



JAPAN ACADEMIC WORKSHOP

Friday, 17 November | 1:00pm -5:15pm



D4CG

Transforming Human Health Through Data
The Pediatric Cancer Data Commons

Samuel Volchenboum, MD, PhD



A perfect storm

Rare diseases, siloed data,
manual processes, lack of standards,
difficult research questions



Pediatric Cancer Data Commons (PCDC)

1. Provides a hub for researchers across **most pediatric cancers**
2. Facilitates **cross-disease research**
3. Develops **common core data dictionaries**
4. Creates **common governance structures**
5. Builds on an **active international network** of researchers
6. **Educates, trains, and increases data literacy** of the community
7. Sources data from **clinical trials, registries**, and (soon) **EHRs**



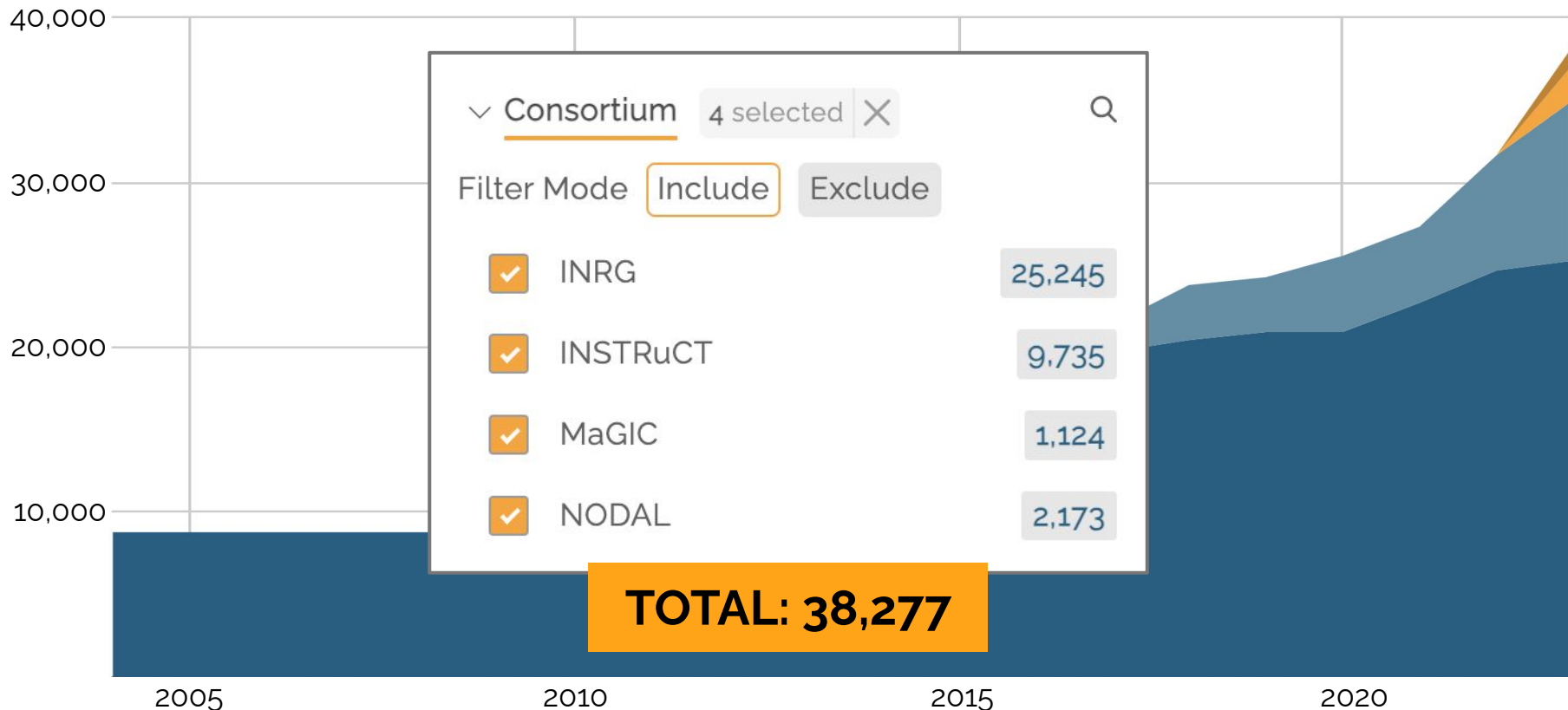
DATA FOR THE COMMON GOOD

We build communities, platforms, and ecosystems
that maximize the potential of data
to drive discovery and improve human health.



