih.gov/25697820/) · PMCID: [PMC4765878](https://pubmed.ncbi.nlm.nih.gov/PMC4765878/)

62. **Scaling accurate genetic variant discovery to tens of thousands of samples**

Ryan Poplin, Valentin Ruano-Rubio, Mark A. DePristo, Tim J. Fennell, Mauricio O. Carneiro, Geraldine A. Van der Auwera, David E. Kling, Laura D. Gauthier, Ami Levy-Moonshine, David Roazen, ... Eric Banks

*Genomics* (2017-11-14) <https://doi.org/ggmrvr>

DOI: [10.1101/201178](https://doi.org/10.1101/201178)

63. **NGSCheckMate: software for validating sample identity in next-generation sequencing studies within and across data types**

Sejoon Lee, Soohyun Lee, Scott Ouellette, Woong-Yang Park, Eunjung A. Lee, Peter J. Park

*Nucleic Acids Research* (2017-06-20) <https://doi.org/f9xrq4>

DOI: [10.1093/nar/gkx193](https://doi.org/nar/gkx193) · PMID: [28369524](https://pubmed.ncbi.nlm.nih.gov/28369524/) · PMCID: [PMC5499645](https://pubmed.ncbi.nlm.nih.gov/PMC5499645/)

64. **Strelka2: fast and accurate calling of germline and somatic variants**

Sangtae Kim, Konrad Scheffler, Aaron L. Halpern, Mitchell A. Bekritsky, Eunho Noh, Morten Källberg, Xiaoyu Chen, Yeonbin Kim, Doruk Beyter, Peter Krusche, Christopher T. Saunders

*Nature Methods* (2018-08) <https://doi.org/gdwrp4>

DOI: [10.1038/s41592-018-0051-x](https://doi.org/10.1038/s41592-018-0051-x) · PMID: [30013048](https://pubmed.ncbi.nlm.nih.gov/30013048/)

65. **Calling Somatic SNVs and Indels with Mutect2**

David Benjamin, Takuto Sato, Kristian Cibulskis, Gad Getz, Chip Stewart, Lee Lichtenstein

*Bioinformatics* (2019-12-02) <https://doi.org/ggnntw>

DOI: [10.1101/861054](https://doi.org/10.1101/861054)

66. **Genome-wide somatic variant calling using localized colored de Bruijn graphs**

Giuseppe Narzisi, André Corvelo, Kanika Arora, Ewa A. Bergmann, Minita Shah, Rajeeva Musunuri, Anne-Katrin Emde, Nicolas Robine, Vladimir Vacic, Michael C. Zody

*Communications Biology* (2018-12) <https://doi.org/gfcfr8>

DOI: [10.1038/s42003-018-0023-9](https://doi.org/10.1038/s42003-018-0023-9) · PMID: [30271907](https://pubmed.ncbi.nlm.nih.gov/30271907/) · PMCID: [PMC6123722](https://pubmed.ncbi.nlm.nih.gov/PMC6123722/)

67. **VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research**

Zhongwu Lai, Aleksandra Markovets, Miika Ahdesmaki, Brad Chapman, Oliver Hofmann, Robert McEwen, Justin Johnson, Brian Dougherty, J. Carl Barrett, Jonathan R. Dry

*Nucleic Acids Research* (2016-06-20) <https://doi.org/f8v6qz>

DOI: [10.1109/nar/gkw227](https://doi.org/nar/gkw227) · PMID: [27060149](https://pubmed.ncbi.nlm.nih.gov/27060149/) · PMCID: [PMC4914105](https://pubmed.ncbi.nlm.nih.gov/PMC4914105/)

68. **Deep sequencing of 3 cancer cell lines on 2 sequencing platforms**

Kanika Arora, Minita Shah, Molly Johnson, Rakesh Sanghvi, Jennifer Shelton, Kshithija Nagulapalli, Dayna M. Oschwald, Michael C. Zody, Soren Germer, Vaidehi Jobanputra, ... Nicolas Robine

*Genomics* (2019-04-30) <https://doi.org/ggc9vx>  
DOI: [10.1101/623702](https://doi.org/10.1101/623702)

## 69. The Ensembl Variant Effect Predictor

William McLaren, Laurent Gil, Sarah E. Hunt, Harpreet Singh Riat, Graham R. S. Ritchie, Anja Thormann, Paul Flicek, Fiona Cunningham  
*Genome Biology* (2016-12) <https://doi.org/gdz75c>  
DOI: [10.1186/s13059-016-0974-4](https://doi.org/10.1186/s13059-016-0974-4) · PMID: [27268795](https://pubmed.ncbi.nlm.nih.gov/27268795/) · PMCID: [PMC4893825](https://pmcid.ncbi.nlm.nih.gov/pmc/articles/PMC4893825/)

## 70. The mutational constraint spectrum quantified from variation in 141,456 humans

Konrad J. Karczewski, Laurent C. Francioli, Grace Tiao, Beryl B. Cummings, Jessica Alföldi, Qingbo Wang, Ryan L. Collins, Kristen M. Laricchia, Andrea Ganna, Daniel P. Birnbaum, ... Daniel G. MacArthur  
*Nature* (2020-05-28) <https://doi.org/ggztw9>  
DOI: [10.1038/s41586-020-2308-7](https://doi.org/10.1038/s41586-020-2308-7) · PMID: [32461654](https://pubmed.ncbi.nlm.nih.gov/32461654/) · PMCID: [PMC7334197](https://pmcid.ncbi.nlm.nih.gov/pmc/articles/PMC7334197/)

