An Open Pediatric Brain Tumor Atlas

Authors

- **Joshua A. Shapiro**
  0000-0002-6224-0347 · jashapiro · jashapiro
  Childhood Cancer Data Lab, Alex's Lemonade Stand Foundation, Philadelphia, 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, Philadelphia, 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, Philadelphia, 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
  Department of Bioinformatics and Health Informatics, Children's Hospital of Philadelphia

- **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
  Rowan University, Glassboro, NJ, USA · Funded by Alex's Lemonade Stand Foundation Childhood Cancer Data Lab (CCDL)

- **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

- **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

- **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

- **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

- **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 · 🐙 cgreene · 🐦 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, Philadelphia, 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, Philadelphia, 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, rokita@chop.edu

In Brief

Highlights

Summary

Keywords

pediatric brain tumors, somatic variation, open science, classification

Introduction

Pediatric brain and spinal cord tumors are the second most common tumors in children after leukemia, yet 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]. Despite their relative rarity, the years of potential life lost due to brain tumors in 2009 was estimated at 47,631 years for children and adolescents aged 0-19 in the United States [**???**]. The low survival rates for some tumors are clearly multifactorial but can be explained in part by our lack of understanding of the ever-evolving array of brain tumor molecular subtypes, difficulty drugging these entities, 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 line and mouse models has been a barrier to research. Furthermore, the incidence of any single molecular tumor entity is relatively low due to the rarity of pediatric tumors in general. Together, these factors have hindered research progress and have led to multiple national and international center and consortia efforts to collaboratively share specimens and data to accelerate breakthroughs and clinical translation.

There has been significant progress in recent years to elucidate the landscape of somatic variation responsible for pediatric brain tumor formation and progression, however, translation of therapeutic agents to phase II or III clinical trials and subsequent FDA approval has been slow. 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; anti-PD-1 immunotherapy, pembrolizumab, for microsatellite instability–high or mismatch repair–deficient tumors; NTRK inhibitors larotrectinib and entrectinib for tumors with an NTRK 1/2/3 gene fusions; MEK1/2 inhibitor, selumetinib, for neurofibromatosis type 1 (NF1) and symptomatic, inoperable plexiform neurofibromas. This is, in part, due to pharmaceutical company priorities and/or concerns regarding toxicity that

have resulted in an inability to obtain drugs for pediatric clinical trials, ultimately delaying access to new agents. An amendment to the Pediatric Research Equity Act called the Research to Accelerate Cures and Equity (RACE) for Children Act mandates that as of August 18, 2020 all new adult oncology drugs also be tested in children when the molecular targets are relevant to a particular childhood cancer. Here, we present a comprehensive, collaborative, open genomic analysis of 943 patient tumors from 59 distinct brain tumor histologies which can be used to support the RACE Act and accelerate rational clinical trial design.

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) [3], consisting of samples from the Children's Brain Tumor Network (CBTN) [4] and the PNOC003 DMG clinical trial [5] of the Pacific Pediatric Neuro-oncology Symposium (PNOC) (**Figure ?? A**). We then harnessed the benchmarking efforts of the KidsFirst Data Resource Center to develop a robust and reproducible data analysis workflow within the CAVATICA platform to perform primary somatic analyses: variant calling of single nucleotide variants (SNVs), copy number variants (CNVs), structural variants (SVs), and fusions (**Figure 5** - red boxes and **STAR Methods**). Next, we created a Github analysis repository (https://github.com/AlexsLemonade/OpenPBTA-analysis) with continuous integration to ensure analysis reproducibility and a GitHub manuscript repository (https://github.com/AlexsLemonade/OpenPBTA-manuscript) with ManuBot [6,7] integration to enable manuscript creation using Markdown within GitHub. We maintained a data release folder on Amazon S3 containing merged files for each analysis, downloadable from the GitHub repository or open access CAVATICA project.

Stub for the process for contributing analytical code (**Figure 5** and **STAR Methods**) and to the manuscript. Also add a quick note about consensus SNV and CNV to reference **Figure 6A-F** here.

Kids First
PEDIATRIC RESEARCH PROGRAM
Gabriella Miller
Data Resource Center
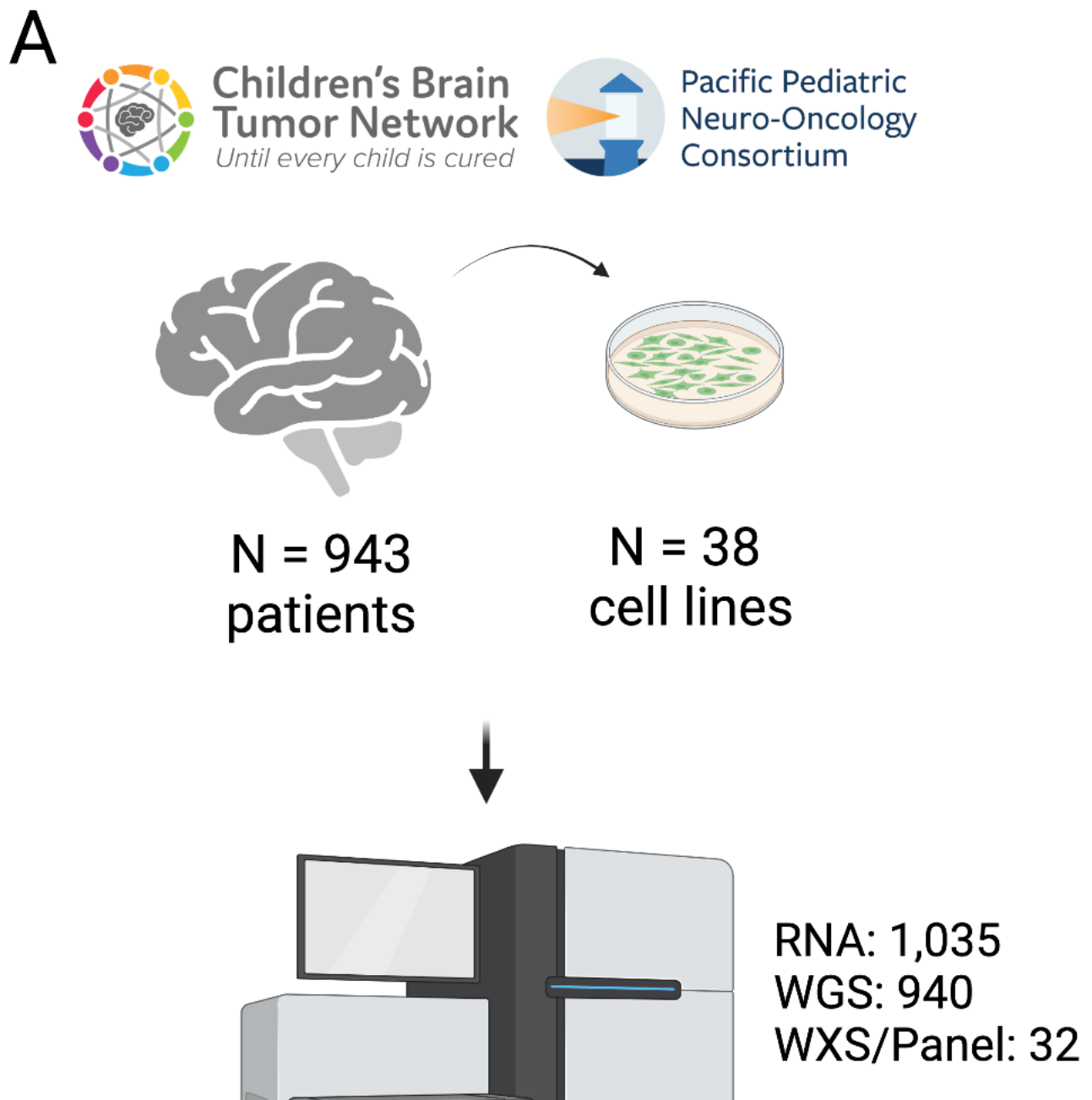
CAVATICA

Amazon S3

Molecular Subtyping of OpenPBTA CNS Tumors

Over the past two decades, together with the World Health Organization (WHO), experts in neurooncology have iteratively redefined the classifications of central nervous system (CNS) tumors [**???** 10.1093/jnen/61.3.215,8]. More recently, in 2016 and 2021 [9,10], molecular subtypes have been integrated into these entities. The Children's Brain Tumor Tissue Consortium (CBTTC), currently the Children's Brain Tumor Network (CBTN), opened its protocol for brain tumor and matched normal sample collection in 2011 and as such, the majority of the samples within the OpenPBTA dataset lack molecular subtype annotation. In the absence of methylation arrays for tumor classification, we utilized key genomic features of entities described by the WHO in 2016, as well as Ryall and colleagues [11], coupled with clinician and pathologist review to subtype 64% (1,281/2,007) of tumor biospecimens with high confidence (**Supplemental Table S1**). Importantly, this collaborative molecular subtyping process allowed us to identify data entry errors (e.g., an ETMR entered as a medulloblastoma) and mis-identified specimens (e.g., Ewing sarcoma sample labeled as a craniopharyngioma), update diagnoses using modern terms (e.g., 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** {#tbl:subtypes} lists the subtypes we defined within the OpenPBTA, comprising of LGATs (N = 476), high-grade gliomas (N = 314), glialneuronal tumors (N = 14), medulloblastomas (N = 190), chordomas (N = 12), ependymomas (N = 65), other embryonal (N = 39), craniopharyngiomas (N = 51), neurocytoma (N = 6), gangliogliomas (N = 93), and Ewing sarcoma (N = 21). For methods, see **STAR Methods** and **Supplemental Figure 5S1**.

Broad Histology	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
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

Broad Histology	Molecular Subtype	N
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
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. {#tbl:subtypes}

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 (**Figures 5** and **6A-F**), we observed lower expected tumor mutation burden (TMB) **Figure 6G** in pediatric tumors compared to adult brain tumors from The Cancer Genome Atlas (TCGA), with hypermutant (> 10 Mut/Mb) and ultra-hypermutant (> 100 Mut/Mb) tumors only found within HGATs.

Low-grade astrocytic tumors

Figure 1A depicts an oncoprint of driver genes for 227 primary low-grade astrocytic tumors. As expected, the majority (62%, 140/227) of these tumors harbor a somatic alteration in *BRAF*, with canonical *BRAF::KIAA1549* fusions as the major oncogenic driver. 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 [11,12]. Indeed, we observed significant upregulation (ANOVA p < 0.01) of the KRAS signaling pathway in LGATs (**Figure 3B**).

Embryonal tumors

Figure 1B shows the mutational landscape for 128 primary embryonal tumors. The majority (N = 95) are medulloblastomas and span the spectrum of molecular subtypes: WNT, SHH, Group3, and Group 4 (see **Molecular Subtyping of CNS Tumors**), with their 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) [13,14,15,16].

Diffuse astrocytic and oligodendroglial tumors (N = 61)

In **Figure 1C**, we show genomic alterations in diffuse midline gliomas (DMGs, N = 34) and non-midline high-grade gliomas (N = 26) biopsied at diagnosis. The single oligodendroglioma sample in the OpenPBTA does not contain mutations in the genes shown and is therefore not present in this oncoprint. Across HGATs, we found *TP53* (57%, 35/61) and *H3F3A* (52%, 32/61) to be the most mutated and co-occurring genes (**Figure 1A**), 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 [5,17,18,19]. Gene set enrichment analysis showed upregulation (ANOVA p < 0.01) of DNA repair, G2M checkpoint, and MYC pathways as well as downregulation of the TP53 pathway (**Figure 3B**). The two tumors with ultra-high tumor mutation burden (TMB) (> 100 Mutations/Mb) were from patients with known mismatch repair deficiency syndrome [3].

