

# Navigating St. Jude's PeCan v2 & Survivorship Data Sharing Tools

*Childhood Cancer Data Initiative Webinar Series*

1. Introductions
2. *PeCan v2*
3. *Live Demo*
4. *St. Jude Survivorship Portal*
5. *Live Demo*

# Introductions

*Gregory Reaman*

# Today's Speakers



**Clay McLeod**  
Director of Product and Engineering  
St. Jude Department of Computational  
Biology



**Dr. Xin Zhou**  
Faculty Member  
St. Jude Department of Computational  
Biology

# *PeCan v2*

*Clay McLeod*

# What is PeCan?

**Pediatric Cancer** reference knowledge base within St. Jude Cloud and available at <https://pecan.stjude.cloud>.

Goal: collate, harmonize, and make available a knowledge base of pediatric cancer data in the browser.

Version 1.0 was largely focused on somatic mutations and was published in 2021 as a part of the St. Jude Cloud ecosystem.

RESEARCH ARTICLES | AUTHOR CHOICE | MAY 01 2021

## St. Jude Cloud: A Pediatric Cancer Genomic Data-Sharing Ecosystem **FREE**

Clay McLeod; Alexander M. Gout; Xin Zhou; Andrew Thrasher ; Delaram Rahbarinia; Samuel W. Brady; Michael Macias ; Kirby Birch; David Finkelstein; Jobin Sunny; Rahul Mudunuri; Brent A. Orr ; Madison Treadway; Bob Davidson; Tracy K. Arr; Arthur Chiao; Andrew Swistak; Stephanie Wiggins; Scott Foy; Jian Wang; Edgar Sloson ; Shuoguo Wang; J. Robert Michael ; Yu Liu; Xiaotu Ma ; Aman Patel; Michael N. Edmonson; Mark R. Wilkinson; Andrew M. Frantz; Ti-Cheng Chang; Liqing Tian ; Shaohua Lei; S.M. Ashiqui Islam ; Christopher Meyer; Naina Thangaraj; Pamela Tater; Vijay Kandali ; Singer Ma ; Tuan Nguyen; Omar Serang ; Irina McGuire; Nedra Robison; Darrell Gentry ; Xing Tang; Lance E. Palmer; Gang Wu; Ed Suh; Leigh Tanner; James McMurry; Matthew Lear; Alberto S. Pappo; Zhaoming Wang; Carmen L. Wilson ; Yong Cheng; Soheil Meshinchi; Ludmil B. Alexandrov ; Mitchell J. Weiss; Gregory T. Armstrong; Leslie L. Robison; Yutaka Yasui; Kim E. Nichols ; David W. Ellison; Chaitanya Bangur; Charles G. Mullighan ; Suzanne J. Baker; Michael A. Dyer; Gerilyn Miller; Scott Newman; Michael Rusch ; Richard Daly; Keith Perry ; James R. Downing ; Jinghui Zhang

 Check for updates

+ Author & Article Information

*Cancer Discov* (2021) 11 (5): 1082–1099.

<https://doi.org/10.1158/2159-8290.CD-20-1230> [Article history](#)

### Related Content

A related article has been published: [In This Issue](#)

