

# The Open Pediatric Cancer Project

This manuscript ([permalink](#)) was automatically generated from [d3b-center/OpenPedCan-methods@24e6ad9](#) on June 13, 2023.

## Authors

---

- **Eric Wafula**

 [0000-0001-8073-3797](#)

Department of Bioinformatics and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Krutika S. Gaonkar**

 [0000-0003-0838-2405](#)

Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA; Department of  
Bioinformatics and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Run Jin**

 [0000-0002-8958-9266](#)

Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Komal S. Rathi**

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

Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Department of Bioinformatics and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Yuankun Zhu**

 [0000-0002-2455-9525](#)

Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Bailey K. Farrow**

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


Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Daniel P. Miller**

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

Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Mariarita Santi**

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

Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine,  
Philadelphia, PA, 19104, USA

- **Adam A. Kraya**

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

Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Xiaoyan Huang**

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

Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Bo Zhang**

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

Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Brian M. Ennis**

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

Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Ryan J. Corbett**

 [0000-0002-3478-0784](#)

Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Sharon J. Diskin**

 [0000-0002-7200-8939](#)

Division of Oncology, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA; Department of Pediatrics,  
University of Pennsylvania, Philadelphia, PA, 19104, USA

- **Nicholas Van Kuren**

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

Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Noel Coleman**

 [0000-0001-6454-1285](#)

Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Christopher Blackden**

Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Jennifer L. Mason**

Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Miguel A. Brown**

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



Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;  
Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

- **Adam C. Resnick**

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

Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA;

Division of Neurosurgery, Children’s Hospital of Philadelphia, Philadelphia, PA, 19104, USA · Funded by 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

- **Jo Lynne Rokita**<sup>^</sup>   [0000-0003-2171-3627](#)

Center for Data-Driven Discovery in Biomedicine, Children’s Hospital of Philadelphia, Philadelphia, PA, 19104, USA; Division of Neurosurgery, Children’s Hospital of Philadelphia, Philadelphia, PA, 19104, USA; Department of Bioinformatics and Health Informatics, Children’s Hospital of Philadelphia, Philadelphia, PA, 19104, USA · Funded by NCI/NIH Contract No. 75N91019D00024, Task Order No. 75N91020F00003

## Contact information

---

<sup>^</sup>Lead Contact: Jo Lynne Rokita [rokita@chop.edu](mailto:rokita@chop.edu)

✉Correspondence: Jo Lynne Rokita [rokita@chop.edu](mailto:rokita@chop.edu)

## In Brief

---

## Highlights

---

## Summary

---

## Keywords

---

## Introduction

---

## Results

---

## Discussion

---

## Acknowledgments

---

## Author Contributions

---

Author	Contributions
Eric Wafula	Formal analysis, Software
Krutika S. Gaonkar	Data curation, Formal analysis, Investigation, Methodology, Software, Writing – Original draft, Writing - Review and editing
Run Jin	Data curation, Formal analysis, Visualization, Writing – Original draft, Writing - Review and editing
Komal S. Rathi	Formal analysis, Investigation, Methodology, Writing – Original draft

Author	Contributions
Yuankun Zhu	Data curation, Formal analysis, Investigation, Methodology, Supervision
Bailey K. Farrow	Data curation, Software
Daniel P. Miller	Formal analysis
Mariarita Santi	Investigation, Validation, Writing - Review and editing
Adam A. Kraya	Methodology
Xiaoyan Huang	Formal analysis
Bo Zhang	Data curation, Formal analysis
Brian M. Ennis	Data curation, Formal analysis
Ryan J. Corbett	Formal analysis
Sharon J. Diskin	Investigation, Supervision, Validation, Funding acquisition, Writing - Review and editing
Nicholas Van Kuren	Data curation, Software
Noel Coleman	Data curation
Christopher Blackden	Resources
Jennifer L. Mason	Supervision
Miguel A. Brown	Data curation, Methodology, Formal analysis
Adam C. Resnick	Conceptualization, Funding acquisition, Resources, Supervision
Jo Lynne Rokita <sup>^</sup>	Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Supervision, Writing – Original draft, Writing - Review and editing