Other CNS tumors

Figure 1D depicts an oncoprint for the remaining primary CNS tumors (N = 195). We observed 8% (16/195) of tumors to be *C11orf95::RELA* fusion positive ependymomas and 12% (23/195) to be adamantinomatous craniopharyngiomas driven by mutations in *CTNNB1*. Multiple cancer types contained somatic mutations or fusions in *NF2* (7%, comprised of ependymomas, desmoplastic infantile astrocytoma and ganglioglioma, ..., and ...) and *BRAF* (4%), in addition to rare ependymoma fusions (< 4% of all embryonal tumors) in *YAP1, KRAS, ERBB4,* and *MYB*. The majority of dysembryoplastic tumors harbored alterations in *FGFR1*, including *FGFR1::TACC1* fusions.

Figure 1: 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 = 131): medulloblastomas (N = 97), atypical teratoid rhabdoid tumors (N = 24), embryonal tumors with multilayer rosettes (N = 2), other CNS embryonal tumors (N = 6), 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), oligodendroglioma (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 (N < 5 each).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.

Mutational co-occurrence and signatures highlight key oncogenic drivers

The top 50 mutated genes in primary tumors are shown **Figure 2** by tumor type (**A**, bar plots), with co-occurrence scores illustrated in the heatmap (**B**). We observed *TP53* to be the most frequently mutated gene across OpenPBTA tumors (8.4%, 56/666), significantly co-occurring with *H3F3A* (25/71 or 35.2% of tumors with a mutation in either gene, OR = 32, q = 8.46e-17), *ATRX* (14/66 or 21.2%, OR = 20, q = 4.43e-8), *NF1* (10/71 or 14.1%, OR = 8.62, q = 5.45e-5), and *EGFR* (6/60 or 10%, OR = 18.2, q = 1.6e-4). Other canonical cancer driver genes were frequently mutated: *BRAF*, *H3F3A*, *CTNNB1*, *NF1*, *ATRX*, *FGFR1*, and *PIK3CA*. Although LGG and embryonal tumors make up the majority of tumor types within the OpenPBTA, most of the significant gene interactions stem from HGATs (N = 847/872). At the broad histology level, *CTNNB1* significantly co-occurs with *TP53* (5/14 or 35.7%, OR = 42.9%, q = 1.63e-3) and *DDX3X* (5/16 or 31.3%, OR = 21.1, q = 4.46e-3) in embryonal tumors, *FGFR1* and *PIK3CA* significantly co-occur in LGGs (3/9 or 33.3%, OR = 76.1, q = 3.26e-3), consistent with previous findings [20; 10.1186/s40478-020-01027-z]. *TP53* and *PPM1D* mutations have been shown to be mutually exclusive in HGATs, and our data recapitulates that trend (52/54 or 96.3% of tumors have a mutation in either gene, OR = 0.188, p = 4.13e-2, q = 5.87e-2) [21].
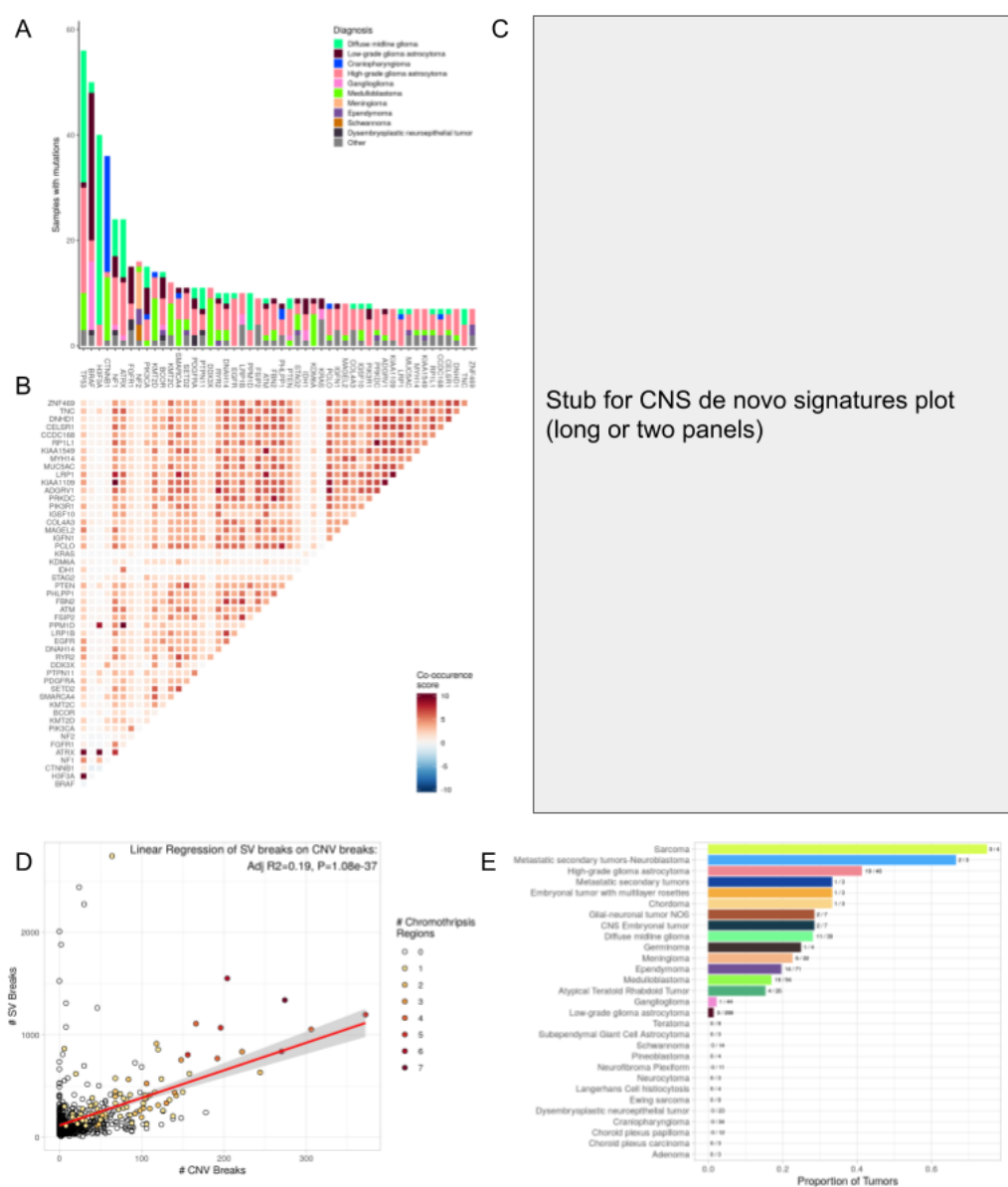
Figure 2: **Figure 3. Mutational co-occurrence and signatures highlight key oncogenic drivers.** A, Bar plot of occurrence and co-occurrence of nonsynonymous mutations for the 50 most commonly mutated genes across all tumor types (annotated from `cancer_group` if N >= 10 or `Other` if N < 10); B, Co-occurrence 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, *De novo* CNS signatures; D, The number of SV breaks significantly correlate with CNV breaks (Adjusted R = 0.436, p = 1.08e-37). E, Chromothripsis frequency across pediatric brain tumors shown by `cancer_group` with N >= 3.

Genomic instability of pediatric brain tumors

We developed a method for generating consensus copy number altered regions from MantaSV, CNVkit, and ControlFREEC (**STAR Methods**). This resulted in high-confidence gains, losses, amplifications, and deep deletion calls used as input for breakpoint density (**Table S2**) and chromothripsis analysis. We observed that HGATs, followed by medulloblastomas, have the most unstable genomes (**Figure S3A**). By contrast, craniopharyngiomas generally lack somatic copy number variation. These patterns of copy number variation largely align with estimates of tumor mutational burden (**Figure S2G**). The number of SV and CNV breakpoints were significantly correlated across tumors (p = 1.08e-37) (**Figure 2D**) and as expected, the number of chromothripsis regions called increases as breakpoint density increases (**Figure S3B-C**). Chromothripsis events were observed in 41% (N = 19/46) of non-midline high-grade gliomas and 28.2% (N = 11/39) of DMGs (**Figure 2E**). We also found evidence of chromothripsis in over 15% of embryonal tumors, ependymomas, meningiomas, germinomas, glial-neuronal tumors, chordomas, metastatic secondary tumors, and sarcomas, highlighting the genomic instability and complexity of pediatric brain tumors.

Transcriptomic Landscape of Pediatric Brain Tumors

Histologic and oncogenic pathway clustering

UMAP visualization of gene expression variation across brain tumors (**Figure 3A**) 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 8A**), as expected. Additionally, *C11orf95::RELA* fusion positive ependymomas formed distinct clusters (**Figure 8B**). However, within low-grade (**Figure 8C** and high-grade (**Figure 8D** astrocytic tumors, tumors were dispersed by molecular subtype across their respective clusters.

We performed gene set variant analysis (GSVA) for Hallmark cancer gene sets, demonstrating activation of underlying oncogenic pathways (**Figure 3B**).

Figure 3: Figure 4. Transcriptomic overview A, First two dimensions from UMAP of sample transcriptome data. Points are colored by `cancer_group` of the samples they represent. B, Heatmap of GSVA scores for Hallmark gene sets with significant differences, with samples ordered by `cancer_group`. C,

Prediction of *TP53* oncogenicity and telomerase activity

To understand the *TP53* phenotype in each tumor, we ran a classifier previously trained on TCGA [22] 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 9A-B**), potentially due to the low number of *TP53* altered (N = 7) 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 [23]. We annotated *TP53* alterations as "activated" if samples harbored one of p.R273C or p.R248W mutations [24], "loss" if the patient had a Li Fraumeni Syndrome (LFS) predisposition diagnosis or if any other SNV, CNV, SV, or fusion fell within a functional domain. 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". Interestingly, we observed that samples annotated as either "activated" or "loss" had significantly higher *TP53* scores than those annotated as "other" (**Figure 4B**, p_{adj} loss vs. other < 2e-16, p_{adj} activated vs. other = 4.0e-5), suggesting that the classifier detects an oncogenic, or altered, *TP53* phenotype (scores > 0.5) rather than solely *TP53* inactivation, as interpreted previously [22]. 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 plotted *TP53* scores for patients with LFS **Figure 4C**, hypothesizing all of these tumors would have high scores. Indeed, we observed significantly higher scores (Wilcoxon p = 4.4e-2) in LFS tumors for which we detected high-confidence *TP53* somatic alterations. Of note, we were unable to detect canonical somatic *TP53* mutations in two LFS patient tumors, annotated as "other" in **Figure 4D**. We confirmed the LFS diagnosis from pathology reports and found each to have a germline variant in *TP53*, however, the tumor purity of these 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. Additionally, we found a significant positive correlation between *TP53* scores and CNV breakpoint density (**Figure 4E**), consistent with a previous report [23]. Tumors with the highest median *TP53* scores were those known to harbor somatic *TP53* alterations: choroid plexus tumors, embryonal tumors, HGATs, and pineal tumors (**Figure 4F**), while melanocytic tumors, meningiomas, and tumors of cranial and paraspinal nerves had the lowest scores.

