Learning from Every Child: The NCI Childhood Cancer Data Initiative

May 23, 2022

Big Data Training for Cancer Research

Warren A. Kibbe, Ph.D.
Vice Chair & Professor of Biostatistics and Bioinformatics
Duke University School of Medicine
warren.kibbe@duke.edu

@wakibbe
NCI Childhood Cancer Data Initiative (CCDI) – Building A Community

Childhood and AYA cancers are rare diseases: only **4%** of diagnosed cancers

About 16,000 cases for children from birth to 19 years of age

Acute need for data to support research

**CCDI** aims to build a community focused on Pediatric and AYA Cancer
CCDI improves treatment, quality of life, and survivorship by learning from every child.
Personal & Professional Background

• PhD in Chemistry at Caltech, Postdoc in molecular genetics of RAS at the Max-Planck-Institute for Biophysical Chemistry
• Cancer research for 25+ years - cancer informatics, data science, healthcare – open science, open data advocate
• Feinberg School of Medicine at Northwestern for 15+ years
• Director NCI CBIIT 2013-2017; NCI CIO 2013-2017; Acting NCI Deputy Director for Data Science 2016-2017
• At Duke since 2017
• Currently on IPA with NCI to co-lead the CCDI
• Lost three grandparents to cancer, father to cancer in 2019
Disclaimer

The views I express are mine and not endorsed by Duke or NCI.
The importance of Open Science

Calls for greater transparency and ‘open data access’ in clinical research continue actively.

“Open science is the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society”*

Open Science Project**: “If we want open science to flourish, we should raise our expectations to: Work. Finish. Publish. Release.”

FAIR Principles: Findability, Accessibility, Interoperability, and Reusability***

TRUST Principles: Transparency, Responsibility, User focus, Sustainability and Technology****

* https://www.fosteropenscience.eu/resources
** http://openscience.org/
*** https://www.nature.com/articles/sdata201618
**** https://www.nature.com/articles/s41597-020-0486-7
Open Science and Patient Data Access

Some of the challenges are:

- Patient privacy
- Academic credit
- Commercial sensitivity and intellectual property
- Data standards
- Resources (money and people)

There should be room for researchers and patients alike to gain from this effort.

Informatics experts and data scientists are essential elements of this discussion.
One problem with Clinical Trials Data Sharing

“The tendency for researchers to “sit” on their data for an unduly long period of time is neither desirable from a scientific point of view nor acceptable from an ethical perspective.

After all, the data belong to the patients who agreed to participate in the research, not to the investigators who coordinated it, as the new European General Data Protection Regulation emphasizes.”*

Access to patient-level data is important for research

There are certainly challenges, but question is not whether data should be shared, but rather how and when access should be granted.

Responsible open access enables secondary analyses that:

• **Enhance reproducibility** of clinical research
• **Honor** the **contributions** of trial participants,
• **Improve the design** of future trials
• **Generate new research** findings

This journey of making patient data available is part of an evolution in transparency and not a sudden awakening.
The goal of the CCDI is to build a community of pediatric cancer researchers, advocates, families, hospitals, and networks committed to sharing data to improve treatments, quality of life, and survivorship of every child with cancer.
Pediatric/AYA data from multiple sources

- Development of new research and analytical tools
- Culture change towards improved collaboration and data sharing
- Improved understanding of why some cancers develop resistance or don’t respond to treatment
- Generation of new ideas for intervention

CCDI
Our ability to generate biomedical data continues to grow in terms of variety and volume.
What is Real World Data?

Collected in the context of patient care. Real World Data was called out as part of the 21st Century Cures Act.

Patient Data Touchpoints

Screening, Diagnosis, Treatment

- **Decision-making / Clinical Care**: Molecular Tumor Board, Review notes, treatment decisions
- **Confirmation / Refinement**: Molecular Sequencing, proteomics, etc.
- **Diagnosis**: Clinical presentation, biopsy, imaging
- **Progression**
- **Recurrence**
- **Secondary Cancer**

**Outcomes**
Short-term outcomes

**Longitudinal Follow-up**
Long-term outcomes

**Treatment**
Treatments, procedures, adverse events

**Epidemiology / Population Sciences Data**
Familial data, environmental, registry, population studies, disease cohorts
Cancer Data “Scorecard”

- Diagnosis
- Molecular Characterization
- Molecular Tumor Board
- Treatment
- Outcomes
- Long-term follow-up
- Recurrence
- Follow-on Cancer
Cancer Data “Scorecard”