### Abstract

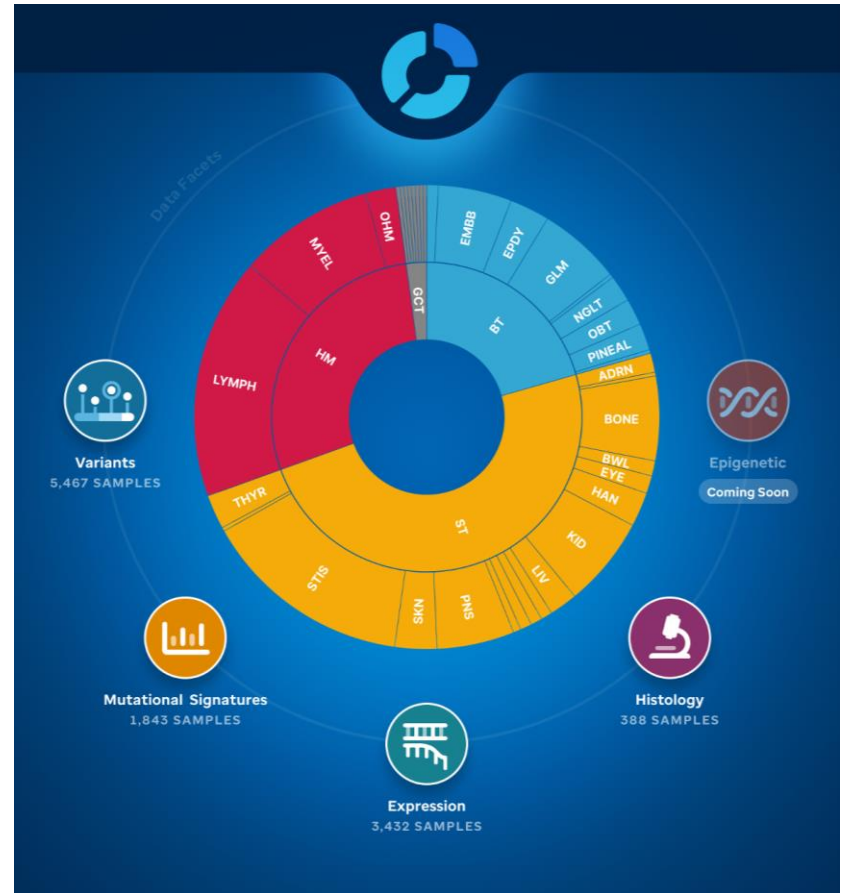
Effective data sharing is key to accelerating research to improve diagnostic precision, treatment efficacy, and long-term survival in pediatric cancer and other childhood catastrophic diseases. We present St. Jude Cloud (<https://www.stjude.cloud>), a cloud-based data-sharing ecosystem for accessing, analyzing, and visualizing genomic data from >10,000 pediatric patients with cancer and long-term survivors, and >800 pediatric sickle cell patients. Harmonized genomic data totaling 1.25 petabytes are freely available, including 12,104 whole genomes, 7,697 whole exomes, and 2,202 transcriptomes. The resource is expanding rapidly, with regular data uploads from St. Jude's prospective clinical genomics programs. Three interconnected apps within the ecosystem

McLeod et. al, *Cancer Discovery*, May 2021

# Motivation for Version 2.0

Shortly after our publication, we started ideating on the next version with these goals:

- Expand beyond somatic mutations to cover areas such as mutational signatures, gene expression, and histology images.
- Curate and organize the scientific content to enable new discovery and represent the most up-to-date knowledge/standards.
- Integrate all of this into a single, cohesive platform within the web browser.



# What have we achieved so far?



Built upon what already existed in PeCan v1 for genomic and epigenomic mutations, including:

- A new **oncoprint view** that summarizes the mutational landscape within each subtype.
- A new **mutational prevalence view** within which the frequency of mutations by gene and mutation type can be explored.
- Integration with **GenomePaint** (Zhou et. al, Cancer Cell, 2021).



Developed and made publicly available data facets for the new kinds of data we originally set out to incorporate in v2, including:

- Mutational signatures
- Gene expression
- Histological images



Greatly expanded the diagnosis ontology by incorporating the broader St. Jude Cloud ontology (contains 481 molecular subtypes with 388 represented in PeCan).



Rearchitected the client and API from the ground up for scalability and flexibility.



*Demo*

# Future Directions

- Continue to improve existing data facets
  - Expression plots for individual genes
  - Image search using machine learning for histology
- Cohort building
- Epigenetic data facet
- Subject and sample pages
- Update brain tumor ontology to match WHO CNS5

You have 98 samples selected

Choose the data you'd like to add to your cohort.

Samples <sup>98</sup>

ST.JUDE SAMPLE ID	VARIANTS	MUTATIONAL SIGNATURES	EXPRESSION	HISTOLOGY	EPIGENETICS
SJBT076364_D1	<input checked="" type="checkbox"/>	—	—	<input checked="" type="checkbox"/>	—
SJTALL002035_D1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
SJALL015285_D1	<input checked="" type="checkbox"/>	—	—	<input checked="" type="checkbox"/>	—
SJHGG078_A	<input checked="" type="checkbox"/>	—	—	<input checked="" type="checkbox"/>	—
SJWLM040001_D1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
SJHGG101_A	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
SJTALL013800_R1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
SJNBL017231_D1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

[Download all data](#) [Add to cohort](#)