## 71. Activating Telomerase TERT Promoter Mutations and Their Application for the Detection of Bladder Cancer

Maria Zvereva, Eduard Pisarev, Ismail Hosen, Olga Kisil, Simon Matskeplishvili, Elena Kubareva, David Kamalov, Alexander Tivtikyan, Arnaud Manel, Emmanuel Vian, ... Florence Le Calvez-Kelm  
*International Journal of Molecular Sciences* (2020-08-21) <https://doi.org/gmf45b>  
DOI: [10.3390/ijms21176034](https://doi.org/10.3390/ijms21176034) · PMID: [32839402](https://pubmed.ncbi.nlm.nih.gov/32839402/) · PMCID: [PMC7503716](https://pmcid.ncbi.nlm.nih.gov/pmc/articles/PMC7503716/)

## 72. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data

Valentina Boeva, Tatiana Popova, Kevin Bleakley, Pierre Chiche, Julie Cappo, Gudrun Schleiermacher, Isabelle Janoueix-Lerosey, Olivier Delattre, Emmanuel Barillot  
*Bioinformatics* (2012-02-01) <https://doi.org/ckt4vz>  
DOI: [10.1093/bioinformatics/btr670](https://doi.org/10.1093/bioinformatics/btr670) · PMID: [22155870](https://pubmed.ncbi.nlm.nih.gov/22155870/) · PMCID: [PMC3268243](https://pmcid.ncbi.nlm.nih.gov/pmc/articles/PMC3268243/)

## 73. Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization

Valentina Boeva, Andrei Zinovyev, Kevin Bleakley, Jean-Philippe Vert, Isabelle Janoueix-Lerosey, Olivier Delattre, Emmanuel Barillot  
*Bioinformatics* (2011-01-15) <https://doi.org/c6bcps>  
DOI: [10.1093/bioinformatics/btq635](https://doi.org/10.1093/bioinformatics/btq635) · PMID: [21081509](https://pubmed.ncbi.nlm.nih.gov/21081509/) · PMCID: [PMC3018818](https://pmcid.ncbi.nlm.nih.gov/pmc/articles/PMC3018818/)

## 74. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing

Eric Talevich, A. Hunter Shain, Thomas Botton, Boris C. Bastian  
*PLOS Computational Biology* (2016-04-21) <https://doi.org/c9pd>  
DOI: [10.1371/journal.pcbi.1004873](https://doi.org/10.1371/journal.pcbi.1004873) · PMID: [27100738](https://pubmed.ncbi.nlm.nih.gov/27100738/) · PMCID: [PMC4839673](https://pmcid.ncbi.nlm.nih.gov/pmc/articles/PMC4839673/)

## 75. Quantifying tumor heterogeneity in whole-genome and whole-exome sequencing data

Layla Oesper, Gryte Satas, Benjamin J. Raphael  
*Bioinformatics* (2014-12-15) <https://doi.org/f6rmgt>  
DOI: [10.1093/bioinformatics/btu651](https://doi.org/10.1093/bioinformatics/btu651) · PMID: [25297070](https://pubmed.ncbi.nlm.nih.gov/25297070/) · PMCID: [PMC4253833](https://pmcid.ncbi.nlm.nih.gov/pmc/articles/PMC4253833/)

## 76. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers

Craig H Mermel, Steven E Schumacher, Barbara Hill, Matthew L Meyerson, Rameen Beroukhim, Gad Getz

**77. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications**

Xiaoyu Chen, Ole Schulz-Trieglaff, Richard Shaw, Bret Barnes, Felix Schlesinger, Morten Källberg, Anthony J. Cox, Semyon Kruglyak, Christopher T. Saunders

*Bioinformatics* (2016-04-15) <https://doi.org/gf3ggb>

DOI: [10.1093/bioinformatics/btv710](https://doi.org/10.1093/bioinformatics/btv710) · PMID: [26647377](#)

**78. AnnotSV: an integrated tool for structural variations annotation**

Véronique Geoffroy, Yvan Herenger, Arnaud Kress, Corinne Stoetzel, Amélie Piton, Hélène Dollfus, Jean Muller

*Bioinformatics* (2018-10-15) <https://doi.org/gdcsh3>

DOI: [10.1093/bioinformatics/bty304](https://doi.org/10.1093/bioinformatics/bty304) · PMID: [29669011](#)

**79. STAR: ultrafast universal RNA-seq aligner**

Alexander Dobin, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, Thomas R. Gingeras

*Bioinformatics* (2013-01) <https://doi.org/f4h523>

DOI: [10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635) · PMID: [23104886](#) · PMCID: [PMC3530905](#)

**80. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome**

Bo Li, Colin N Dewey

*BMC Bioinformatics* (2011-12) <https://doi.org/cwg8n5>

DOI: [10.1186/1471-2105-12-323](https://doi.org/10.1186/1471-2105-12-323) · PMID: [21816040](#) · PMCID: [PMC3163565](#)