Cases integrated to date

INRG NODAL
INSTRuCT MaGIC



Japan data in the PCDC

The screenshot shows the Pediatric Cancer Data Commons (PCDC) interface. The top navigation bar includes "About PCDC" and "Our Sponsors". The main header displays "PEDIATRIC CANCER DATA COMMONS". On the left, a "Filters" sidebar shows "Data Contributor" with a list of institutions, including JCCG (checked) with 970 subjects. A red arrow points to the JCCG checkbox. The main content area shows "Summary View" and "Subjects 970". Below this, four horizontal bar charts display demographic data: Sex (Unknown: 442, Male: 298, Female: 230), Race (Not Reported: 441, Asian: 309, Unknown: 150), Ethnicity (Not Hispanic or Latino: 771, Unknown: 199), and Consortium (INRG: 970). At the bottom, two researchers are listed: Akira Nakagawara (Saga Medical Center Koseikan) and Takehiko Kamijo (Saitama Cancer Center), with Miki Ohira (Saitama Cancer Center) listed below.

Demographic	Category	Count
Sex	Unknown	442
	Male	298
	Female	230
Race	Not Reported	441
	Asian	309
	Unknown	150
Ethnicity	Not Hispanic or Latino	771
	Unknown	199
Consortium	INRG	970

Researchers:

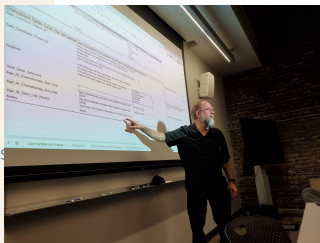
- Akira Nakagawara, Saga Medical Center Koseikan
- Takehiko Kamijo, Saitama Cancer Center
- Miki Ohira, Saitama Cancer Center

Building a data commons—the D4CG way

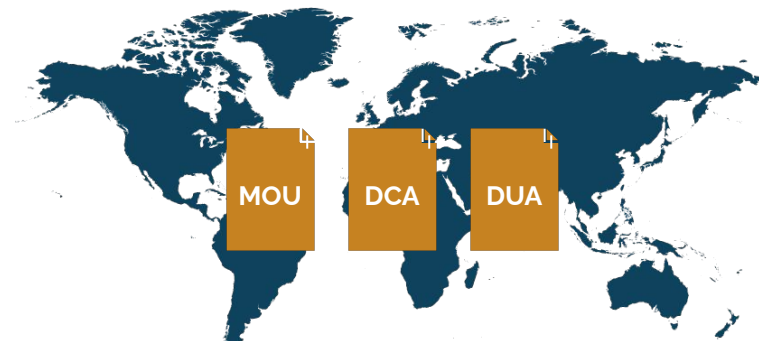
Step 1: Establish a **consortium**



- PCDC Master Data Dictionary
- Acute Lymphoblastic Leukemia (ALL)
- Acute Myeloid Leukemia (AML)
- Central Nervous System Tumors (CNS)
- Ewing Sarcoma (EWS)
- Germ Cell Tumors (GCT)
- Hodgkin Lymphoma (HL)
- Neuroblastoma (NBL)
- Non-Rhabdomyosarcoma Soft Tissue Sarcoma (NRSTS)
- Osteosarcoma (OS)
- Predisposition
- Retinoblastoma (RB)
- Rhabdomyosarcoma (RMS)