- **Diagnosis**: EHRs, LIMS, Radiology, Pathology Systems
- **Molecular Characterization**: LIMS, Sequencing Systems
- **Follow-on Cancer**: EHRs, LIMS, Radiology, Pathology Systems
- **Recurrence**: EHRs, LIMS, Radiology, Pathology Systems
- **Long-term follow-up**: EHRs, Registries
- **Outcomes**: EHRs, CDMS, CTMS
- **Treatment**: EHRs, CDMS, Adverse Events, CTMS
- **Clinical Decision, Molecular Tumor Board**: Tumor Board Systems, Hospital Systems, EHRs, Clinical Trial Matching

**Federated Data Ecosystem**
History of Initiatives and Policy Leading up to CCDI

**Initiatives**
- Gabriella Miller Kids First Research Act
- Cancer Moonshot Blue Ribbon Panel (Oct)
- 21st Century CURES Act (Dec)

**Policies**
- NIH Genomic Data Sharing Policy (Jan)
- NIH Intramural Human Data Sharing Policy (July/Aug)
- NCBI Moonshot Public Access & Data Sharing Policy (Jan)
- Cancer Moonshot Blue Ribbon Panel (Oct)
- RACE for Children Act (Aug)
- STAR Act (Jun)
- Childhood Cancer Data Initiative (Feb)
- NCAB Ad Hoc WG on Data Science Report (Jun)

**Holdren Memo**

**2013**

**2014**

**2015**

**2016**

**2017**

**2018**

**2019**

**2020**
Ad Hoc Working Group in Support of the CCDI

Kevin Shannon (Co-Chair)
UCSF

Otis Brawley (Co-Chair)
Johns Hopkins

Peter Adamson
Sanofi

Tom Curran
Children’s Mercy

James Downing
St. Jude

Julie Guillot
Leukemia & Lymphoma Society

Amanda Haddock
Dragon Master Foundation

Douglas Hawkins
Seattle Children’s

Andrea Hayes-Jordan
UNC

Katherine Janeway
DFCI

Warren Kibbe
Duke

Andrew Kung
MSKCC

John Maris
CHOP

Samuel Volchenboum
U Chicago
CCDI Activities: Building On Recommendations

**Enhance** (Reduce Barriers) to **Data Sharing**

Establish **Federated Infrastructure** (*cancer data ecosystem*) – sharing/analyzing research and clinical data:

- Identify/connect data (clinical, preclinical, research, surveillance) and tools
- Develop enabling cancer technologies to enrich/automate – data capture, curation, harmonization, use, and publication

Create **Central Resource Catalog** of available pediatric/AYA data, biospecimens, and tools

Aggregate preclinical testing and cancer model data to inform **Rapid Clinical Translation** (*FDA Relevant Molecular Targets List*)

Develop a **National Strategy** of biospecimen collection/archiving (germline & tumor) and genomic testing (tx data → outcome; relapse seq) for **every child**
Foundational Goals for CCDI

• Gather data from every child/AYA diagnosed with cancer, regardless of where they receive their care
• Develop core data from consented patients, including tumor and germline molecular characteristics, to enable research using patient-level data in a secure and de-identified way
• Create a system that can bring data of different types, from different sources together in a way that incentivizes researchers to query the available data in new ways
A national strategy, building on efforts including COG’s Project:EveryChild, to offer appropriate clinical and molecular characterization to every child with cancer that:

- Enables discovery when these and other data are connected
- Defines a minimum set of molecular diagnostics to be collected for every pediatric and AYA cancer patient
- Is accessible to all children with cancer, including those treated at community-based institutions; provide access to underserved pediatric cancer patients
- Clinical sequencing of ~3,000 patients/year
- Align with Rare Pediatric Tumor Cell Atlas

CCDI Childhood Molecular Characterization Protocol
Childhood Molecular Characterization Protocol

• Expand access to comprehensive molecular sequencing as a step towards the goal of reaching all children with pediatric cancer
• Develop NCI-recommended guidelines for clinical and molecular data collection as part of standard of care
• Create a comprehensive, harmonized, and integrated database of clinical, genomic, and phenotypic data for research
• First patient was enrolled in April 2022!

Clinical Service: Diagnostic clinical molecular characterization services for patients who might not otherwise have access to them

Data to be collected (CLIA certified)

• DNA: CLIA WES or NGS targeted panel
• Fusion panel
• Methylation: CLIA DNA Methylation array
• Clinical annotation

Research Discovery: Molecular characterization to learn more about disease subtypes and rare cancers

Data to be collected (in addition to clinical/seq data on selected populations)

• WGS/deep molecular (DNA) profiling
• RNAseq
• Longitudinal data
CCDI Molecular Characterization Initiative

Launched March 21, 2022

First patient enrolled April 2022

CCDI Molecular Characterization Initiative

CCDI Molecular Characterization Initiative

What is the CCDI Molecular Characterization Initiative?