We next used gene expression data to predict telomerase activity using EXpression-based Telomerase ENzymatic activity Detection (EXTEND) [25] as a surrogate measure of malignant potential [25; 10.1093/carcin/bgp268]. EXTEND scores significantly correlated with *TERT* and *TERC* expression (**Figure 9C-D**).

We found aggressive tumors such as CNS lymphoma, ETMR, ATRT, DMG, and HGG had high EXTEND scores (**Figure 4G**), while benign lesions such as teratomas, dysplasias, and hemangioblastomas had the lowest scores (**Table S3**).
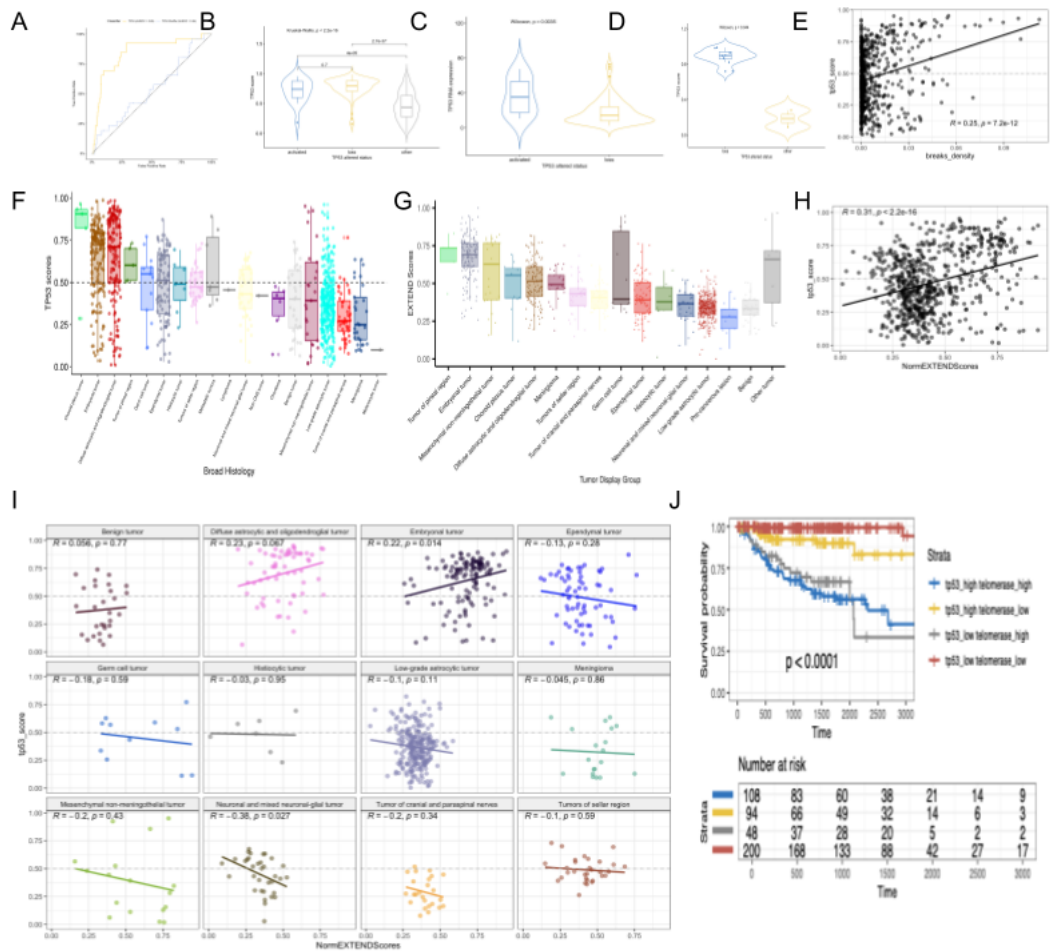
Figure 4: Figure 5. *TP53* **and telomerase activity** A, Receiver Operating Characteristic for *TP53* classifier run on FPKM of stranded RNA-Seq samples. B, Violin and box plots of *TP53* scores plotted by *TP53* alteration type. C, Violin and box plots of *TP53* RNA expression plotted by *TP53* activation status. D, Violin and box plots of *TP53* scores in Li Fraumeni patient tumors. E, *TP53* scores significantly correlate with CNV breakpoint density across PBTA tumors (Pearson R = 0.25, p = 7.5e-12). F, Box plots of *TP53* scores grouped by `broad_histology` . G, Box plots of EXTEND scores grouped by `broad_histology` . H, *TP53* scores significantly correlate with EXTEND scores across PBTA tumors (Pearson R = 0.31, p <2.2e-16). I, Histology-specific Pearson correlations between *TP53* and EXTEND scores. J, Kaplan-Meier

Discussion

Stub in discussion section

Acknowledgments

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
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
Bailey Farrow	Data curation, Software
Nicolas Van Kuren	Data curation, Software
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	Supervison
Derek Hanson	Validation
Hongbo M. Xie	Methodology, Supervision
Siyuan Zheng	Formal analysis, Visualization, Writing - Original draft, Supervision
Cassie N. Kline	Supervision
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

Figure Titles and Legends

Tables with Titles and Legends

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 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.sbgenomics.com/u/cavatica/openpbta or via download script from https://github.com/AlexsLemonade/OpenPBTA-analysis/. Summary data are visible within Pedcbioportal at https://pedcbioportal.kidsfirstdrc.org/study/summary?id=openpbta. The DOIs are listed in the *Key Resources Table*.

All original code has been deposited in the following repositories and is publicly available as of the date of the publication: - Primary data analyses: https://github.com/d3b-center/OpenPBTA-workflows/ - Downstream data analyses: https://github.com/AlexsLemonade/OpenPBTA-analysis/ - Manuscript code: https://github.com/AlexsLemonade/OpenPBTA-manuscript The DOIs are listed in 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 [4] is a collaborative, multi-institutional (18 institutions worldwide) research program dedicated to the study of childhood brain tumors. The Pacific Pediatric Neuro-Oncology Consortium (PNOC) [26] is an international consortium dedicated to bringing new therapies to children and young adults with brain tumors. PNOC collected blood and tumor biospecimens from newly-diagnosed diffuse intrinsic pontine glioma (DIPG) patients as part of the clinical trial PNOC003/NCT02274987 [5].

All CBTN data can be download from the Gabriella Miller Kids First Data Resource Center, [27]. The de-identified patient's blood and tumor tissue were prospectively collected by the consortium from patients enrolled within the CBTN.

The cell lines were generated by the CBTN from either fresh tumor tissue obtained directly from surgery performed at Children's Hospital of Philadelphia (CHOP) or from prospectively collected tumor specimens stored in Recover Cell Culture Freezing media (cat# 12648010, Gibco). The tissue was dissociated using enzymatic method with papain as described [28]. Briefly, tissue was washed with HBSS (cat# 14175095, Gibco), minced and incubated with activated papain solution (cat# LS003124, SciQuest) for up to 45 minutes. The papain was inactivated using ovomucoid solution (cat# 542000, SciQuest), tissue was briefly treated with DNase (cat# 10104159001, Sigma) and passed through the 100µm cell strainer (cat# 542000, Greiner Bio-One). Two cell culture conditions were initiated based on the number of cells available. For cultures utilizing the fetal bovine serum (FBS), a minimum density of 3×10^5 cells/ml were plated 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 the serum-free media conditions cells were plated at minimum density of 1×10^6 cells/ml in DMEM/F12 media 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), 2.5µg/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 7 but no QC thresholds were implemented for the DNA. For library preparation, 500ng of nucleic acids were used as input for RNA-Seq, WXS, and targeted DNA panel (panel). The RNA prep was performed using the TruSeq RNA Sample Prep Kit (Illumina, #FC-122-1001) and the exome prep was performed using KAPA Library Preparation Kit (Kapa Biosystems, #KK8201) using Agilent's SureSelect Human All Exon V5 backbone with custom probes. The targeted DNA panel developed by Ashion (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 2×10^6 cells pellet was used for extractions. Tissues were lysed using a Qiagen TissueLyser II (Qiagen) with 2×30 sec at 18Hz settings using 5 mm steel beads (cat# 69989, Qiagen). Both tissue and cell pellets processes included a CHCl3 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 QIAsymphony automated platform (Qiagen) using the QIAsymphony 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). Library preparation and sequencing was performed by the NantHealth sequencing center. Briefly, DNA sequencing libraries were prepared for tumor and matched-normal DNA using the KAPA HyperPrep kit (cat# KK8541, Roche); tumor RNA-Seq libraries were prepared using KAPA Stranded RNA-Seq with RiboErase kit (cat# KK8484, Roche). Whole genome sequencing (WGS) was performed at an average depth of coverage of 60X for tumor samples and 30X for germline. The panel tumor sample was sequenced to 470X and the normal panel sample was sequenced to 308X. RNA samples were sequenced to an average of 200M reads. All samples were sequenced on the Illumina HiSeq platform (X/400) (Illumina) with 2 × 150bp read length.

Data generation

NantHealth Sequencing Center (Culver City, CA) performed whole genome sequencing (WGS) on all paired tumor (~60X) and constitutive (~30X) DNA samples. WGS libraries were 2x150 bp and sequenced on an Illumina X/400. NantHealth Sequencing Center performed ribosomal-depleted whole transcriptome stranded RNA-Seq to an average depth of 100M reads for CBTN tumor samples. 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. PNOC WXS and RNA-Seq libraries 2x100 bp and sequenced on an Illumina HiSeq 2500.

DNA WGS Alignment

We used BWA-MEM [29] v0.7.17 for alignment of paired-end DNA-seq reads. We used version 38, patch release 12 of the *Homo sapiens* genome as our alignment reference, which we obtained as a FASTA file from UCSC [30]. Alignments were further processed using following the Broad Institute's Best Practices [31] for processing Binary Alignment/Map files (BAMs) in preparation for variant discovery. Duplicates were marked using SAMBLASTER [32] v0.1.24, BAMs merged and sorted using Sambamba [33] v0.6.3. Resultant BAMs were processing using Broad's Genome Analysis Tool Kit [GATK] (https://software.broadinstitute.org/gatk/) v4.0.3.0, BaseRecalibrator submodule. Lastly, for normal/germline input, we run the GATK HaplotypeCaller [34] submodule on the recalibrated BAM, generating 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. References can be obtained from the Broad Genome References on AWS bucket, with a general description of references here [35].

Quality Control of Sequencing Data

NGSCheckmate [36] was performed on matched tumor/normal CRAM files to confirm sample matches and remove mismatched samples from the dataset. CRAM inputs were preprocessed using BCFtools to filter and call 20k common single nucleotide polymorphisms (SNPs) using default parameters[37] and the resulting VCFs were used to run NGSCheckmate using this workflow in the D3b GitHub repository. Per author guidelines, <= 0.61 was used as a correlation coefficient cutoff at sequencing depths >10 to predict mismatched samples. For RNA-Seq, read strandedness was determined by running the `infer_experiment.py` script on the first 200k mapped reads. If calculated strandedness did not match strandedness information received from the sequencing center, samples were removed from analysis. We required at least 60% of RNA-Seq reads mapped to the human reference or samples were removed from analysis. MEND QC [38] was performed on aligned RNA-Seq reads using this workflow to identify mapped exonic non-duplicate reads.