**81. Accurate and efficient detection of gene fusions from RNA sequencing data**

Sebastian Uhrig, Julia Ellermann, Tatjana Walther, Pauline Burkhardt, Martina Fröhlich, Barbara Hutter, Umut H. Toprak, Olaf Neumann, Albrecht Stenzinger, Claudia Scholl, ... Benedikt Brors

*Genome Research* (2021-03) <https://doi.org/gjvdvp>

DOI: [10.1101/gr.257246.119](https://doi.org/10.1101/gr.257246.119) · PMID: [33441414](#) · PMCID: [PMC7919457](#)

**82. STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq**

Brian J. Haas, Alex Dobin, Nicolas Stransky, Bo Li, Xiao Yang, Timothy Tickle, Asma Bankapur, Carrie Ganote, Thomas G. Doak, Nathalie Pochet, ... Aviv Regev

*Bioinformatics* (2017-03-24) <https://doi.org/gf5pc5>

DOI: [10.1101/120295](https://doi.org/10.1101/120295)

**83. Maftools: efficient and comprehensive analysis of somatic variants in cancer**

Anand Mayakonda, De-Chen Lin, Yassen Assenov, Christoph Plass, H. Phillip Koeffler

*Genome Research* (2018-11) <https://doi.org/gfmnwf>

DOI: [10.1101/gr.239244.118](https://doi.org/10.1101/gr.239244.118) · PMID: [30341162](#) · PMCID: [PMC6211645](#)

**84. BEDTools: a flexible suite of utilities for comparing genomic features**

Aaron R. Quinlan, Ira M. Hall

*Bioinformatics* (2010-03-15) <https://doi.org/cmrrms3>

DOI: [10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033) · PMID: [20110278](#) · PMCID: [PMC2832824](#)

**85. The UCSC Genome Browser database: extensions and updates 2013**

Laurence R. Meyer, Ann S. Zweig, Angie S. Hinrichs, Donna Karolchik, Robert M. Kuhn, Matthew Wong, Cricket A. Sloan, Kate R. Rosenbloom, Greg Roe, Brooke Rhead, ... W. James Kent

*Nucleic Acids Research* (2012-11-15) <https://doi.org/f4jr4v>  
DOI: [10.1093/nar/gks1048](https://doi.org/10.1093/nar/gks1048) · PMID: [23155063](https://pubmed.ncbi.nlm.nih.gov/23155063/) · PMCID: [PMC3531082](https://pubmed.ncbi.nlm.nih.gov/PMC3531082/)

## 86. Software for Computing and Annotating Genomic Ranges

Michael Lawrence, Wolfgang Huber, Hervé Pagès, Patrick Aboyou, Marc Carlson, Robert Gentleman, Martin T. Morgan, Vincent J. Carey

*PLoS Computational Biology* (2013-08-08) <https://doi.org/f5cmfg>

DOI: [10.1371/journal.pcbi.1003118](https://doi.org/10.1371/journal.pcbi.1003118) · PMID: [23950696](https://pubmed.ncbi.nlm.nih.gov/23950696/) · PMCID: [PMC3738458](https://pubmed.ncbi.nlm.nih.gov/PMC3738458/)

## 87. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing

Isidro Cortés-Ciriano, Jake June-Koo Lee, Ruibin Xi, Dhawal Jain, Youngsook L. Jung, Lixing Yang, Dmitry Gordenin, Leszek J. Klimczak, Cheng-Zhong Zhang, David S. Pellman, ... PCAWG Consortium  
*Nature Genetics* (2020-03-02) <https://doi.org/ggkkpt>

DOI: [10.1038/s41588-019-0576-7](https://doi.org/10.1038/s41588-019-0576-7) · PMID: [32025003](https://pubmed.ncbi.nlm.nih.gov/32025003/) · PMCID: [PMC7058534](https://pubmed.ncbi.nlm.nih.gov/PMC7058534/)

## 88. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology.

Gregor Sturm, Francesca Finotello, Florent Petitprez, Jitao David Zhang, Jan Baumbach, Wolf H Fridman, Markus List, Tatsiana Aneichyk

*Bioinformatics (Oxford, England)* (2019-07-15) <https://www.ncbi.nlm.nih.gov/pubmed/31510660>  
DOI: [10.1093/bioinformatics/btz363](https://doi.org/btz363) · PMID: [31510660](https://pubmed.ncbi.nlm.nih.gov/31510660/) · PMCID: [PMC6612828](https://pubmed.ncbi.nlm.nih.gov/PMC6612828/)

## 89. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data

Francesca Finotello, Clemens Mayer, Christina Plattner, Gerhard Laschober, Dietmar Rieder, Hubert Hackl, Anne Krogsdam, Zuzana Loncova, Wilfried Posch, Doris Wilflingseder, ... Zlatko Trajanoski

*Genome Medicine* (2019-12) <https://doi.org/gf5jwc>

DOI: [10.1186/s13073-019-0638-6](https://doi.org/10.1186/s13073-019-0638-6) · PMID: [31126321](https://pubmed.ncbi.nlm.nih.gov/31126321/) · PMCID: [PMC6534875](https://pubmed.ncbi.nlm.nih.gov/PMC6534875/)