The Molecular Characterization Initiative is part of NCI's Childhood Cancer Data Initiative (CCDI) and is a national collaboration between the childhood cancer community, including the Children's Oncology Group, advocates, pediatric oncologists, researchers, data scientists, children and adolescent and young adults (AYAs) with cancer, and families. It provides state-of-the-art molecular characterization at the time of diagnosis that helps participants and doctors select the best and most appropriate treatment.

How it works
A specialized and accredited lab analyzes tumor and blood samples that may have already been provided for initial diagnosis to learn about a participant's cancer at its most basic, molecular level. Results will be shared with them and their doctors to inform treatment options and potential clinical trial participation.

Later, these data will be made accessible to researchers for future studies. Data that are shared with researchers have no personal information and cannot be connected directly to participants.

Participation in the initiative is voluntary and the costs to participate are covered by the program.

Helping advance childhood cancer research
These data will also enable researchers to create better clinical trials, learn about the origins and drivers of childhood cancer, and make faster progress in the development of new and better treatments, especially for childhood cancers with limited effective treatment options.

Who can participate in the Molecular Characterization Initiative?

Currently, children and AYAs who are both newly diagnosed with a central nervous system tumor and receiving care at a hospital affiliated with the Children’s Oncology Group are eligible to participate in the initiative.

To join at this time, a child or AYA must:

• be newly diagnosed (have not had any treatments for their cancer yet)
• be 25 years old or younger
• be diagnosed with a central nervous system tumor (tumors of the brain and spine)
• get cancer care from a hospital affiliated with the Children's Oncology Group

If a child or AYA meets the above criteria, they can get more information on how to participate by talking with their Children’s Oncology Group-affiliated doctor.

The initiative will expand to include children and AYAs with soft tissue sarcomas and other rare tumors later in 2022. In late 2022 and 2023, it will continue to expand to children and AYAs outside of Children's Oncology Group-affiliated hospitals, those with other childhood cancers, and those whose cancer has returned.

How will personal health information be protected?

Privacy is extremely important to NCI, and personal information will only be shared with participants and their medical team.

Data from the Molecular Characterization Initiative and from any clinical trials will have personal health information removed before it is made accessible to researchers. There will be no way for researchers or anyone to connect the data to participants directly.

CCDI Molecular Characterization Initiative

Where can you learn more about molecular characterization?

Talk with a doctor

To participate in the Molecular Characterization Initiative, an eligible child or AYA must be receiving care at a Children’s Oncology Group–affiliated hospital. To check a hospital’s affiliation, visit the group’s institution locator page.

If a child or AYA is not eligible for the initiative at this time, they can talk with their doctor or oncology team about molecular characterization. They can also discuss clinical trials that might be studying new ways to treat specific types of cancer. Additional clinical trial information is available on trials.cancer.gov.

Reach out to NCI’s Cancer Information Service

Please contact NCI’s Cancer Information Service via chat, email, or phone (both in English and Spanish) with additional questions. Information specialists can help find information on cancer in children and AYAs, clinical trials, and relevant community resources.

Updated: March 21, 2022

• Gather data from every child diagnosed with cancer in the United States
• It will:
  ✓ Capture the cancer care trajectory of children and AYAs, including care provided outside of COG and other networks, to identify gaps and disparities in care and outcome
  ✓ Track biospecimen availability
  ✓ Provide access to data from underserved patients
  ✓ Provide for consistent research consent
  ✓ Allow for long-term follow up of childhood cancer patients
• A critical component of this effort will be the **National Childhood Cancer Registry (NCCR)**
The National Childhood Cancer Registry (NCCR)

• Slides from Dr. Lynne Penberthy, NCI
The National Childhood Cancer Registry (NCCR)

• NCCR is a centralized infrastructure that brings together existing data on all cancer patients ages 0 to 39
  • Core data are derived from cancer registries
  • Data from registries are extended and expanded to include additional relevant information such as
    • Detailed treatment
    • Genomic characterization of the tumor
    • Risk of recurrence
    • Risk of subsequent primary cancers
  • With a goal to capture the complete trajectory of care from diagnosis throughout the patient’s life

• Currently includes patients representing 77% of all childhood cancers with 21 state cancer registries participating
What is a cancer registry?