Germline Variant Calling

SNP calling for B-allele Frequency (BAF) generation

Germline haplotype calls were performed following the GATK Joint Genotyping Workflow, except the workflow was run on an individual sample basis. This workflow was applied to the GVCF output from the alignment workflow on normal/germline samples. 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. This filtered VCF was used as input to Control-FREEC and CNVkit (below) for generation of BAF files. GATK v4.0.12.0 was used for all steps except `VariantFiltration`, which used 3.8.0 because as of GATK 4.0.12.0, this tool was beta and known to be unreliable for this purpose. This single-sample workflow can be found in the Kids First GitHub repository. References can be obtained from the Broad Genome References on AWS bucket, with a general description of references here [35].

Somatic Mutation Calling

SNV and indel calling

For PBTA samples, we used four variant callers to call SNVs and indels from targeted DNA panel, WXS, and WGS data: Strelka2 [39], Mutect2 [40], Lancet [41], and VarDict [42]. WXS samples from TCGA were run using Strelka2, Mutect2 and Lancet. TCGA samples were captured using different WXS target capture kits and all the BED files were downloaded from `GDC_portal`. The input interval BED files for both panel and WXS data for PBTA samples were provided by the manufacturers. For both PBTA and TCGA, all panel and WXS BED files were padded by 100 bp on each side during Strelka2, Mutect2, and VarDict runs and 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. Strelka2 [39] v2.9.3 was run using default parameters for canonical chromosomes (chr1-22, X,Y,M), as recommended by the authors. The final Strelka2 VCF was filtered for PASS variants. Mutect2 from GATK v4.1.1.0 was run following Broad best practices outlined from their Workflow Description Language (WDL) [43]. The final Mutect2 VCF was filtered for PASS variants. To manage memory issues, VarDictJava [42] v1.58 [44] was run using 20Kb 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) >= 0.05 (instead of >= 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. The final VCF was filtered for PASS variants with TYPE=StronglySomatic. Lancet v1.0.7 was run using default parameters, except for those noted below. For input intervals to Lancet WGS, a reference BED was created by using 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 [45]. These intervals were then padded by 300 bp on each side during Lancet variant calling. Per recommendations by the New York Genome Center [45], for WGS samples, the Lancet input intervals described above were augmented with PASS variant calls from Strelka2 and Mutect2 as validation.

VCF annotation and MAF creation

Normalization of INDELs using `bcftools norm` [46] was performed on all PASS VCFs using the following subworkflow [47], release v3. The ENSEMBL Variant Effect Predictor [48], reference release 93, was used to annotate variants and bcftools was used to add population allele frequency (AF) from gnomAD. SNV and INDEL hotspots from v2 of MSKCC's database [49] plus the C228T and C250T TERT promoter mutations [50] were annotated. SNVs were annotated by matching amino acid position (`Protein_position` column in MAF file) with SNVs in the MSKCC database, splice sites were matched

to `HGVSp_Short` values in the MSKCC database, and INDELs were matched based on amino acid present within the range of INDEL hotspots values in the MSKCC database. Non-hotspot annotated variants with a normal depth of <= 7 and/or gnomAD AF > 0.001 were removed as potential germline variants. TERT promoter mutations were matched using hg38 coordinates from [50]: C228T occurs at 5:1295113, is annotated as existing variant `s1242535815`, `COSM1716563`, `COSM1716558`, and is 66bp away from TSS and C250T occurs at Chr5:1295135, is annotated as existing variant `COSM1716559`, and is 88 bp away from TSS.

The final set of variants were retained if annotated as PASS or HotSpotAllele=1. MAFs were created using MSKCC's vcf2maf [51] v1.6.17.

Gather SNV and INDEL Hotspots

All variant calls from Strelka2, Mutect2, or Lancet that overlap with an SNV or INDEL hotspot from v2 of MSKCC's database [49] or the C228T and C250T TERT promoter mutations [50] were retained in a hotspot-specific MAF file, which was used for select analyses as described in the methods below. VarDict-only calls were not retained since ~ 39M calls with low VAF were uniquely called and may be potential false positives.

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 with Strelka2. We examined the variant allele frequencies produced by each caller and compared their overlap with each other [52]. VarDict calls included many variants that were not identified by other callers [53], 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 [54]. We termed mutations shared among the other three callers (Strelka2, Mutect2, and Lancet) to be consensus mutation calls and dropped VarDict due to concerns about it calling a large number of false positives. In practice, because our filtered set was based on the intersection of these three sets and because VarDict called nearly every mutation from the other three callers plus many that were unique to it, the decision to not consider VarDict calls has little impact on the results.

For some downstream analyses, only coding sequence SNVs (based on GENCODE v27 [55]) are used, to enhance comparability to other studies. 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. This definition of *effectively surveyed* base pairs is what is used to calculate effective genome size for calculations for tumor mutation burden and mutational signatures.

Recurrently mutated genes and co-occurrence of gene mutations

Using the consensus SNV calls, we identified genes that were recurrently mutated in the cohort, including nonsynonymous mutations with a variant allele frequency greater than 5% among the set of independent samples. The set of nonsynonymous mutations was determined using ENSEMBL Variant Effect Predictor [48] annotations, including High and Moderate consequence types as defined in `maftools` [56]. 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 the $I \times -\log_{10}(P)$ 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.

Somatic Copy Number Variant Calling (WGS samples only)

We used Control-FREEC [57,58] v11.6 and CNVkit [59] v0.9.3 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 [60] 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 [61] v.2.0.23 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 OpenPBTA Analysis repository.

Consensus CNV Calling

For each caller and sample, CNVs were called based on consensus among Control-FREEC [57,58], CNVkit [59], and Manta [62]. CNVs called significant by Control-FREEC (p-value < 0.01) and Manta calls that passed all filters [63] were included in consensus calling. Sample and caller combination files with more than 2500 CNVs called were removed from the set; we expect these to be noisy and poor quality samples based on cutoffs used in GISTIC [61]. For each sample, the following regions are included in the final consensus set: 1) regions with reciprocal overlap of at least 50% between two of the three callers; 2) smaller CNV regions that are at least 90% covered by another caller. Any copy number alteration that was not called by two or more callers was not included in the consensus file. For the samples that are included in the consensus file, if a certain region has a neutral call, copy number of `NA` is defined for that region. CNV regions within 10,000 bp of each other with the same direction of gain or loss were merged into single region. We filtered out any CNVs that overlapped 50% or more with immunoglobulin, telomeric, centromeric, segment duplicated regions or were shorter than 3000 bp.

Focal Copy Number Calling

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 [64,65] to add cytoband status using the UCSC cytoband file [66,67]. 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. The weighted means of each status call fraction were calculated using band length. We used the weighted means to define the dominant status at the chromosome arm-level.

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

We also wanted 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. For the determination of the most focal units, we first considered the dominant

status calls at the chromosome arm-level. If the chromosome arm dominant status was not clearly defined as a gain or loss (and was callable) we looked to include the cytoband-level status call. Similarly, if a cytoband dominant status call was not clearly defined as a gain or loss (and was callable) we looked to include the gene-level status call. To obtain the gene-level data, we used the mergeByOverlaps function [68] from the IRanges package [69] to find overlaps between the segments in the consensus CNV file and the exons in the GENCODE v27 annotation file [55]. 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 is greater than two times the ploidy value.

Somatic Structural Variant Calling (WGS samples only)

We used Manta SV [62] v1.4.0 for structural variant (SV) calls. Manta SV calling was also limited to regions used in Strelka2. The hg38 reference for SV calling used was limited to canonical chromosome regions. The somatic DNA workflow for SNV, indel, copy number, and SV calling can be found in the KidsFirst Github repository. Manta SV output was annotated using AnnotSV v2.1 [70] and the workflow can be found in the D3b GitHub repository.

Chromothripsis Analysis (WGS samples only)

Candidate chromothripsis regions were identified in the set of independent tumor WGS samples with ShatterSeek [71], using Manta SV calls that passed all filters and consensus CNV calls. Only chromosomes 1-22 and X were considered. The consensus CNV data were modified to fit ShatterSeek input requirements: 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. Candidate chromothripsis regions were classified as high- or low-confidence by applying the statistical criteria described by the ShatterSeek authors.

Gene Expression

Abundance Estimation

We used STAR [72] v2.6.1d to align paired-end RNA-seq reads. This output was used for all subsequent RNA analysis. We used Ensembl GENCODE 27 [55], "Comprehensive gene annotation" as a reference. We used RSEM [73] v1.3.1 for both FPKM and TPM transcript- and gene-level quantification. We also added a second method of quantification using kallisto [74] v0.43.1. This method differs in that it uses pseudoaligments using FASTQ reads directly to the aforementioned GENCODE 27 reference.

Gene Expression Matrices with Unique HUGO Symbols

Algorithms that perform gene set enrichment, molecular subtyping, or immune-profiling, for example, require an RNA-seq gene expression matrix as input, with HUGO gene symbols as row names and sample names as column names. There is a small proportion of gene symbols that map to multiple Ensembl gene identifiers (in GENCODE v27, 212 gene symbols map to 1866 Ensembl gene identifiers), termed multi-mapped gene symbols.

We first removed genes with no expression from the RSEM abundance data using a cut-off of FPKM > 0 in at least 1 sample across the PBTA cohort. We computed the mean FPKM across all samples per gene and for each multi-mapped gene symbol, we chose the Ensembl identifier corresponding to the maximum mean FPKM with the goal of choosing the identifier that best represented the expression of the gene. After collapsing gene identifiers, there were a total of 46,400 unique expressed genes in the poly-A dataset and a total of 53,011 unique expressed genes remaining in the stranded dataset. More detail can be found in the `collapse-rnaseq` analysis module.

Immune Profiling/Deconvolution

We used the R package immunedeconv [75,76] to deconvolute, quantify and compare various immune cell types across 21 histologies from the PBTA cohort with xCell [77] and CIBERSORT [78] in the stranded and poly-A collapsed FPKM RNA-seq datasets (`immune-deconv` analysis module). Both methods allow between samples (inter-sample), between cell types (intra-sample) and between cancer type (inter-histology) comparisons.

Gene Set Variation Analysis

We performed Gene Set Variation Analysis (GSVA) [79] on collapsed, log2-transformed RSEM FPKM data using the GSVA Bioconductor package [80] with setting `mx.diff=TRUE` to obtain Gaussian-distributed scores (`gene-set-enrichment-analysis` analysis module) for each of the MSigDB hallmark gene sets [81]. We compared GSVA scores among histology groups (`short_histology`) using ANOVA and subsequent Tukey tests; p-values were Bonferroni-corrected for multiple hypothesis testing.

Dimension reduction

We applied Uniform Manifold Approximation and Projection (UMAP) [82] to log2-transformed FPKM data using the umap R package [83]. We set the number of neighbors to 15 (`transcriptomic-dimension-reduction` analysis module).