## 90. GSVA: gene set variation analysis for microarray and RNA-Seq data

Sonja Hänelmann, Robert Castelo, Justin Guinney

*BMC Bioinformatics* (2013-12) <https://doi.org/gb8vx5>

DOI: [10.1186/1471-2105-14-7](https://doi.org/10.1186/1471-2105-14-7) · PMID: [23323831](https://pubmed.ncbi.nlm.nih.gov/23323831/) · PMCID: [PMC3618321](https://pubmed.ncbi.nlm.nih.gov/PMC3618321/)

## 91. GSVA

Justin Guinney [Aut, Cre], Robert Castelo [Aut], Joan Fernandez[Ctb]

*Bioconductor* (2017) <https://doi.org/ggxrgs>

DOI: [10.18129/b9.bioc.gsva](https://doi.org/10.18129/b9.bioc.gsva)

## 92. The Molecular Signatures Database Hallmark Gene Set Collection

Arthur Liberzon, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, Pablo Tamayo

*Cell Systems* (2015-12) <https://doi.org/gf78hq>

DOI: [10.1016/j.cels.2015.12.004](https://doi.org/10.1016/j.cels.2015.12.004) · PMID: [26771021](https://pubmed.ncbi.nlm.nih.gov/26771021/) · PMCID: [PMC4707969](https://pubmed.ncbi.nlm.nih.gov/PMC4707969/)

## 93. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction

Leland McInnes, John Healy, James Melville

*arXiv* (2018-02-09) <https://arxiv.org/abs/1802.03426v2>

## 94. The Human Transcription Factors

Samuel A. Lambert, Arttu Jolma, Laura F. Campitelli, Pratyush K. Das, Yimeng Yin, Mihai Albu,

Xiaoting Chen, Jussi Taipale, Timothy R. Hughes, Matthew T. Weirauch  
*Cell* (2018-02) <https://doi.org/gcw8rb>  
DOI: [10.1016/j.cell.2018.01.029](https://doi.org/10.1016/j.cell.2018.01.029) · PMID: [29425488](https://pubmed.ncbi.nlm.nih.gov/29425488/)

**95. Genomic analysis of diffuse pediatric low-grade gliomas identifies recurrent oncogenic truncating rearrangements in the transcription factor *MYBL1***

Lori A. Ramkissoon, Peleg M. Horowitz, Justin M. Craig, Shakti H. Ramkissoon, Benjamin E. Rich, Steven E. Schumacher, Aaron McKenna, Michael S. Lawrence, Guillaume Bergthold, Priscilla K. Brastianos, ... Keith L. Ligon  
*Proceedings of the National Academy of Sciences* (2013-05-14) <https://doi.org/f42gg4>  
DOI: [10.1073/pnas.1300252110](https://doi.org/10.1073/pnas.1300252110) · PMID: [23633565](https://pubmed.ncbi.nlm.nih.gov/23633565/) · PMCID: [PMC3657784](https://pubmed.ncbi.nlm.nih.gov/PMC3657784/)

**96. Subgroup-specific structural variation across 1,000 medulloblastoma genomes**

Paul A. Northcott, David J. H. Shih, John Peacock, Livia Garzia, A. Sorana Morrissy, Thomas Zichner, Adrian M. Stütz, Andrey Korshunov, Jüri Reimand, Steven E. Schumacher, ... Michael D. Taylor  
*Nature* (2012-08) <https://doi.org/ggdhk3>  
DOI: [10.1038/nature11327](https://doi.org/10.1038/nature11327) · PMID: [22832581](https://pubmed.ncbi.nlm.nih.gov/22832581/) · PMCID: [PMC3683624](https://pubmed.ncbi.nlm.nih.gov/PMC3683624/)

**97. New Brain Tumor Entities Emerge from Molecular Classification of CNS-PNETs**

Dominik Sturm, Brent A. Orr, Umut H. Toprak, Volker Hovestadt, David T.W. Jones, David Capper, Martin Sill, Ivo Buchhalter, Paul A. Northcott, Irina Leis, ... Marcel Kool  
*Cell* (2016-02) <https://doi.org/f3t869>  
DOI: [10.1016/j.cell.2016.01.015](https://doi.org/10.1016/j.cell.2016.01.015) · PMID: [26919435](https://pubmed.ncbi.nlm.nih.gov/26919435/) · PMCID: [PMC5139621](https://pubmed.ncbi.nlm.nih.gov/PMC5139621/)

**98. Fusion of TTYH1 with the C19MC microRNA cluster drives expression of a brain-specific DNMT3B isoform in the embryonal brain tumor ETMR**

Claudia L Kleinman, Noha Gerges, Simon Papillon-Cavanagh, Patrick Sin-Chan, Albena Pramatarova, Dong-Anh Khuong Quang, Véronique Adoue, Stephan Busche, Maxime Caron, Haig Djambazian, ... Nada Jabado  
*Nature Genetics* (2014-01) <https://doi.org/ggdhk4>  
DOI: [10.1038/ng.2849](https://doi.org/10.1038/ng.2849) · PMID: [24316981](https://pubmed.ncbi.nlm.nih.gov/24316981/)