- Cancer registry data provide the base for the NCCR
- Registries:
  - collect, store, and manage data on every person with cancer from diagnosis until death
  - are population based (capture all cancers within a defined geographic area)
  - have the legal authority to capture data on every cancer
  - require health care providers to report information on cancer patients to state registries
Routine linkages will be performed centrally with external data sources including:

- Complete **registry abstracts** for each cancer case (1995-2019+)
- Survival data from **National Death Index (NDI)** & **State vital records**
- **Residential history data**
  - essential to enable data to be linked over time for treatment, recurrence, subsequent cancers or adverse effects
- **Virtual Pooled Registry (VPR)**
  - A national system of virtual connections for all state registries the supports
    - de-duplication (many pediatric patients receive care across state boundaries) and
    - allows capture of subsequent cancers as patients age
Routinely link with *external data sources* via central linkage infrastructure

- **Pharmacy Data**—
  - Real time feeds from CVS/Walgreens/Riteaid (Real Time)/ PBM United Health Care
  - Provides data on oral antineoplastics often used to treat recurrent disease
- **Longitudinal Radiation oncology treatment data**
  - Initial and treatment of recurrence
- **Claims data** linkages (United HealthCare, Medicaid etc.)
  - Treatment (chemotherapy, surgery etc.)
  - Comorbidity (pre-diagnosis and post diagnosis/treatment)
- **Radiology reports + images**
  - case finding/ recurrence
- **Genomic Data**
  - Enables better classification of cancers to understand disparities in outcome
Data sources currently used

- Birth Records, Blood spots
- Selected state vital records
- Long-term follow-up center, AACCR Children’s Oncology Group
- SEER Cancer Registries
- Leveraging Existing SEER*DMS servers holding PII for Secure Data Platform

NCCR Database***
Combines de-identified data submitted from participating registries plus linked data from additional sources

Selected State Cancer Registries (including TX, TN, PA, IL, NJ, OH, FL)
Instance of DMS*Lite
For each registry to hold PII for linkages

SEER Cancer Registries
Leveraging Existing SEER*DMS servers holding PII for NCCR Database***
Combines de-identified childhood cancer patient data submitted from participating registries. Infrastructure to support research on childhood cancers

New NCCR data sources

- Selected state vital records
- Birth Records, Blood spots
- Children’s Oncology Group

***NCCR Database— holds de-identified childhood cancer patient data submitted from participating registries. Infrastructure to support research on childhood cancers
NCCR is Supported/Guided by multiple Working Groups

A broad general/scientific working group
- Including clinicians, epidemiologists, advocates, IRB members, registrars
- Representing cancer centers, registries etc.
- Participating on focused working groups

Focused working Groups
- Meta-Data
  - Harmonizing existing and recommending new data to collect
- Data Access & Release
  - Developing processes for appropriate data release & access – protecting patient privacy
- Data Products
  - Informing key analyses and data sets
Why is NCCR important?

• NCCR provides:
  • Patient-level data to understand cancer in every child
  • Accessible data on how patients are diagnosed, treated, and associated outcomes
  • A measurement of how well we are doing in reducing mortality from childhood cancers over time

• NCCR allows us to answer important questions:
  • Where do we need to focus attention to improve outcomes?
  • Who may not be participating in clinical trials or receiving state of the art care?
  • How far are patients traveling to receive care?
  • Many more questions…
How does NCCR fit into CCDI?

• NCCR provides valuable data on every child, adolescent, and young adult with cancer related to diagnosis, treatment and survivorship, directing the focus of CCDI activities.
  • For example, NCCR data
    • can guide the Molecular Characterization Protocol, showing where we have the most opportunity for impact
    • direct CCDI to focus on patients who may not have access to clinical trials or COG facilities and state of the art care
Why are the data from NCCR important: *Providing a Report Card* - Leukemia as an example*

- Registries provide de-identified data **on every patient with cancer** in a defined geographic region
- These data allow us to
  - **Monitor progress** (report card) of clinical care on outcomes
  - Help us identify groups of patients who may not be experiencing the benefit of the new treatments by age, racial and ethnic subgroups, or geographic area
- Similar trends are being developed for survival and mortality

NCCR Explorer

https://nccrexplorer.ccdi.cancer.gov
NCCR Explorer

https://nccrexplorer.ccdi.cancer.gov
NCCR Explorer

https://nccrexplorer.ccdi.cancer.gov
Designed to federate data from multiple children’s cancer institutions and community-based and NCI-supported childhood/AYA data resources, featuring:

- Patient-level data from all available sources
- Easy access to data to enable deep analytics
- Supports interoperability among existing data resources and with tools and other resources for use by researchers
- Provide a central portal to find and analyze childhood/AYA cancer data
Using data to achieve the goals of CCDI

**Piece it together:** CCDI is completing the puzzle to learn from and help heal children, teens, and young adults with cancer.