RNA Fusion Calling and Prioritization

Gene fusion detection

We set up Arriba v1.1.0 and STAR-Fusion 1.5.0 [84] fusion detection tools using CWL on CAVATICA. For both 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 GRCh38_v27_CTAT_lib_Feb092018.plug-n-play.tar.gz provided in the STAR-fusion release. For Arriba, we used a blacklist file (blacklist_hg38_GRCh38_2018-11-04.tsv.gz) from the Arriba release tarballs to remove recurrent fusion artifacts and transcripts present in healthy tissue. We also provided Arriba with strandedness information or set it to auto-detection for poly-A samples. We used FusionAnnotator on Arriba fusion calls in order to harmonize annotations with those of STAR-Fusion. The RNA expression and fusion workflows can be found in the KidsFirst GitHub repository and the FusionAnnotator workflow found in the D3b GitHub repository.

Fusion prioritization

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 a minimum of 1 junction reads and at least one gene partner expressed (TPM > 1) to be true calls. If a fusion call had 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` (>2 samples) or the fusion was specific to the `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 [[85](#)], [COSMIC genes](#), and/or known [TCGA fusions](#) from curated [references](#). *MYBL1* [[86](#)], *SNCAIP* [[87](#)], *FOXR2* [[88](#)], *TTYH1* [[89](#)], and *TERT* [[90](#),[91](#),[92](#),[93](#)] were added to the oncogene list and *BCOR* [[88](#)] and *QKI* [[94](#)] were added to the tumor suppressor gene list based on pediatric cancer literature review. The fusion filtering workflow can be found in the [OpenPBTA Analysis repository](#).

Mutational Signatures

We obtained weights for signature sets by applying deconstructSigs [[95](#),[96](#)] to consensus SNVs with the BSgenome.Hsapiens.UCSC.hg38 annotations [[97](#)]. We estimated how many mutations contributed to each signature for each sample using each sample's signature weights. Weights for signatures fall in the range zero to one inclusive. For a given sample and signature combination, we estimated the number of contributing mutations per Mb of the genome by multiplying the signature weight by the total number of trinucleotide mutations identified by deconstructSigs and then dividing by the size of the effectively surveyed genome.

$$\frac{\text{\# of contributing mutations}}{Mb} = \frac{\text{weight} * \sum \text{\# Trinucleotide mutations}}{\text{Size in Mb of effectively surveyed genome}}$$

These results do not include signatures with small contributions; deconstructSigs drops signature weights that are less than 6% [[95](#)]. We used these methods to calculate signature scores for each sample with both COSMIC [[98](#)] and Alexandrov et al, 2013 [[99](#)] signature sets.

PBTA Tumor Mutation Burden

We consider tumor mutation burden (TMB) to be the number of consensus SNVs per *effectively surveyed* base of the genome.

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

We used the total number coding sequence consensus SNVs for the numerator and the size of the intersection of the regions considered by Lancet, Strelka2, and Mutect2 with coding regions (CDS from GENCODE v27 annotation [[55](#)]) as the denominator.

TCGA Tumor Mutation Burden

We calculated tumor mutation burden in TCGA using MC3 mutation calls [[100](#)] for TCGA brain-related tumor projects including: LGG (lower-grade glioma) [[101](#)], GBM (glioblastoma multiforme) [[102](#)], and PCPG (pheochromocytoma and paraganglioma) [[103](#)]. The MC3 project provided an exome BED file. All SNVs fell within these regions. We considered the regions covered by the MC3 BED file (based on GENCODE v19 annotation [[104](#)]) to have been effectively surveyed.

Clinical Data Harmonization

WHO Classification of Disease Types

For the CBTN cohort, the `pathology_diagnosis` field in the `pbta-histologies.tsv` file contains a harmonized diagnosis based on the patient's pathology report. For the PNOC003 cohort, the `pathology_diagnosis` field in the `pbta-histologies.tsv` file contains the diagnosis gathered from the patient's pathology report. For the CBTN cohort, the `pathology_free_text_diagnosis` field in the `pbta-histologies.tsv` file contains additional free text diagnosis information gathered from the patient's pathology report. The `integrated_diagnosis` field in the `pbta-histologies.tsv` file is the standardized 2016 WHO diagnosis [[9](#)] based on `pathology_diagnosis`, molecular subtyping, and in some cases, additional pathology review. The `broad_histology` denotes the broad 2016 WHO classification for each tumor. The `short_histology` is an abbreviated version of either the `broad_histology` or `integrated_diagnosis` for plotting purposes. The `CNS_region` was subtyped into hemispheric, midline, mixed, or other based on specimen location (see table below).

Clinical and Histology Metadata	Definition	Possible values
age_at_diagnosis_days	Patient age at diagnosis in days	numeric
age_last_update_days	Patient age at the last clinical event/update in days	numeric
aliquot_id	External aliquot identifier	variable
broad_composition	Broad classification of sample type	cell-line;cyst;non-tumor;tumor
broad_histology	Broad WHO 2016 classification of cancer type	text
cancer_predispositions	Reported cancer predisposition syndromes	text
cohort	Scientific cohort	CBTN;PNOC
cohort_participant_id	Scientific cohort participant ID	C#####-C######

Clinical and Histology Metadata	Definition	Possible values
composition	Sample composition	Derived Cell Line;Not Reported;Peripheral Whole Blood;Saliva;Solid Tissue
ethnicity	Patient reported ethnicity	text
experimental_strategy	Sequencing strategy	WGS;WXS;RNA-Seq;Panel
germline_sex_estimate	Predicted sex of patient based on germline X and Y ratio calculation (described in methods)	Female;Male;Unknown
CNS_region	Brain region based on `primary_site`	
integrated_diagnosis	2016 WHO diagnosis integrated from pathology diagnosis and molecular subtyping	text
Kids_First_Biospecimen_ID	KidsFirst biospecimen identifier	BS_########
Kids_First_Participant_ID	KidsFirst patient identifier	PT_########
molecular_subtype	Molecular subtype defined by WHO 2016 guidelines	text
normal_fraction	Theta2 normal DNA fraction estimate	numeric
Notes	Free text field describing changes from `pathology_diagnosis` to `integrated_diagnosis` or manner in which molecular_subtype was determined	text
OS_days	Overall survival in days	numeric
OS_status	Overall survival status	DECEASED;LIVING
parent_aliquot_id	External identifier combining sample_id, sample_type, aliquot_id, and sequencing_strategy for some samples	text
pathology_diagnosis	Reported and/or harmonized patient diagnosis from pathology reports	text
pathology_free_text_diagnosis	Free text patient diagnosis from pathology reports	text
primary_site	Bodily site(s) from which specimen was derived	text
race	Patient reported race	text
reported_gender	Patient reported gender	text
RNA_library	Type of RNA-Sequencing library preparation	stranded;poly-A
sample_id	External biospecimen identifier	variable
sample_type	Broad sample type	Normal;Tumor
seq_center	Sequencing center	BGI@CHOP Genome Center;Genomic Clinical Core at Sidra Medical and Research Center;NantOmics;TGEN
short_histology	Abbreviated `integrated_diagnosis` or `broad_histology` for plotting purposes	text
tumor_descriptor	Phase of therapy from which tumor was derived	Initial CNS Tumor;Progressive Progressive Disease Post-Mortem;Recurrence;Second Malignancy;Unavailable
tumor_fraction	Theta2 tumor DNA fraction estimate	numeric
tumor_ploidy	Control-FREEC ploidy	numeric

Table S1. Clinical metadata collected for OpenPBTA. {#tbl:S1}

CNS_region	primary_site
Hemispheric	Frontal Lobe,Temporal Lobe,Parietal Lobe,Occipital Lobe
Midline	Pons/Brainstem,Brain Stem- Midbrain/Tectum,Brain Stem- Pons,Brain Stem-Medulla,Thalamus,Basal Ganglia,Hippocampus,Pineal Gland
Spine	Spinal Cord- Cervical,Spinal Cord- Thoracic,Spinal Cord- Lumbar/Thecal Sac,Spine NOS
Ventricles	Ventricles
Posterior fossa	Cerebellum/Posterior Fossa
Optic pathway	Optic Pathway
Suprasellar	Suprasellar/Hypothalamic/Pituitary

CNS_region	primary_site
Other	Meninges/Dura,Other locations NOS,Skull,Cranial Nerves NOS,Brain

Table S2. Harmonized CNS brain regions derived from primary site values. {#tbl:S2}

Molecular Subtyping

The `molecular_subtype` column in the `pbta-histologies.tsv` file 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 [9].

Medulloblastoma (MB) subtypes SHH, MYC, Group 3, and Group 4 were predicted using the consensus of two RNA expression classifiers: [Medulloblastoma Classifier]](https://github.com/d3b-center/medullo-classifier-package) and MM2S Classifier [105] on the RSEM FPKM data.

High-grade glioma (HGG) subtypes were derived using the criteria below (additional details in the analysis README):

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 HGAT, 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 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 [106]. 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 using the criteria below [107,108,109,110]. Additional details can be found in the analysis notebook.

1. Any RNA-seq biospecimen with *LIN28A* overexpression, plus a *TYH1* 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) [111; 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) [89,111].
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 [112]) 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) [**???**]
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 based on the primary site of the tumor [113]. If `primary_site` of the tumor was `Ventricles`, it was subtyped as `CNC`; otherwise, it was subtyped as `EVN`.

Craniopharyngiomas (CRANIO) were subtyped into adamantinomatous (`CRANIO, ADAM`), papillary (`CRANIO, PAP`) or undetermined (`CRANIO, To be classified`) based on the following criteria [114,115]: 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 do 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`.

Low-grade astrocytic tumors (LGAT) or glialneuronal tumors (GNT) were subtyped based on SNV, fusion and CNV status based on 11, 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 [116], 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 [117], 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 LGAT 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 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`.

Ependymoma (EPN) were subtyped 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. [118]. Briefly, fusion, CNV and gene expression data were used to subtype EPN as followed: 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 through `molecular-subtyping-pathology` module. The compilation was executed by the following steps:

Firstly, `broad_histology`, `short_histology`, and `integrated_diagnosis` columns in the result files from the above subtyping modules (i.e., `CRANIO_molecular_subtype.tsv`, `EWS_results.tsv`, `EPN_all_data_withsubgroup.tsv`, `HGG_molecular_subtype.tsv`, `lgat_subtyping.tsv`, `MB_molecular_subtype.tsv`, `embryonal_tumor_molecular_subtypes.tsv`, and `neurocytoma_subtyping.tsv`) were updated based on the molecular subtype of the tumor. Detailed information about the updating procedure were included in the analysis [notebook](#). Notes were also added to indicated that the changes in `broad_histology`, `short_histology` and `integrated_diagnosis` were from OpenPBTA subtyping modules.