## Declarations of Interest

---

## Figure Titles and Legends

---

## Table Titles and Legends

---

## OPENPEDCAN METHODS

---

## RESOURCE AVAILABILITY

### Lead contact

Requests for access to OpenPedCan 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**

Merged summary files for OpenPedCan v12 are openly accessible in [CAVATICA](#) or via download script in the <https://github.com/PediatricOpenTargets/OpenPedCan-analysis> repository. Cancer group summary data are visible within the NCI's pediatric [Molecular Targets Platform](#) and cohort, cancer group, and individual data are visible within [PedicBioPortal](#)

OpenPedCan analysis modules were developed within OpenPBTA [1], modified based on OpenPBTA, or newly created and can be found within the following publicly available repositories. OpenPBTA module analyses can be found at <https://github.com/AlexsLemonade/OpenPBTA-analysis>.

OpenPedCan module analyses can be found at

<https://github.com/PediatricOpenTargets/OpenPedCan-analysis>. OpenPedCan api code can be found at <https://github.com/PediatricOpenTargets/OpenPedCan-api>.

Software versions are documented in **Table XX**.

## **Data releases**

## **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**

## **METHOD DETAILS**

### **Nucleic acids extraction and library preparation**

### **Data generation**

### **DNA WGS Alignment**

Please refer to the OpenPBTA manuscript for details [1].

### **Quality Control of Sequencing Data**

Please refer to the OpenPBTA manuscript for details [1].

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

Please refer to the OpenPBTA manuscript for details [1].

### **Somatic Mutation Calling**

### **SNV and indel calling**

Please refer to the OpenPBTA manuscript for details [1].

### **VCF annotation and MAF creation**

### **Gather SNV and INDEL Hotspots**

### **Consensus SNV Calling**

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

## **Consensus CNV Calling**

## **Somatic Structural Variant Calling (WGS samples only)**

Please refer to the OpenPBTA manuscript for details [\[1\]](#).

## **Methylation Analysis**

## **Gene Expression**

## **Abundance Estimation**

## **Gene Expression Matrices with Unique HUGO Symbols**

## **Gene fusion detection**

## **QUANTIFICATION AND STATISTICAL ANALYSIS**

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

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

Please refer to the OpenPBTA manuscript for details [\[1\]](#).

## **Fusion prioritization ( `fusion_filtering` analysis module)**

## **Mutational Signatures ( `mutational-signatures` analysis module)**

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

## **Clinical Data Harmonization**

## **WHO Classification of Disease Types**

## **Molecular Subtyping**

Here, we build upon the molecular subtyping performed in OpenPBTA [\[1\]](#).

High-grade gliomas..

Atypical teratoid rhabdoid tumors..

Neuroblastoma tumors...

## **Integration of brain tumor methylation classifications**

## **TP53 Alteration Annotation ( `tp53_nf1_score` analysis module)**

Please refer to the OpenPBTA manuscript for details [\[1\]](#).

## **Prediction of participants' genetic sex**

Please refer to the OpenPBTA manuscript for details [1].

## **Selection of independent samples ( [independent-samples](#) analysis module)**

For analyses that require all input biospecimens to be independent, we use the OpenPedCan-analysis [independent-samples](#) module to select only one biospecimen from each input participant. For each input participant of an analysis, the independent biospecimen is selected based on the analysis-specific filters and preferences for the biospecimen metadata, such as experimental strategy, cancer group, and tumor descriptor.

## **Supplemental Information Titles and Legends**

---

### **Consortia**

---

## References

---

1. **OpenPBTA: The Open Pediatric Brain Tumor Atlas**  
Joshua A Shapiro, Krutika S Gaonkar, Stephanie J Spielman, Candace L Savonen, Chante J Bethell, Run Jin, Komal S Rath, Yuankun Zhu, Laura E Egolf, Bailey K Farrow, ... Jaclyn N Taroni  
*Cell Genomics* (2023-05) <https://doi.org/gr92p6>  
DOI: [10.1016/j.xgen.2023.100340](https://doi.org/10.1016/j.xgen.2023.100340)