**Build a strong base:** Progress requires data from many sources that is connected and easy to access.

**Assemble better data:** Complete data sets are needed to understand each type of cancer.

**Make data easy to use:** More thoughtful tools for analyzing data will help answer important questions.

**Improve treatments:** Data is the foundation that informs new treatments and improves lives faster.

Slides thanks to Dr Tony Kerlavage, NCI
Establish Infrastructure to Manage & Share Data

NCI Activities:

- Develop a **Federated Pediatric Cancer Data Ecosystem** of research repositories & patient registries:
  - **National Childhood Cancer Registry (NCCR)** – link clinical patient data (e.g. SEER, VPR, PPCR, CCRN)
  - **Pediatric Preclinical Data Commons (PPDC)**
  - **Pediatric/ AYA Data/Tools** from Care Centers:
    - NCI data repositories (rare tumors/ clinically-relevant variants)
    - NCCR patient-linked registry
  - Searchable **catalog** of pediatric data, tools, & resources

Slides thanks to Dr Tony Kerlavage, NCI
Establish a **Federated Pediatric Cancer Data Ecosystem**:

- Underlying data science infrastructure
- Enhanced cloud-computing
- Services linking clinical, image, & molecular data
- Standards & tools for data interoperability
- Data repositories (e.g. **Pediatric Preclinical Data Commons**)
- Linked data (**National Childhood Cancer Registry**)

*Discovery Science ➔ Clinical Studies/Care ➔ Surveillance*
Develop Resources to Analyze & Use Data

NCI Activities:

- Develop or transfer **tools & pipelines** to NCI resources
- Automated curation of data *(e.g. Natural Language Processing)* for refining, scaling & real-world data capture
- **Computational methods** & tools – *annotate, integrate, translate data from multiple repositories or registries*  
  - Interpreting pathology images, patient reports & complex genotypes  
  - Identifying & validating molecular targets
- Interactive **data portals** – engagement & resource accessibility
- Pediatric **data model** & terminology **harmonization** *(PCDC)*
Create Comprehensive & Meaningful Data Sets for Discovery

NCI Activities:

- **Sequencing** & characterization:
  - Germline & tumor samples (*clinical trials: diagnosis, relapse: Pediatric MATCH, CNS/sarcomas*)
  - Pediatric pre-clinical models (*PDX, cell lines*)
  - Secondary cancers from survivors (CCSS)

- Explore **risk & susceptibility** in patients/ survivors:
  - Environmental exposures
  - Metabolomic profiling (*inherited predisposition*)

- Develop & characterize **cancer models** (*pediatric brain & solid cancers: cell lines, organoids*)
Moving Discoveries Into the Clinic

NCI Activities:

- Analyze molecular & clinical data to validate variants on FDA Relevant Molecular Targets List
- Establish & translate Rare Pediatric Tumor Cell Atlas data to the clinic *(MyPART)*
- Improve & enhance efficiency of clinical trials data:
  - Comprehensive central monitoring & safety reports
  - Increased quality of data *(self-reports, PRO)* collected
- Exploring genetic susceptibility to adverse events *(post treatment)*; & risks/benefits of proton therapy
- Molecular Characterization Protocol *(CNS, Soft tissue sarcomas, Rare Cancers)*
Childhood Cancer Data Catalog

CCDI Annual Symposium

A guide to finding childhood cancer programs, registries, and repositories

Slides thanks to Dr. Patrick Dunn, Frederick National Laboratory
Pediatric Cancer Data Resources
A Catalog of Data Resources for Childhood Cancer Research

A guide to relevant research programs, clinical trials, data sets and repositories to support the RACE for Children Act
An expanding list organized by data repository and funding opportunities
**PedcBioPortal for Integrated Childhood Cancer Genomics (pedcBio portal)**

[https://pedcbiportal.org](https://pedcbiportal.org)

Point of Contact: Ethan Cerami, cerami@hms.harvard.edu

### ABOUT THIS RESOURCE

PedcBioPortal is an NIH-funded cloud-based cancer visualization and analytics application to empower analysis of pediatric cancer data. It is a childhood cancer visualization tool modeled upon and partnered with the highly successful, adult-focused, dbioPortal, which was originally developed at Memorial Sloan-Kettering Cancer Center (MSK). PedcBioPortal provides access to high-level processed data types and supports the curation and cross-cancer integration of public, childhood cancer genomic information. The portal provides integration of public, pediatric cancer genomics datasets as well as "open science" initiatives integrated within the Kids First Data Resource Center as well as data from consortia-based efforts, including the Children's Brain Tumor Tissue Consortium (CBTTC), the Pediatric NeuroOncology Consortium (PNOC), the St. Baldrick's Pediatric Stand Up 2 Cancer Dream Team, and the Pediatric Precisional Testing Consortium (PPTC).