**99. TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors**

Linda J Valentijn, Jan Koster, Danny A Zwijnenburg, Nancy E Hasselt, Peter van Sluis, Richard Volckmann, Max M van Noesel, Rani E George, Godelieve AM Tytgat, Jan J Molenaar, Rogier Versteeg  
*Nature Genetics* (2015-12) <https://doi.org/ggdhk5>  
DOI: [10.1038/ng.3438](https://doi.org/10.1038/ng.3438) · PMID: [26523776](https://pubmed.ncbi.nlm.nih.gov/26523776/)

**100. Recurrent pre-existing and acquired DNA copy number alterations, including focal *TERT* gains, in neuroblastoma central nervous system metastases: Copy Number Changes in Neuroblastoma Cns Metastases**

David Cobrinik, Irina Ostrovnaya, Maryam Hassimi, Satish K. Tickoo, Irene Y. Cheung, Nai-Kong V. Cheung  
*Genes, Chromosomes and Cancer* (2013-12) <https://doi.org/f5gd94>  
DOI: [10.1002/gcc.22110](https://doi.org/10.1002/gcc.22110) · PMID: [24123354](https://pubmed.ncbi.nlm.nih.gov/24123354/)

**101. Activation of human telomerase reverse transcriptase through gene fusion in clear cell sarcoma of the kidney**

Jenny Karlsson, Henrik Lilljebjörn, Linda Holmquist Mengelbier, Anders Valind, Marianne Rissler, Ingrid Øra, Thoas Fioretos, David Gisselsson  
*Cancer Letters* (2015-02) <https://doi.org/f25ck5>  
DOI: [10.1016/j.canlet.2014.11.057](https://doi.org/10.1016/j.canlet.2014.11.057) · PMID: [25481751](https://pubmed.ncbi.nlm.nih.gov/25481751/)

**102. New Molecular Considerations for Glioma: IDH, ATRX, BRAF, TERT, H3 K27M**

Michael Karsy, Jian Guan, Adam L. Cohen, Randy L. Jensen, Howard Colman

*Current Neurology and Neuroscience Reports* (2017-02) <https://doi.org/ggdhk2>

DOI: [10.1007/s11910-017-0722-5](https://doi.org/s11910-017-0722-5) · PMID: [28271343](https://pubmed.ncbi.nlm.nih.gov/28271343/)

**103. MYB-QKI rearrangements in angiocentric glioma drive tumorigenicity through a tripartite mechanism**

Pratiti Bandopadhyay, Lori A Ramkissoon, Payal Jain, Guillaume Bergthold, Jeremiah Wala, Rhamy Zeid, Steven E Schumacher, Laura Urbanski, Ryan O'Rourke, William J Gibson, ... Adam C Resnick

*Nature Genetics* (2016-03) <https://doi.org/f8bwqn>

DOI: [10.1038/ng.3500](https://doi.org/ng.3500) · PMID: [26829751](https://pubmed.ncbi.nlm.nih.gov/26829751/) · PMCID: [PMC4767685](https://pubmed.ncbi.nlm.nih.gov/PMC4767685/)

**104. Atypical Teratoid/Rhabdoid Tumors Are Comprised of Three Epigenetic Subgroups with Distinct Enhancer Landscapes**

Pascal D. Johann, Serap Erkek, Marc Zapatka, Cornelius Kerl, Ivo Buchhalter, Volker Hovestadt, David T. W. Jones, Dominik Sturm, Carl Hermann, Maia Segura Wang, ... Marcel Kool

*Cancer Cell* (2016-03) <https://doi.org/f8fmz4>

DOI: [10.1016/j.ccr.2016.02.001](https://doi.org/10.1016/j.ccr.2016.02.001) · PMID: [26923874](https://pubmed.ncbi.nlm.nih.gov/26923874/)

**105. Chromosome 19 microRNA cluster enhances cell reprogramming by inhibiting epithelial-to-mesenchymal transition**

Ezinne F. Mong, Ying Yang, Kemal M. Akat, John Canfield, Jeffrey VanWye, John Lockhart, John C. M. Tsibris, Frederick Schatz, Charles J. Lockwood, Thomas Tuschl, ... Hana Totary-Jain

*Scientific Reports* (2020-12) <https://doi.org/gg329d>

DOI: [10.1038/s41598-020-59812-8](https://doi.org/s41598-020-59812-8) · PMID: [32080251](https://pubmed.ncbi.nlm.nih.gov/32080251/) · PMCID: [PMC7033247](https://pubmed.ncbi.nlm.nih.gov/PMC7033247/)

**106. H3 K27M mutations are extremely rare in posterior fossa group A ependymoma**

Scott Ryall, Miguel Guzman, Samer K. Elbabaa, Betty Luu, Stephen C. Mack, Michal Zapotocky, Michael D. Taylor, Cynthia Hawkins, Vijay Ramaswamy

*Child's Nervous System* (2017-07) <https://doi.org/gbn2dn>

DOI: [10.1007/s00381-017-3481-3](https://doi.org/s00381-017-3481-3) · PMID: [28623522](https://pubmed.ncbi.nlm.nih.gov/28623522/)