## PeCan Team

Alex Gout  
Stephanie Sandor  
James Madson  
Jobin Sunny  
Kevin Benton  
Delaram Rahbarinia  
Sue Qiu  
Brian Curran  
Andrew Swistak  
Wayne Wang  
Pranaya Shrestha  
Ira Baranova

## Clinical Genomics

Victor Pastor  
David Wheeler

## Broader St. Jude Cloud

Andrew Thrasher  
Andrew Frantz  
David Finkelstein  
Michael Macias  
Mark Wilkinson  
Michael Rusch

## COMET Project

Abbas Shirinifard  
Ben Lansdell  
Cynthia Cline  
Brent Orr  
Gang Wu  
Elizabeth Stewart  
Michael Dyer

## Data Visualization

Colleen Reilly  
Edgar Sioson  
Jian Wang  
Xin Zhou

## Pathology

Larissa Furtado  
Jeff Klco  
David Ellison

## Leadership

Jinghui Zhang  
Keith Perry  
Charlie Roberts  
Jim Downing



# *St. Jude Survivorship Portal*

*Dr. Xin Zhou*

# St. Jude Survivorship Portal

## Department of Computational Biology

- Xin Zhou, Ph.D., assistant member
- Jinghui Zhang, Ph.D., member, Chair
- Clay McLeod, director
- Stephanie Sander, product manager

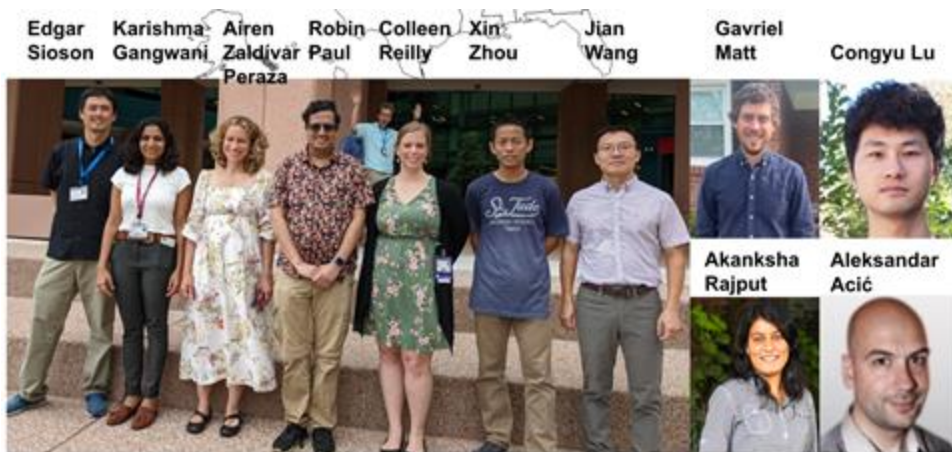
## Department of Epidemiology and Cancer Control

- Yutaka Yasui, Ph.D., member
- Les Robison, Ph.D., member emeritus
- Melissa Hudson, M.D., member
- Kiri Ness, P.T., Ph.D., member
- Greg Armstrong, M.D., M.S.C.E., member, Chair
- Kyla Shelton, manager

[survivorship.stjude.cloud](https://survivorship.stjude.cloud)

## The ProteinPaint Team

- 6 Ph.D. staffs
- 4 Web developers
- 1 Postdoc



# Background

- Pediatric cancer 5-year survival rate has increased significantly from less than 30% in the 1950s to over 85% today.
- The survivor population is an emerging clinical population that is growing fast and is at higher risk of adverse outcomes compared to the general population.
- To eliminate or mitigate these outcomes, survivorship research needs to analyze large cohort, multi-modality datasets to understand causes and develop risk-stratified intervention approaches.
- The St. Jude Survivorship Portal is designed to address this data access need, enabling data visualization and analysis.