The PedcBioPortal for Childhood Cancer Genomics is an instance of dbioPortal (http://www.dbioportal.org) supporting the curation and pan-cancer integration of public, pediatric cancer genomics datasets as well as "open science" initiatives integrated within the Kids First Data Resource Center (https://kidsfirstdc.org), as well as data from consortia-based efforts including the Children's Brain Tumor Tissue Consortium (CBTTC - [https://cbttc.org](https://cbttc.org)), the Pediatric NeuroOncology Consortium (PNOC - [http://www.pnoc.org](http://www.pnoc.org)), the St. Baldrick's Pediatric Stand Up 2 Cancer Dream Team ([http://www.standup2cancer.com](http://www.standup2cancer.com)), and other initiatives.

### Resource Description

<table>
<thead>
<tr>
<th>DATA RESOURCE TYPE</th>
<th>Data Commons</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPECIALIZATION</td>
<td>Mixed Adult and Pediatric</td>
</tr>
</tbody>
</table>

### Resource Tools

<table>
<thead>
<tr>
<th>VISUALIZATION TOOLS</th>
<th>YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANALYTIC TOOLS</td>
<td>YES</td>
</tr>
</tbody>
</table>

### Resource Data Content Types

- Clinical Data
- Omics Data

### Data Access

- [API (INTERNAL)](https://www.pndcbioportal.org/swagger-u.html)
Browsing and Searching the Data Catalog

Sharing clinical care and research data generated by the pediatric cancer research community

Home Search Catalog Participating Resources About

Search Results

Search the Catalog

Advanced Search

Card View Table View

SORT BY Resource Name

Showing 1 - 10 of 29 Previous 1 2 3 Next

The Childhood Cancer Survivor Study

Update Date: 2021-11-03

Pl: Gregory T. Armstrong, M.D., MSCE.

Cases: Acute lymphocytic leukemia, Acute myeloid leukemia, Other Leukemia, Astrocytoma, Medulloblastoma/PNET, Other CNS

Malignancy: Hodgkin Lymphoma, Non-Hodgkin Lymphoma, Kidney Tumors, Neuroblastoma ...

Case Count: 25665

Description: The Childhood Cancer Survivor Study (CCSS) includes all participants with a confirmed diagnosis of cancer and 5 year survival, a cohort of 35,923 childhood cancer survivors diagnosed between 1970 and 1999. The CCSS cohort has been assembled through the efforts of 31 participating centers in the United States and Canada. CCSS is a resource in which to investigate current and future questions regarding consequences of therapy, genetic associations, disease processes and causation, interventions, a ...

Burkitt Lymphoma Genome Sequencing Project (BLGSP)

Update Date: 2021-11-03

Pl: TBD

CGCI
Looking for Resources in More Detail
Contribute to the Childhood Cancer Data Catalog

Contribute to the CCDI Data Catalog

NCI is interested in expanding this resource. Submit summaries about your data resource makes the existence known to a broader community and helps to promote the use of the data.

If you would like to include your resource in this data catalog, complete the summary submission template and send it to Childhood Cancer Data Initiative. Summaries quantify one or more data element values such as count, minimum, maximum, or average. A data element is a unit of data such as a disease diagnosis, or case age. Each assertion in a summary pertains to a data element and its value or set of values in a dataset.

For questions, please contact Childhood Cancer Data Initiative.
The NCI Molecular Targets Platform: Integrating Data to Catalyze Childhood Cancer Drug Development

CCDI Annual Symposium

Slides thanks to Dr. Deanne Taylor, Children’s Hospital of Philadelphia
Motivation for a Molecular Targets Platform

- The RACE for Children Act provides significant opportunities and challenges for developing new drugs for childhood cancer.
- The FDA maintains an evolving list of "molecular targets" relevant to childhood cancer (Pediatric Molecular Targets List [PMTL]).
- The RACE Act mandates that any drug being developed for an adult cancer indication be developed in children’s cancers if the mechanism of action aligns with the PMTL.
- The PMTL is a document disconnected from the wealth of data used to generate the list.
Proposed Solution

- Develop a node within the CCDI to provide the computational infrastructure for academics, industry-partners and advocates to adjudicate pediatric cancer drug development decisions

- Adopt the open-source Open Targets Portal for pediatric cancer drug development
  - [https://www.opentargets.org/](https://www.opentargets.org/)