**107. Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups**

Kristian W. Pajtler, Hendrik Witt, Martin Sill, David T.W. Jones, Volker Hovestadt, Fabian Kratochwil, Khalida Wani, Ruth Tatevossian, Chandanamali Punchihewa, Pascal Johann, ... Stefan M. Pfister

*Cancer Cell* (2015-05) <https://doi.org/f7ct8f>

DOI: [10.1016/j.ccr.2015.04.002](https://doi.org/10.1016/j.ccr.2015.04.002) · PMID: [25965575](https://pubmed.ncbi.nlm.nih.gov/25965575/) · PMCID: [PMC4712639](https://pubmed.ncbi.nlm.nih.gov/PMC4712639/)

**108. C11orf95-RELA fusions drive oncogenic NF-κB signalling in ependymoma**

Matthew Parker, Kumarasamypet M. Mohankumar, Chandanamali Punchihewa, Ricardo Weinlich, James D. Dalton, Yongjin Li, Ryan Lee, Ruth G. Tatevossian, Timothy N. Phoenix, Radhika Thiruvenkatam, ... Richard J. Gilbertson

*Nature* (2014-02-27) <https://doi.org/f5t87h>

DOI: [10.1038/nature13109](https://doi.org/nature13109) · PMID: [24553141](https://pubmed.ncbi.nlm.nih.gov/24553141/) · PMCID: [PMC4050669](https://pubmed.ncbi.nlm.nih.gov/PMC4050669/)

**109. Genomic landscape of high-grade meningiomas**

Wenya Linda Bi, Noah F. Greenwald, Malak Abedalthagafi, Jeremiah Wala, Will J. Gibson, Pankaj K. Agarwalla, Peleg Horowitz, Steven E. Schumacher, Ekaterina Esaulova, Yu Mei, ... Rameen Beroukhim

*npj Genomic Medicine* (2017-12) <https://doi.org/gbnhbw>

DOI: [10.1038/s41525-017-0014-7](https://doi.org/s41525-017-0014-7) · PMID: [28713588](https://pubmed.ncbi.nlm.nih.gov/28713588/) · PMCID: [PMC5506858](https://pubmed.ncbi.nlm.nih.gov/PMC5506858/)

**110. Correlations between genomic subgroup and clinical features in a cohort of more than 3000 meningiomas**

Mark W. Youngblood, Daniel Duran, Julio D. Montejo, Chang Li, Sacit Bulent Omay, Koray Ozduman, Amar H. Sheth, Amy Y. Zhao, Evgeniya Tyrtova, Danielle F. Miyagishima, ... Murat Günel  
*Journal of Neurosurgery* (2020-11) <https://doi.org/gm3q3r>  
DOI: [10.3171/2019.8.jns191266](https://doi.org/10.3171/2019.8.jns191266) · PMID: [31653806](#)

**111. Genetic alterations in uncommon low-grade neuroepithelial tumors: BRAF, FGFR1, and MYB mutations occur at high frequency and align with morphology**

Ibrahim Qaddoumi, Wilda Orisme, Ji Wen, Teresa Santiago, Kirti Gupta, James D. Dalton, Bo Tang, Kelly Haupfear, Chandanamali Punchihewa, John Easton, ... David W. Ellison  
*Acta Neuropathologica* (2016-06) <https://doi.org/f8m3pn>  
DOI: [10.1007/s00401-016-1539-z](https://doi.org/10.1007/s00401-016-1539-z) · PMID: [26810070](#) · PMCID: [PMC4866893](#)

**112. The genetic landscape of choroid plexus tumors in children and adults**

Christian Thomas, Patrick Soschinski, Melissa Zwaig, Spyridon Oikonomopoulos, Konstantin Okonechnikov, Kristian W Pajtler, Martin Sill, Leonille Schweizer, Arend Koch, Julia Neumann, ... Martin Hasselblatt  
*Neuro-Oncology* (2021-04-12) <https://doi.org/gppwxt>  
DOI: [10.1093/neuonc/noaa267](https://doi.org/10.1093/neuonc/noaa267) · PMID: [33249490](#) · PMCID: [PMC8041331](#)

**113. Langerhans cell histiocytosis in children**

Jolie Krooks, Milen Minkov, Angela G. Weatherall  
*Journal of the American Academy of Dermatology* (2018-06) <https://doi.org/gdmdp2>  
DOI: [10.1016/j.jaad.2017.05.059](https://doi.org/10.1016/j.jaad.2017.05.059) · PMID: [29754885](#)

**114. EZHIP is a specific diagnostic biomarker for posterior fossa ependymomas, group PFA and diffuse midline gliomas H3-WT with EZHIP overexpression**

C. Antin, A. Tauziède-Espriat, M.-A. Debily, D. Castel, J. Grill, M. Pagès, O. Ayrault, F. Chrétien, A. Gareton, F. Andreiuolo, ... P. Varlet  
*Acta Neuropathologica Communications* (2020-12) <https://doi.org/gm3q3q>  
DOI: [10.1186/s40478-020-01056-8](https://doi.org/10.1186/s40478-020-01056-8) · PMID: [33153494](#) · PMCID: [PMC7643397](#)