Subsequently, `broad_histology`, `short_histology` and `harmonized_diagnosis` columns of tumors with particular pathology diagnosis free text were updated as specified in the following table:

pathology_diagnosis	subtyping module	pathology_free_text_diagnosis	broad_histology	short_histology	harmonized_diagnosis
Primary CNS lymphoma	NA	contains `burkitt's lymphoma`	Lymphoma	CNS lymphoma	CNS Burkitt's lymphoma
Other	NA	contains `xanthogranuloma` or `jxg`	Histiocytic tumor	JXG	Juvenile xanthogranuloma
Meningioma	NA	contains `atypical`	Meningioma	Meningioma	Atypical meningioma
Meningioma	NA	contains `anaplastic`	Meningioma	Meningioma	Anaplastic (malignant) meningioma
Meningioma	NA	contains `clear cell meningioma`	Meningioma	Meningioma	Clear cell meningioma
Meningioma	NA	contains `meningothelial`	Meningioma	Meningioma	Meningothelial meningioma
Meningioma	NA	does not contain `atypical`, `anaplastic`, `clear cell`, or `meningothelial`	Meningioma	Meningioma	Meningioma
Choroid plexus papilloma	NA	contains `atypical`	Choroid plexus tumor	Choroid plexus tumor	Atypical choroid plexus papilloma
Craniopharyngioma	CRANIO	contains `adamantinomatous`	Tumors of sellar region	Craniopharyngioma	Adamantinomatous craniopharyngioma

Similarly, `broad_histology`, `short_histology`, `integrated_diagnosis` and `harmonized_diagnosis` columns of tumors with following pathology diagnosis free text were updated as specified in the table below:

pathology_diagnosis	subtyping module	pathology_free_text_diagnosis	broad_histology	short_histology	integrated_diagnosis
Low-grade glioma/astrocytoma (WHO grade I/II)	LGAT	contains `sega` or `subependymal giant cell astrocytoma`	Low grade astrocytic tumor	LGAT	Subependymal Giant Cell Astrocytoma,
Low-grade glioma/astrocytoma (WHO grade I/II)	LGAT	contains `fibrillary`	Low grade astrocytic tumor	LGAT	Diffuse fibrillary astrocytoma,
Low-grade glioma/astrocytoma (WHO grade I/II)	LGAT	contains `gliomatosis cerebri, type 1, ia`	Low grade astrocytic tumor	LGAT	Gliomatosis cerebri,

pathology_diagnosis	subtyping module	pathology_free_text_diagnosis	broad_histology	short_histology	integrated_diagnosis
Low-grade glioma/astrocytoma (WHO grade I/II)	LGAT	contains `jpa` or `juvenile astrocytoma` or `pilocytic` or `pilocystic` (typo) or `pilomyxoid` but does not contain `fibrillary`	Low grade astrocytic tumor	LGAT	Pilocytic astrocytoma,
Low-grade glioma/astrocytoma (WHO grade I/II)	LGAT	contains `oligodendroglioma who ii`	Diffuse astrocytic and oligodendroglial tumor	Oligodendroglioma	Oligodendroglioma,
Low-grade glioma/astrocytoma (WHO grade I/II)	LGAT	contains `pxa` or `pleomorphic xanthoastrocytoma`	Low grade astrocytic tumor	LGAT	Pleomorphic xanthoastrocytoma,

Additionally, `broad_histology` , `short_histology` , `integrated_diagnosis` and `harmonized_diagnosis` columns of tumors with following pathology diagnosis free text were updated as specified in the table below:

pathology_diagnosis	subtyping module	pathology_free_text_diagnosis	broad_histology	short_histology	integrated_diagnosis	harmonized_diagnosis
Low-grade glioma/astrocytoma (WHO grade I/II)	NA, remove from LGAT module	contains `desmoplastic infantile astrocytoma`	Neuronal and mixed neuronal-glial tumor	GNT	Desmoplastic infantile astrocytoma and ganglioglioma,	Desmoplastic infantile astrocytoma and ganglioglioma
Low-grade glioma/astrocytoma (WHO grade I/II)	NA, remove from LGAT module	contains `diffuse leptomeningeal glioneuronal tumor`	Neuronal and mixed neuronal-glial tumor	GNT	Diffuse leptomeningeal glioneuronal tumor,	Diffuse leptomeningeal glioneuronal tumor
Low-grade glioma/astrocytoma (WHO grade I/II)	NA, remove from LGAT module	contains `glioneuronal`	Neuronal and mixed neuronal-glial tumor	GNT	Glial-neuronal tumor NOS,	Glial-neuronal tumor NOS
Low-grade glioma/astrocytoma (WHO grade I/II)	NA, remove from LGAT module	contains `rosette forming glioneuronal tumor`	Neuronal and mixed neuronal-glial tumor	GNT	Rosette-forming glioneuronal tumor,	Rosette-forming glioneuronal tumor

Notes were also added to indicated that the changes in `broad_histology` , `short_histology` , `integrated_diagnosis` and `harmonized_diagnosis` were from pathology diagnosis free text.

For samples with subtype discrepancies, `molecular_subtype` and `integrated_diagnosis` were updated following pathology or clinical review. Detailed information can be found in the analysis notebooks for clinical and pathology feedback. Finally, the newly compiled subtypes were integrated into the `pbta-histologies.tsv` file in the `molecular-subtyping-integrate` module.

TP53 Alteration Annotation

In addition to tumor types mentioned above, TP53 altered status is also annotated for all samples and if a sample is determined to be either `TP53 loss` or `TP53 activated` , this annotation will be included in the `molecular_subtype` column. We applied a *TP53* inactivation classifier originally trained on TCGA PanCan data [22] to the matched RNA expression data for each sample. Along with the *TP53* classifier scores, consensus SNV and CNV, SV, and references databases that list TP53 hotspot mutations [49,119,120] and functional domains [121] were used collectively to determine TP53 alteration status for each sample. The rules for calling either `TP53 loss` or `TP53 activated` are as follows: If a sample has any of the two well-characterized *TP53* gain-of-function mutations, p.R273C or p.R248W [24], `TP53 activated` status will be assigned. A sample will be annotated as `TP53 loss` if any of the following conditions is met: 1) It contains a *TP53* hotspot mutation as defined by IARC TP53 database [49,119,120] 2) It contains two *TP53* alterations, including SNV, CNV or SV, which is indicative of probable bi-allelic alterations 3) It contains one *TP53* somatic alteration, including SNV, CNV, or SV and a germline *TP53* mutation indicated by the diagnosis of Li-Fraumeni syndrome [122] 4) It contains one germline *TP53* mutation indicated by Li-Fraumeni syndrome and the *TP53* classifier score for matched RNA-Seq is over 0.5.

Survival

Overall survival, denoted `OS_days` , was calculated as days since initial diagnosis.

Prediction of participants' genetic sex

The clinical metadata provided included a reported gender. We used DNA data, in concert with the reported gender, to predict participant genetic sex so that we could identify sexually dimorphic outcomes. This analysis could also reveal samples that may have been contaminated in certain circumstances. We used the idxstats utility from SAMtools [123] 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 reviewed this statistic in the context of reported gender and determined that a threshold of less than 0.2 clearly delineated female samples. Fractions greater than 0.4 were predicted to be males. Samples with values in the range [0.2, 0.4] were marked as unknown. We ran this analysis through [CWL](CWL) on CAVATICA. Resulting calls were added to the clinical metadata as `germline_sex_estimate` .

Selection of independent samples

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

Quantify Telomerase Activity using Gene Expression Data

We predicted telomerase activity of pediatric brain tumor samples using our recently developed method EXTEND. In brief, EXTEND estimates telomerase activity based on the expression of a 13-gene signature. This signature was derived by comparing telomerase positive tumors and tumors with activated alternative lengthening of telomeres pathway, a group presumably negative of telomerase activity. More details about the algorithm can be found in reference [124]. We calculated telomerase activity score for each sample of the PBTA cohort and examined the score distribution across both broad and specific disease histological subtypes. We also compared EXTEND scores across the four molecular subgroups of medulloblastoma (Group3, Group4, SHH and WNT) using a series of pairwise two-sample t-tests with a Bonferroni correction for multiple testing. EXTEND scores have been further compared, using Spearman rank correlation, between counts and FPKM gene expression values from poly-A and stranded protocols.

QUANTIFICATION AND STATISTICAL ANALYSIS

ADDITIONAL RESOURCES

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Bacterial and virus strains		
Biological samples		
Chemicals, peptides, and recombinant proteins		
Critical commercial assays		
Deposited data		
Experimental models: Cell lines		
Experimental models: Organisms/strains		
Oligonucleotides		
Recombinant DNA		
Software and algorithms		
Other		

Supplemental Information Titles and Legends
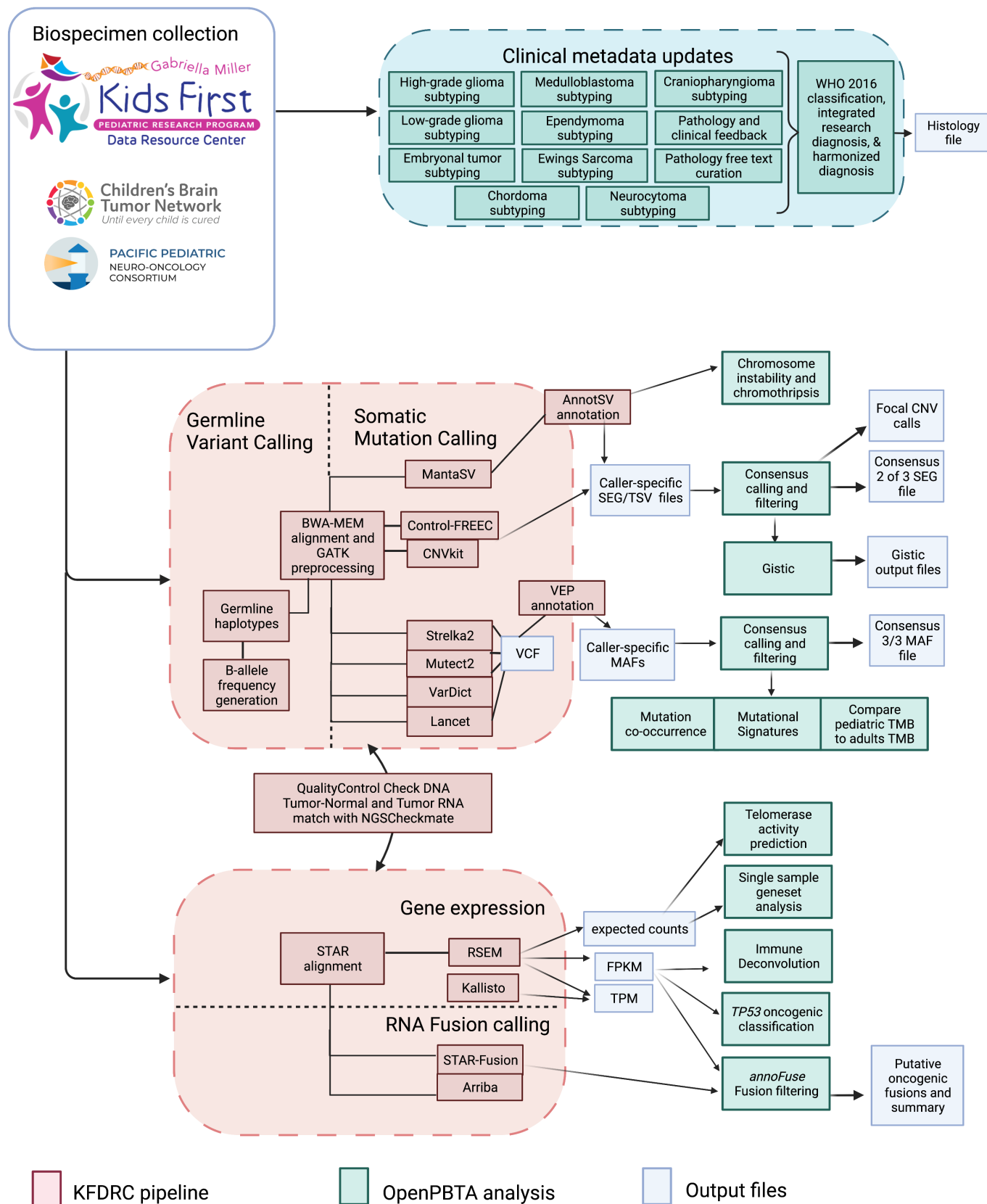
Figure 5: Figure S1. OpenPBTA Project Workflow, Related to Figure 1.