- Developed a collaborative working group of multi-disciplinary investigators from the CCDI, NCI, FNL, and CHOP with kickoff in March 2021
Current Data Sources for Molecular Targets portal

- **Up to date**
  - Open Targets (Wellcome Sanger) Platform & Data
- **Complete**
  - TARGET Project Data (738 samples)
- **In process**
  - PPTC Data (738 samples)
- **Up to date**
  - Kids First Cancer Data (881 samples)
- **Up to date**
  - OpenPBTA (CBTN, PNOC) (2984 Samples)
- **GTEx Complete**
  - Adult reference data: GTEx (17382 samples), TCGA (10414 samples)

**Analysis and Harmonization Pipelines:**
Kids First DRC and the OpenPedCan Project

**Molecular Targets Database**

- Adds to Open Targets Database
- Harmonized Pediatric Data
- Preclinical Treatment Evidence
- Compatible Ontologies
- Pediatric PMTL

**Evidence visualized for potential pediatric cancer targets**

**Tables and visualizations for targets (genes) vs diseases across childhood cancer datasets, vs valuable healthy tissue profiles from GTEx.**

- Evidence visualized for potential pediatric cancer targets

**Delivers**

- Drug/Disease/Gene evidence
- Plots and visualizations
- Ontological categorizations
- Association scores
- PMTL Relationships
- Target Prioritizations

---

**Up to date**

- Kids First Cancer Data (881 samples)
- TCGA (10414 samples)
- GTEx Complete (17382 samples)
- OpenPBTA (CBTN, PNOC) (2984 Samples)

**Complete**

- TARGET Project Data (738 samples)

**In process**

- PPTC Data (738 samples)

---

**Up to date**

- Open Targets (Wellcome Sanger) Platform & Data

---

**GTEx Complete**

- Adult reference data: GTEx (17382 samples), TCGA (10414 samples)
Fully employing the Open Targets-designed platform

Search on diseases, genes/targets, drugs.
On the disease page, MTP provides a scored list of targets.

Curation on FDA Pediatric Relevant Molecular Target List (PMTL):
- (R) is "Relevant"
- (NR) is Not Relevant
- Blank=Unspecified

OT-specific scoring: needs to be updated for pediatric cancers.

Click on a target.
Clicking into ALK from the neuroblastoma page:

Each box on the evidence page represents sources. OpenPedCan analyses represent the federated and harmonized open source datasets in this project.

Evidence added from OpenPedCan

Clicking on box scrolls down to that section

GitHub: PediatricOpenTargets/OpenPedCan-analysis
Disease evidence. Could contain duplicate samples over different projects.

Live plots and matching downloadable data are generated from OpenPedCan data sources.

GTEx data
Can click into the gene name to get a wider view of the gene’s relationship to other diseases.
Information on ALK as a target across all diseases
ALK information across all OpenPedCan-analyzed histologies

Part of the OpenPedCan effort is to harmonize histologies through recent work with the Pediatric Cancer Data Commons and NCIT

Tumors with n<3 not displayed
What do we know about GPC2 as a potential target?
Disease evidence. Could contain duplicate samples over different projects.

GPC2 expression is increased in neuroblastoma versus GTEx samples.
Investigating CAR-T cell therapies for GPC2: cross-histology utility?

If therapies being developed against GPC2 in neuroblastoma are found safe, other cancers with similar profiles could also be targeted by the same therapy.
Exploring Somatic Alterations
Unspecified in the PMTL: **RUNX1**
(known as a fusion partner in AML)
AML expression vs GTEx data

*RUNX1*, a transcription factor, has higher overall expression versus GTEx samples.
Somatic Alterations of RUNX1 in AML

Fusions
Fusions by Gene

SNV by Gene

CNV by Gene

SNV by Variant
**Somatic Alterations of RUNX1 in AML**

<table>
<thead>
<tr>
<th><strong>Fusions</strong></th>
<th><strong>SNV by Gene</strong></th>
<th><strong>CNV by Gene</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fusions by Gene</td>
<td>SNV by Variant</td>
<td>CNV by Gene</td>
</tr>
</tbody>
</table>

### OpenPedCan Somatic Mutations

Somatic mutations associated with RUNX1 in pediatric acute myeloid leukemia. Source: OpenPedCan (v1.0), TARGET (v1.0). Kidd-First Neuroblastoma, OpenPTRA for the CRBN (v1.0), OncXLab (v1.0).