**115. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution**

Rachel Rosenthal, Nicholas McGranahan, Javier Herrero, Barry S. Taylor, Charles Swanton  
*Genome Biology* (2016-12) <https://doi.org/f8bdsg>  
DOI: [10.1186/s13059-016-0893-4](https://doi.org/10.1186/s13059-016-0893-4) · PMID: [26899170](#) · PMCID: [PMC4762164](#)

**116. A polyphenotypic malignant paediatric brain tumour presenting a MN1-PATZ1 fusion, no epigenetic similarities with CNS High-Grade Neuroepithelial Tumour with MN1 Alteration (CNS HGNET-MN1) and related to PATZ1-fused sarcomas**

F. Burel-Vandenbos, G. Pierron, C. Thomas, S. Reynaud, V. Gregoire, G. Duhil de Benaze, S. Croze, N. Chivoret, M. Honavar, D. Figarella-Branger, ... C. Godfraind  
*Neuropathology and Applied Neurobiology* (2020-08) <https://doi.org/gmndph>  
DOI: [10.1111/nan.12626](https://doi.org/10.1111/nan.12626) · PMID: [32397004](#)

**117. Embryonal Tumors of the Central Nervous System in Children: The Era of Targeted Therapeutics.**

David E Kram, Jacob J Henderson, Muhammad Baig, Diya Chakraborty, Morgan A Gardner, Subhasree Biswas, Soumen Khatua  
*Bioengineering (Basel, Switzerland)* (2018-09-23) <https://www.ncbi.nlm.nih.gov/pubmed/30249036>  
DOI: [10.3390/bioengineering5040078](https://doi.org/10.3390/bioengineering5040078) · PMID: [30249036](#) · PMCID: [PMC6315657](#)

**118. LIN28A, a sensitive immunohistochemical marker for Embryonal Tumor with Multilayered Rosettes (ETMR), is also positive in a subset of Atypical Teratoid/Rhabdoid Tumor (AT/RT)**  
Shilpa Rao, R. T. Rajeswarie, T. Chickabasaviah Yasha, Bevinahalli N. Nandeesh, Arimappamagan Arivazhagan, Vani Santosh  
*Child's Nervous System* (2017-11) <https://doi.org/ggnpkn>  
DOI: [10.1007/s00381-017-3551-6](https://doi.org/10.1007/s00381-017-3551-6) · PMID: [28744687](https://pubmed.ncbi.nlm.nih.gov/28744687/)

**119. Childhood Medulloblastoma and Other Central Nervous System Embryonal Tumors Treatment (PDQ®)-Health Professional Version - NCI** (2008-02-13)  
<https://www.cancer.gov/types/brain/hp/child-cns-embryonal-treatment-pdq>

**120. DNA Methylation Profiling for Diagnosing Undifferentiated Sarcoma with Capicua Transcriptional Receptor (CIC) Alterations**  
Evelina Miele, Rita De Vito, Andrea Ciolfi, Lucia Pedace, Ida Russo, Maria Debora De Pasquale, Angela Di Giannatale, Alessandro Crocoli, Biagio De Angelis, Marco Tartaglia, ... Giuseppe Maria Milano  
*International Journal of Molecular Sciences* (2020-03-06) <https://doi.org/ggnn7x>  
DOI: [10.3390/ijms21051818](https://doi.org/10.3390/ijms21051818) · PMID: [32155762](https://pubmed.ncbi.nlm.nih.gov/32155762/) · PMCID: [PMC7084764](https://pubmed.ncbi.nlm.nih.gov/PMC7084764/)

**121. LIN28A immunoreactivity is a potent diagnostic marker of embryonal tumor with multilayered rosettes (ETMR)**  
Andrey Korshunov, Marina Ryzhova, David T. W. Jones, Paul A. Northcott, Peter van Sluis, Richard Volckmann, Jan Koster, Rogier Versteeg, Cynthia Cowdrey, Arie Perry, ... Marcel Kool  
*Acta Neuropathologica* (2012-12) <https://doi.org/f4dxvc>  
DOI: [10.1007/s00401-012-1068-3](https://doi.org/10.1007/s00401-012-1068-3) · PMID: [23161096](https://pubmed.ncbi.nlm.nih.gov/23161096/) · PMCID: [PMC3508282](https://pubmed.ncbi.nlm.nih.gov/PMC3508282/)

**122. ITD assembler: an algorithm for internal tandem duplication discovery from short-read sequencing data**  
Navin Rustagi, Oliver A Hampton, Jie Li, Liu Xi, Richard A. Gibbs, Sharon E. Plon, Marek Kimmel, David A. Wheeler  
*BMC Bioinformatics* (2016-12) <https://doi.org/gmndpj>  
DOI: [10.1186/s12859-016-1031-8](https://doi.org/10.1186/s12859-016-1031-8) · PMID: [27121965](https://pubmed.ncbi.nlm.nih.gov/27121965/) · PMCID: [PMC4847212](https://pubmed.ncbi.nlm.nih.gov/PMC4847212/)