Figure S2: Consensus SNV/TMB, related to Figs 2 and 3
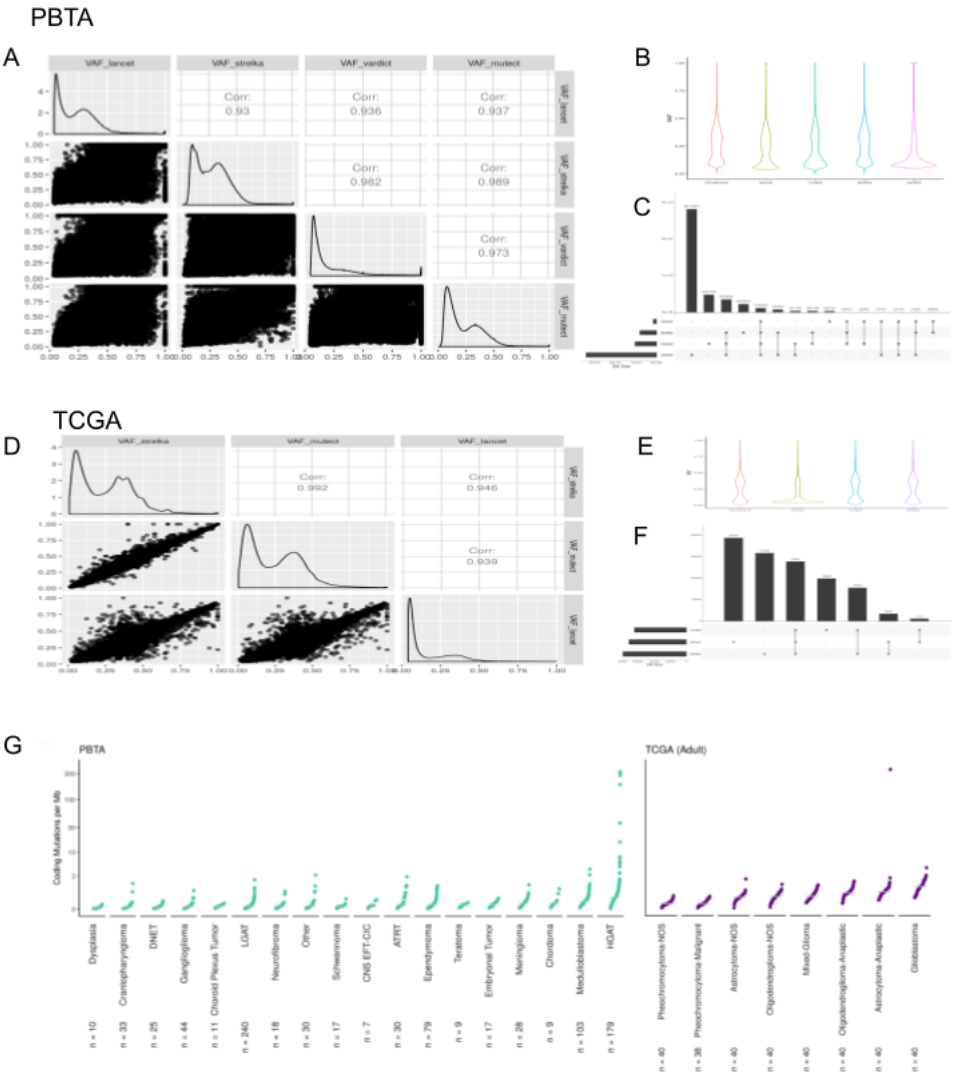
Figure 6: Figure S2. Validation of Consensus SNV calls and Tumor Mutation Burden, Related to Figures 2 and 3.

Figure S3: Other CNS Oncoprint, CNV landscape, related to Figs 2 and 3

Figure 7: Figure S3. Genomic instability of pediatric brain tumors, Related to Figures 2 and 3.

Figure S4: UMAP Subtypes & Immune deconv related to Fig 4

Figure 8: Figure S4. Related to Figure 4. First two dimensions from UMAP of sample transcriptome data with points colored by `molecular_subtype` for A, medulloblastoma, B, ependymoma, C, low-grade glioma, and D, high-grade diffuse astrocytic tumors.

Figure S5: TP53/EXTEND, related to Fig 5
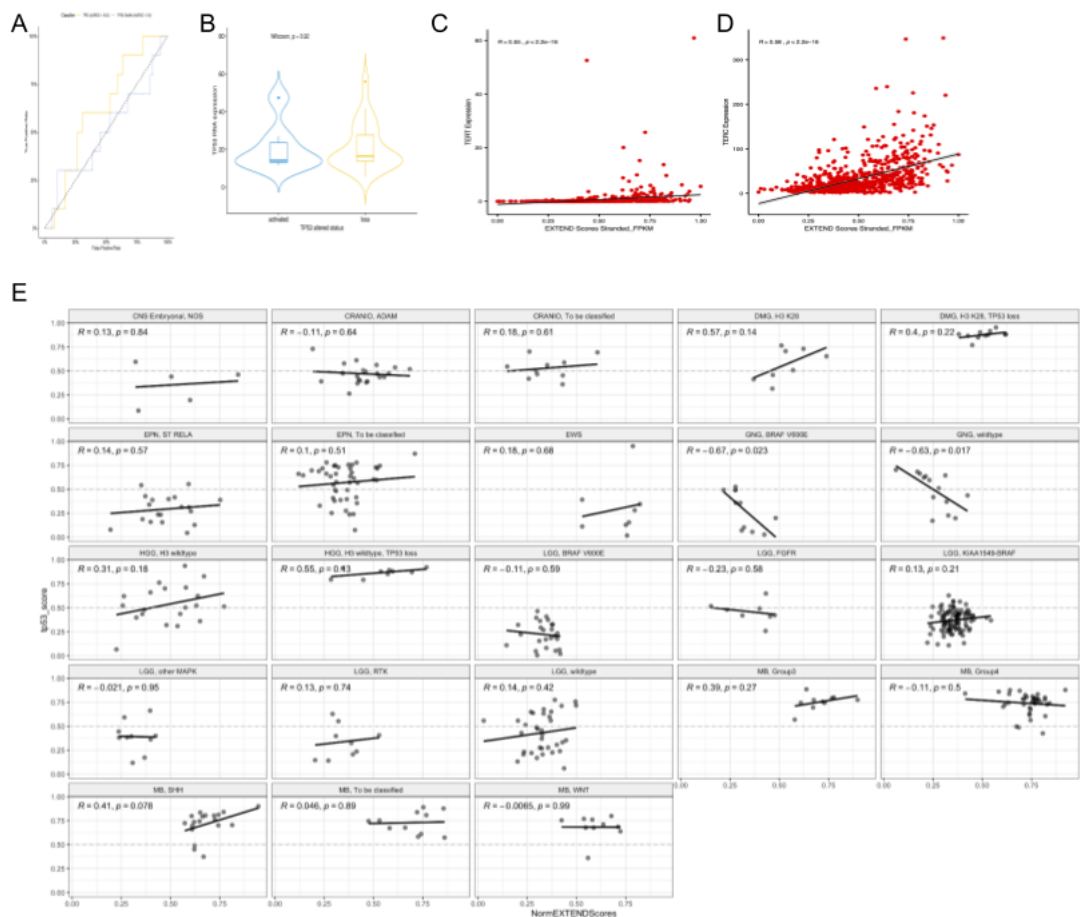
Figure 9: Figure S5. Related to Figure 5.

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 · 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/now207 · PMID: 28475809 · PMCID: PMC8483569

3. **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 · PMID: 31613963 · PMCID: PMC6954395

4. **Children's Brain Tumor Network** https://cbtn.org/

5. **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 · PMID: 30861105

6. **Manubot - Manuscripts, open and automated** https://manubot.org

7. **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 · PMID: 31233491 · PMCID: PMC6611653

8. **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 · PMID: 17618441 · PMCID: PMC1929165

9. **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 · PMID: 27157931

10. **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/gmqhbf
DOI: 10.1093/neuonc/noab106 · PMID: 34185076 · PMCID: PMC8328013

11. **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.ccell.2020.03.011 · PMID: 32289278 · PMCID: PMC7169997

12. **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 · PMID: 32164789 · PMCID: PMC7066826

13. **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 · PMID: 32601913 · PMCID: PMC7423804

14. **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 · PMID: 34272868 · PMCID: PMC8730763

15. **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 · PMID: 32650833 · PMCID: PMC7350623

16. **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 · PMID: 28726821 · PMCID: PMC5905700

17. **WHO classification of tumors of the nervous system: preview of the upcoming 5th edition**
Elisabeth J. Rushing
memo - Magazine of European Medical Oncology (2021-06) https://doi.org/gk2z97
DOI: 10.1007/s12254-021-00680-x

18. **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.ccell.2017.08.017 · PMID: 28966033 · PMCID: PMC5637314

19. **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 · PMID: 32197665 · PMCID: PMC7083001

20. *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 · PMID: 21060032

21. **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 · PMID: 24705251 · PMCID: PMC4056452

22. **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 · PMID: 29617664 · PMCID: PMC5961503

23. **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 · PMID: 31693904 · PMCID: PMC6880934

24. **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 · PMID: 8099841

25. **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/gmfxxd
DOI: 10.1038/s41467-020-20474-9 · PMID: 33420056 · PMCID: PMC7794223

26. **Pacific Pediatric Neuro-Oncology Consortium** https://pnoc.us/

27. **Working Together to Put Kids First** https://kidsfirstdrc.org/

28. **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

29. **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM**
Heng Li
arXiv (2013-05-28) https://arxiv.org/abs/1303.3997

30. **Index of /goldenPath/hg38/bigZips** http://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/

31. https://software.broadinstitute.org/gatk/best-practices/workflow?id

32. **SAMBLASTER: fast duplicate marking and structural variant read extraction**
G. G. Faust, I. M. Hall

Bioinformatics (2014-09-01) https://doi.org/f6kft3
DOI: 10.1093/bioinformatics/btu314 · PMID: 24812344 · PMCID: PMC4147885

33. **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: 10.1093/bioinformatics/btv098 · PMID: 25697820 · PMCID: PMC4765878

34. **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

35. **Broad Genome References** https://s3.amazonaws.com/broad-references/broad-references-readme.html

36. **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 · PMID: 28369524 · PMCID: PMC5499645

37. **GitHub - parklab/NGSCheckMate: Software program for checking sample matching for NGS data**
GitHub
https://github.com/parklab/NGSCheckMate

38. **The case for using Mapped Exonic Non-Duplicate (MEND) read counts in RNA-Seq experiments: examples from pediatric cancer datasets**
Holly C. Beale, Jacquelyn M. Roger, Matthew A. Cattle, Liam T. McKay, Drew K. A. Thomson, Katrina Learned, A. Geoffrey Lyle, Ellen T. Kephart, Rob Currie, Du Linh Lam, … Olena M. Vaske
Genomics (2019-07-30) https://doi.org/gghzj8
DOI: 10.1101/716829

