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

Childhood Cancer Data Initiative Webinar Series

January 23, 2024
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.

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?

- 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

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

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).
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
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

**The ProteinPaint Team**
- 6 Ph.D. staffs
- 4 Web developers
- 1 Postdoc

[Image of team members]
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.
7. Support longitudinal data.
Find Out More About CCDI

Learn about CCDI and subscribe to our monthly newsletter.
cancer.gov/CCDI

Questions? Email us.
NCIChildhoodCancerDataInitiative@mail.nih.gov
Developing Pediatric Cancer Data Standards
Monday, February 26, 2pm - 3pm ET

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

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 Volchenboum
Principal Investigator & Pediatric Oncologist
Data for the Common Good
Thank you!