# Survivorship Portal Data Content

## COHORT

St. Jude Lifetime Study (SJLIFE), n=5,053

Childhood Cancer Survivor Study (CCSS), n=2,688



## PHENOTYPES / EXPOSURES

Demographics, n=36

Cancer diagnosis, n=4

Cancer treatment, n=95

Clinical assessments, n=350

Chronic health conditions, n=400

Self-reported and questionnaire, n=776



## WHOLE-GENOME SEQUENCING

Genotypes for >400 million variants

Polygenic risk scores, >500 traits

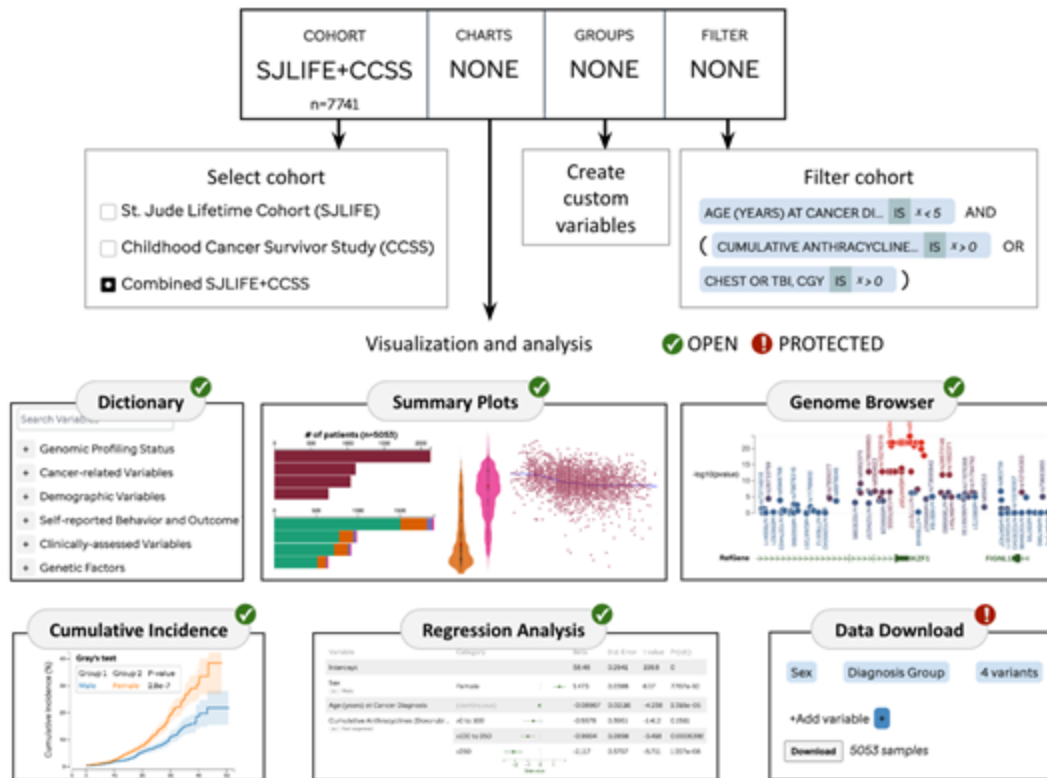
Genetic ancestries

Ancestry principal components

Linkage disequilibrium



# Navigation and Features





*Demo*

# Survivorship Portal Future Work

1. Integrate 25K additional participants from the Childhood Cancer Survivorship Study.
2. Integrate whole-exome sequencing and genotyping array results with whole-genome sequencing (WGS) genotype calling results.
3. Visualize raw sequencing reads (Binary Alignment Map files).
4. Phenome-wide association analysis.
5. Integrate additional types of genomic features from WGS, including copy number variation, structural variation, human leukocyte antigen typing, pharmacogene diplotypes.
6. Multi-omics integration: blood methylome and transcriptome, and single-cell RNA-sequencing.
7. Support longitudinal data.

# Find Out More About CCDI

**Learn about CCDI and subscribe to our monthly newsletter.**

[cancer.gov/CCDI](https://cancer.gov/CCDI)

**Questions? Email us.**

[NCIChildhoodCancerDataInitiative@mail.nih.gov](mailto:NCIChildhoodCancerDataInitiative@mail.nih.gov)



# Q&A

# Developing Pediatric Cancer Data Standards

Monday, February 26, 2pm - 3pm ET



**Dr. Michael Watkins**

Manager of Data Standards and Modeling  
Data for the Common Good



**Brian Furner**

Senior Director of Data and Technology  
Data for the Common Good



**Dr. Sam Volchenbom**

Principal Investigator & Pediatric Oncologist  
Data for the Common Good

Register Here: <https://cbiit.webex.com/weblink/register/r746056f4de0187615bd5bfb01319bcf5>

*Thank you!*

U.S. Department of Health & Human Services  
National Institutes of Health | National Cancer Institute

[cancer.gov](https://cancer.gov)

1-800-4-CANCER