39. **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 · PMID: 30013048

40. **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/ggntwv
DOI: 10.1101/861054

41. **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 · PMID: 30271907 · PMCID: PMC6123722

42. **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: 10.1093/nar/gkw227 · PMID: 27060149 · PMCID: PMC4914105

43. **gatk/mutect2.wdl at 4.1.1.0 · broadinstitute/gatk**
GitHub
https://github.com/broadinstitute/gatk

44. **GitHub - AstraZeneca-NGS/VarDictJava: VarDict Java port**
GitHub
https://github.com/AstraZeneca-NGS/VarDictJava

45. **Deep sequencing of 3 cancer cell lines on 2 sequencing platforms**
Kanika Arora, Minita Shah, Molly Johnson, Rashesh 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

46. **bcftools(1)** http://samtools.github.io/bcftools/bcftools.html#norm

47. **OpenPBTA-workflows/kfdrc_annot_vcf_sub_wf.cwl at master · d3b-center/OpenPBTA-workflows**
GitHub
https://github.com/d3b-center/OpenPBTA-workflows

48. **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 · PMID: 27268795 · PMCID: PMC4893825

49. **Cancer Hotspots** https://www.cancerhotspots.org/

50. **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 · PMID: 32839402 · PMCID: PMC7503716

51. **GitHub - mskcc/vcf2maf: Convert a VCF into a MAF, where each variant is annotated to only one of all possible gene isoforms**
GitHub
https://github.com/mskcc/vcf2maf

52. **OpenPBTA-analysis/compare_snv_callers_plots.Rmd at master · AlexsLemonade/OpenPBTA-analysis**
GitHub
https://github.com/AlexsLemonade/OpenPBTA-analysis

53. https://github.com/AlexsLemonade/OpenPBTA-analysis/blob/master/analyses/snv-callers/plots/comparison/upset_plot.png

54. https://github.com/AlexsLemonade/OpenPBTA-analysis/blob/master/analyses/snv-callers/plots/comparison/vaf_violin_plot.png

55. **GENCODE - Human Release 27** https://www.gencodegenes.org/human/release_27.html

56. **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 · PMID: 30341162 · PMCID: PMC6211645

57. **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 · PMID: 22155870 · PMCID: PMC3268243

58. **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 · PMID: 21081509 · PMCID: PMC3018818

59. **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 · PMID: 27100738 · PMCID: PMC4839673

60. **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 · PMID: 25297070 · PMCID: PMC4253833

61. **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
Genome Biology (2011-04) https://doi.org/dzhjqh
DOI: 10.1186/gb-2011-12-4-r41 · PMID: 21527027 · PMCID: PMC3218867

62. **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 · PMID: 26647377

63. **manta/README.md at 75b5c38d4fcd2f6961197b28a41eb61856f2d976 · Illumina/manta**
GitHub
https://github.com/Illumina/manta

64. **coverage — bedtools 2.30.0 documentation** https://bedtools.readthedocs.io/en/latest/content/tools/coverage.html

65. **BEDTools: a flexible suite of utilities for comparing genomic features**
Aaron R. Quinlan, Ira M. Hall
Bioinformatics (2010-03-15) https://doi.org/cmrms3
DOI: 10.1093/bioinformatics/btq033 · PMID: 20110278 · PMCID: PMC2832824

66. http://hgdownload.cse.ucsc.edu/goldenpath/hg38/database/cytoBand.txt.gz

67. **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 · PMID: 23155063 · PMCID: PMC3531082

68. **findOverlaps-methods function - RDocumentation** https://www.rdocumentation.org/packages/IRanges/versions/2.6.1/topics/findOverlaps-methods

69. **Software for Computing and Annotating Genomic Ranges**
Michael Lawrence, Wolfgang Huber, Hervé Pagès, Patrick Aboyoun, 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 · PMID: 23950696 · PMCID: PMC3738458

70. **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 · PMID: 29669011

71. **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 · PMID: 32025003 · PMCID: PMC7058534

72. **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 · PMID: 23104886 · PMCID: PMC3530905

73. **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 · PMID: 21816040 · PMCID: PMC3163565

74. **Near-optimal probabilistic RNA-seq quantification**
Nicolas L Bray, Harold Pimentel, Páll Melsted, Lior Pachter
Nature Biotechnology (2016-05) https://doi.org/f8nvsp
DOI: 10.1038/nbt.3519 · PMID: 27043002

75. **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, Tatiana Aneichyk
Bioinformatics (Oxford, England) (2019-07-15) https://www.ncbi.nlm.nih.gov/pubmed/31510660
DOI: 10.1093/bioinformatics/btz363 · PMID: 31510660 · PMCID: PMC6612828

76. **GitHub - icbi-lab/immunedeconv: A unified interface to immune deconvolution methods (CIBERSORT, EPIC, quanTIseq, TIMER, xCell, MCPcounter)**
GitHub
https://github.com/icbi-lab/immunedeconv

77. **xCell: digitally portraying the tissue cellular heterogeneity landscape**
Dvir Aran, Zicheng Hu, Atul J. Butte
Genome Biology (2017-12) https://doi.org/gckmjs
DOI: 10.1186/s13059-017-1349-1 · PMID: 29141660 · PMCID: PMC5688663

78. **Robust enumeration of cell subsets from tissue expression profiles**
Aaron M Newman, Chih Long Liu, Michael R Green, Andrew J Gentles, Weiguo Feng, Yue Xu, Chuong D Hoang, Maximilian Diehn, Ash A Alizadeh
Nature Methods (2015-05) https://doi.org/gcp3f5
DOI: 10.1038/nmeth.3337 · PMID: 25822800 · PMCID: PMC4739640

79. **GSVA: gene set variation analysis for microarray and RNA-Seq data**
Sonja Hänzelmann, Robert Castelo, Justin Guinney
BMC Bioinformatics (2013-12) https://doi.org/gb8vx5
DOI: 10.1186/1471-2105-14-7 · PMID: 23323831 · PMCID: PMC3618321

80. **GSVA**
Justin Guinney [Aut, Cre], Robert Castelo [Aut], Joan Fernandez[Ctb]
Bioconductor (2017) https://doi.org/ggxrqs
DOI: 10.18129/b9.bioc.gsva

81. **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 · PMID: 26771021 · PMCID: PMC4707969

82. **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

83. https://cran.r-project.org/package

84. **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

85. **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 · PMID: 29425488

86. **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 · PMID: 23633565 · PMCID: PMC3657784

87. **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 · PMID: 22832581 · PMCID: PMC3683624

88. **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 · PMID: 26919435 · PMCID: PMC5139621

89. **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 · PMID: 24316981

90. **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 · PMID: 26523776

91. **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 · PMID: 24123354

92. **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 · PMID: 25481751

93. **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 · PMID: 28271343

94. **MYB-QKI rearrangements in angiocentric glioma drive tumorigenicity through a tripartite mechanism**
Pratiti Bandopadhayay, 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 · PMID: 26829751 · PMCID: PMC4767685

95. **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/f8bdsq
DOI: 10.1186/s13059-016-0893-4 · PMID: 26899170 · PMCID: PMC4762164

96. **GitHub - raerose01/deconstructSigs: deconstructSigs**
GitHub
https://github.com/raerose01/deconstructSigs

97. **BSgenome.Hsapiens.UCSC.hg38**
Bioconductor
http://bioconductor.org/packages/BSgenome.Hsapiens.UCSC.hg38/

98. **COSMIC - Catalogue of Somatic Mutations in Cancer**
Cosmic
https://cancer.sanger.ac.uk/cosmic

99. **Signatures of mutational processes in human cancer**
Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, ICGC PedBrain, Ludmil B. Alexandrov, Serena Nik-Zainal, David C. Wedge, Samuel A. J. R. Aparicio, Sam Behjati, Andrew V. Biankin, … Michael R. Stratton

Nature (2013-08-22) https://doi.org/f22m2q
DOI: 10.1038/nature12477 · PMID: 23945592 · PMCID: PMC3776390

100. **Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines**
Kyle Ellrott, Matthew H. Bailey, Gordon Saksena, Kyle R. Covington, Cyriac Kandoth, Chip Stewart, Julian Hess, Singer Ma, Kami E. Chiotti, Michael McLellan, … Armaz Mariamidze
Cell Systems (2018-03) https://doi.org/gf9twn
DOI: 10.1016/j.cels.2018.03.002 · PMID: 29596782 · PMCID: PMC6075717

101. **Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas**
The Cancer Genome Atlas Research Network
New England Journal of Medicine (2015-06-25) https://doi.org/f7f82c
DOI: 10.1056/nejmoa1402121 · PMID: 26061751 · PMCID: PMC4530011

102. **The Somatic Genomic Landscape of Glioblastoma**
Cameron W. Brennan, Roel G.W. Verhaak, Aaron McKenna, Benito Campos, Houtan Noushmehr, Sofie R. Salama, Siyuan Zheng, Debyani Chakravarty, J. Zachary Sanborn, Samuel H. Berman, … Roger McLendon
Cell (2013-10) https://doi.org/f5dbzj
DOI: 10.1016/j.cell.2013.09.034 · PMID: 24120142 · PMCID: PMC3910500

103. **Comprehensive Molecular Characterization of Pheochromocytoma and Paraganglioma**
Lauren Fishbein, Ignaty Leshchiner, Vonn Walter, Ludmila Danilova, A. Gordon Robertson, Amy R. Johnson, Tara M. Lichtenberg, Bradley A. Murray, Hans K. Ghayee, Tobias Else, … Erik Zmuda
Cancer Cell (2017-02) https://doi.org/f9vcmf
DOI: 10.1016/j.ccell.2017.01.001 · PMID: 28162975 · PMCID: PMC5643159

104. **GENCODE - Human Release 19** https://www.gencodegenes.org/human/release_19.html

105. **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 · PMID: 27069505 · PMCID: PMC4827218

106. **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 · PMID: 32397004

107. **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 · PMID: 30249036 · PMCID: PMC6315657

108. **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 · PMID: 28744687

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

110. **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 · PMID: 32155762 · PMCID: PMC7084764

111. **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 · PMID: 23161096 · PMCID: PMC3508282

112. **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 · PMID: 27121965 · PMCID: PMC4847212

113. **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 · ISBN: 9783319334301

114. **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 · PMID: 7616262

115. **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 · PMID: 9761047

116. https://pfam.xfam.org/family/PF07714

117. http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/pfamDesc.txt.gz

118. **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.ccell.2015.04.002 · PMID: 25965575 · PMCID: PMC4712639

119. **Accelerating Discovery of Functional Mutant Alleles in Cancer**
Matthew T. Chang, Tripti Shrestha Bhattarai, 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 · PMID: 29247016 · PMCID: PMC5809279

120. **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 Socci, David B Solit, Adam B Olshen, … Barry S Taylor
Nature Biotechnology (2016-02) https://doi.org/gf7vxg
DOI: 10.1038/nbt.3391 · PMID: 26619011 · PMCID: PMC4744099

121. **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 · PMID: 16543939

122. **Inherited**
Tanya Guha, David Malkin
Cold Spring Harbor perspectives in medicine (2017-04-03) https://www.ncbi.nlm.nih.gov/pubmed/28270529
DOI: 10.1101/cshperspect.a026187 · PMID: 28270529 · PMCID: PMC5378014

123. **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 · PMID: 19505943 · PMCID: PMC2723002

124. **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