<table>
<thead>
<tr>
<th>Reciprocal exists</th>
<th>Gene A. amino</th>
<th>Gene B. amino</th>
<th>Gene A. Ensembl ID</th>
<th>Disease</th>
<th>PNTL</th>
<th>Dataset</th>
<th>Total alterations Over Patients in dataset</th>
<th>Total primary tumors mutated</th>
<th>Frequency in overall dataset</th>
<th>Total relapse tumors in dataset</th>
<th>Frequency in relapse tumors</th>
</tr>
</thead>
<tbody>
<tr>
<td>either gene kinase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RUNX1</td>
<td>AML</td>
<td></td>
<td></td>
<td>12/1441</td>
<td>6/1441</td>
<td>4.25%</td>
<td>2/1441</td>
<td>1.41%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RUNX1</td>
<td>AML</td>
<td></td>
<td></td>
<td>12/1441</td>
<td>6/1441</td>
<td>4.25%</td>
<td>2/1441</td>
<td>1.41%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RUNX1</td>
<td>AML</td>
<td></td>
<td></td>
<td>12/1441</td>
<td>6/1441</td>
<td>4.25%</td>
<td>2/1441</td>
<td>1.41%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RUNX1</td>
<td>AML</td>
<td></td>
<td></td>
<td>12/1441</td>
<td>6/1441</td>
<td>4.25%</td>
<td>2/1441</td>
<td>1.41%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RUNX1</td>
<td>AML</td>
<td></td>
<td></td>
<td>12/1441</td>
<td>6/1441</td>
<td>4.25%</td>
<td>2/1441</td>
<td>1.41%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RUNX1</td>
<td>AML</td>
<td></td>
<td></td>
<td>12/1441</td>
<td>6/1441</td>
<td>4.25%</td>
<td>2/1441</td>
<td>1.41%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RUNX1</td>
<td>AML</td>
<td></td>
<td></td>
<td>12/1441</td>
<td>6/1441</td>
<td>4.25%</td>
<td>2/1441</td>
<td>1.41%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RUNX1</td>
<td>AML</td>
<td></td>
<td></td>
<td>12/1441</td>
<td>6/1441</td>
<td>4.25%</td>
<td>2/1441</td>
<td>1.41%</td>
</tr>
</tbody>
</table>
Both AML and ALL have higher expression of *RUNX1* vs other pediatric cancers (due to translocations)
Coming Soon

RNA-seq differential expression for normal vs ped. cancers

More Views/Analyses Planned:
Cancer expression profile heatmaps
Oncoprints views
Splicing heatmaps
Germline variation views
Others as well!

Incoming New Data:
Preclinical data
Whole Exome Sequencing
Whole Genome Sequencing
Targeted Panels
RNA-seq
Methylation Data
Single-Cell Data
Protein Expression
Imaging, Clinical data, Pathology.

Email: ncichildhoodcancerdatainitiative@mail.nih.gov
Summary

- Childhood cancer data is being harmonized through state-of-the-art open-source pipelines.
- Current data will be augmented by harmonizing more NCI-funded and partnered sources in the coming months.
- Interface will be made publicly available for strategic research into childhood cancer therapies, and include negative as well as positive results.
- Additions to the OT platform have yielded childhood-cancer specific views and utility to the research community.
- The Molecular Targets platform provides a target-gene-therapy platform to display these data.
Soliciting feedback!

- Get in touch! Email the team through ncichildhoodcancerdatainitiative@mail.nih.gov

- Check out our code, pipelines in our OpenPedCan GitHub repositories under the PediatricOpenTargets project: https://github.com/PediatricOpenTargets
Credits and Special Thanks

- This has been a productive partnership between:
  - The National Cancer Institute
  - Frederick National Laboratories
  - The Children’s Hospital of Philadelphia

- Special Thanks for help in accelerating this project:
  - The Wellcome Trust
  - The Open Targets Platform Team Members
  - OpenPBTA: the Childhood Cancer Data Lab - Alex’s Lemonade Stand

- Thank you to the following teams for open software, data, support:
  - Kids First & the KF DRC
  - Seven Bridges Genomics (CAVATICA)
  - TARGET
  - CBTN and PNOC
  - Pediatric Cancer Data Commons
  - GTEx
  - OncoKB
  - COSMIC
  - Monarch (MONDO)
  - EMBL-EBI and EFO
  - Jackson Labs
  - PPTC/PIVOT
  - PedcBioPortal
  - Python, R, Bioconductor

Email: nci_childhood_cancer_data_initiative@mail.nih.gov
learn from every child.
engage with CCDI

• Visit the [CCDI website](#) to learn more (on the NCI website)
• Review the [BSA CCDI Working Group final report](#)
• Receive [email updates](#) from NCI on the CCDI
• Contact the [CCDI](#) with questions about engagement, ongoing activities, or funding opportunities
• Look for information about webinars starting in 2022
Thank you!
Questions?