**123. Central Neurocytoma and Extraventricular Neurocytoma**  
Carrie A. Mohila, Ronald A. Rauch, Adekunle M. Adesina  
*Atlas of Pediatric Brain Tumors* (2016) <https://doi.org/gmndn7>  
DOI: [10.1007/978-3-319-33432-5\\_20](https://doi.org/10.1007/978-3-319-33432-5_20) · ISBN: [9783319334301](https://www.ncbi.nlm.nih.gov/books/NBK4334301/)

**124. Papillary craniopharyngioma: a clinicopathological study of 48 cases**  
Thomas B. Crotty, Bernd W. Scheithauer, William F. Young, Dudley H. Davis, Edward G. Shaw, Gary M. Miller, Peter C. Burger  
*Journal of Neurosurgery* (1995-08) <https://doi.org/ccpj22>  
DOI: [10.3171/jns.1995.83.2.0206](https://doi.org/10.3171/jns.1995.83.2.0206) · PMID: [7616262](https://pubmed.ncbi.nlm.nih.gov/7616262/)

**125. The descriptive epidemiology of craniopharyngioma**  
Greta R. Bunin, Tanya S. Surawicz, Philip A. Witman, Susan Preston-Martin, Faith Davis, Janet M. Bruner  
*Journal of Neurosurgery* (1998-10) <https://doi.org/fh9xsh>  
DOI: [10.3171/jns.1998.89.4.0547](https://doi.org/10.3171/jns.1998.89.4.0547) · PMID: [9761047](https://pubmed.ncbi.nlm.nih.gov/9761047/)

**126. Accelerating Discovery of Functional Mutant Alleles in Cancer**  
Matthew T. Chang, Tripti Shrestha Bhattacharai, Alison M. Schram, Craig M. Bielski, Mark T. A. Donoghue, Philip Jonsson, Debyani Chakravarty, Sarah Phillips, Cyriac Kandoth, Alexander Penson,

... Barry S. Taylor  
*Cancer Discovery* (2018-02) <https://doi.org/gf9twp>  
DOI: [10.1158/2159-8290.cd-17-0321](https://doi.org/10.1158/2159-8290.cd-17-0321) · PMID: [29247016](#) · PMCID: [PMC5809279](#)

**127. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity**

Matthew T Chang, Saurabh Asthana, Sizhi Paul Gao, Byron H Lee, Jocelyn S Chapman, Cyriac Kandoth, JianJiong Gao, Nicholas D Soccia, David B Solit, Adam B Olshen, ... Barry S Taylor  
*Nature Biotechnology* (2016-02) <https://doi.org/gf7vxg>  
DOI: [10.1038/nbt.3391](https://doi.org/10.1038/nbt.3391) · PMID: [26619011](#) · PMCID: [PMC4744099](#)

**128. The functional domains in p53 family proteins exhibit both common and distinct properties**

KL Harms, X Chen  
*Cell Death & Differentiation* (2006-06) <https://doi.org/fwgrrt>  
DOI: [10.1038/sj.cdd.4401904](https://doi.org/10.1038/sj.cdd.4401904) · PMID: [16543939](#)

**129. Inherited TP53 Mutations and the Li-Fraumeni Syndrome**

Tanya Guha, David Malkin  
*Cold Spring Harbor Perspectives in Medicine* (2017-04) <https://doi.org/f9s4h3>  
DOI: [10.1101/cshperspect.a026187](https://doi.org/10.1101/cshperspect.a026187) · PMID: [28270529](#) · PMCID: [PMC5378014](#)

**130. The Sequence Alignment/Map format and SAMtools.**

Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin,  
*Bioinformatics (Oxford, England)* (2009-06-08) <https://www.ncbi.nlm.nih.gov/pubmed/19505943>  
DOI: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) · PMID: [19505943](#) · PMCID: [PMC2723002](#)

**131. Inferring Telomerase Enzymatic Activity from Expression Data**

Nighat Noureen, Shaofang Wu, Yingli Lyu, Juechen Yang, WK Alfred Yung, Jonathan Gelfond, Xiaojing Wang, Dimpy Koul, Andrew Ludlow, Siyuan Zheng  
*Bioinformatics* (2020-05-26) <https://doi.org/ggxfhq>  
DOI: <https://doi.org/10.1101/2020.05.21.109249>

**132. Nonparametric Estimation from Incomplete Observations**

E. L. Kaplan, Paul Meier  
*Journal of the American Statistical Association* (1958-06) <https://doi.org/fscrh2>  
DOI: [10.2307/2281868](https://doi.org/10.2307/2281868)

**133. Evaluation of survival data and two new rank order statistics arising in its consideration.**

N Mantel  
*Cancer chemotherapy reports* (1966-03) <https://www.ncbi.nlm.nih.gov/pubmed/5910392>  
PMID: [5910392](#)

**134. <http://www.jstor.org/stable/2985181>**

**135. d3b-center/OpenPBTA-workflows: Release v1.0.1**

Jo Lynne Rokita, Miguel Brown  
*Zenodo* (2022-04-20) <https://doi.org/gpxs5q>  
DOI: [10.5281/zenodo.6474116](https://doi.org/10.5281/zenodo.6474116)