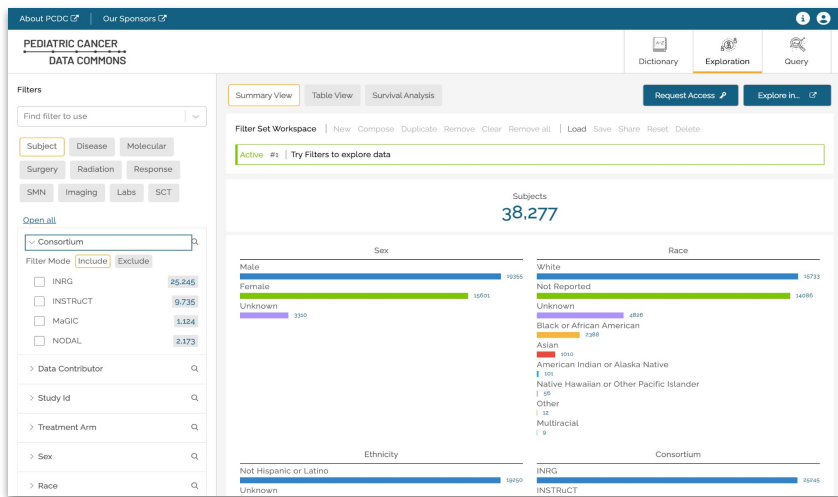


Step 2: Develop and deploy **data operations**



Step 3: Establish and implement **governance**

Building a data commons—the D4CG way



Step 4: Develop and deploy the **technical infrastructure**

PCDC Documentation

- PCDC Data Portal Acceptable Use Policy
- PCDC Data Portal Statistical Manual
- PCDC Data Portal User Guide
- PCDC Data Access and Governance
- PCDC Project Request Forms
- Privacy Notice
- PCDC Terms and Conditions
- GEARBOX Terms and Conditions
- Working with PFB Files



Pediatric Cancer Data Commons Project Request Forms

Before submitting a project request form, please review the PCDC Data Access and Governance document to learn more about the data request process.



ational Soft Tissue Sarcoma Consortium (INSTRuCT), please and complete and e-mail the form as indicated.

ational Neuroblastoma Risk Group (INRG), please click [here](#) and e-mail the form as indicated.

Step 5: **Socialize and sustain** the commons



THE UNIVERSITY OF CHICAGO



DATA FOR THE COMMON GOOD



Japan Academic Workshop
<http://sam.am/Japan2023>

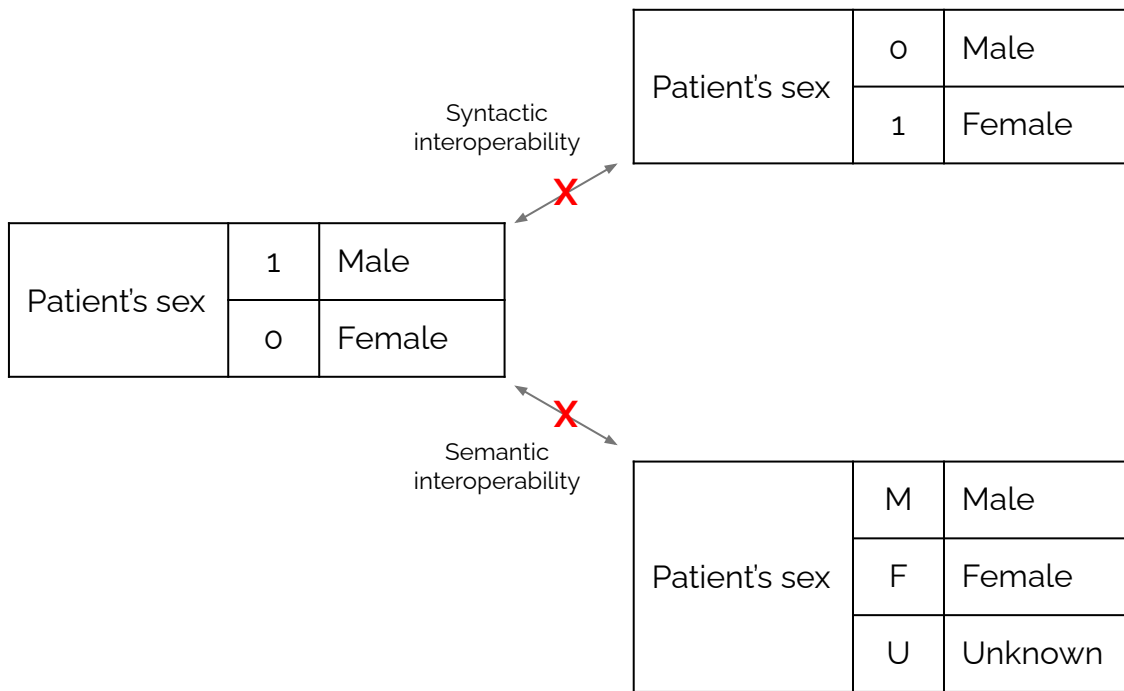
<https://portal.pedscommons.org>
commons.uchicago.edu

Engage stakeholders

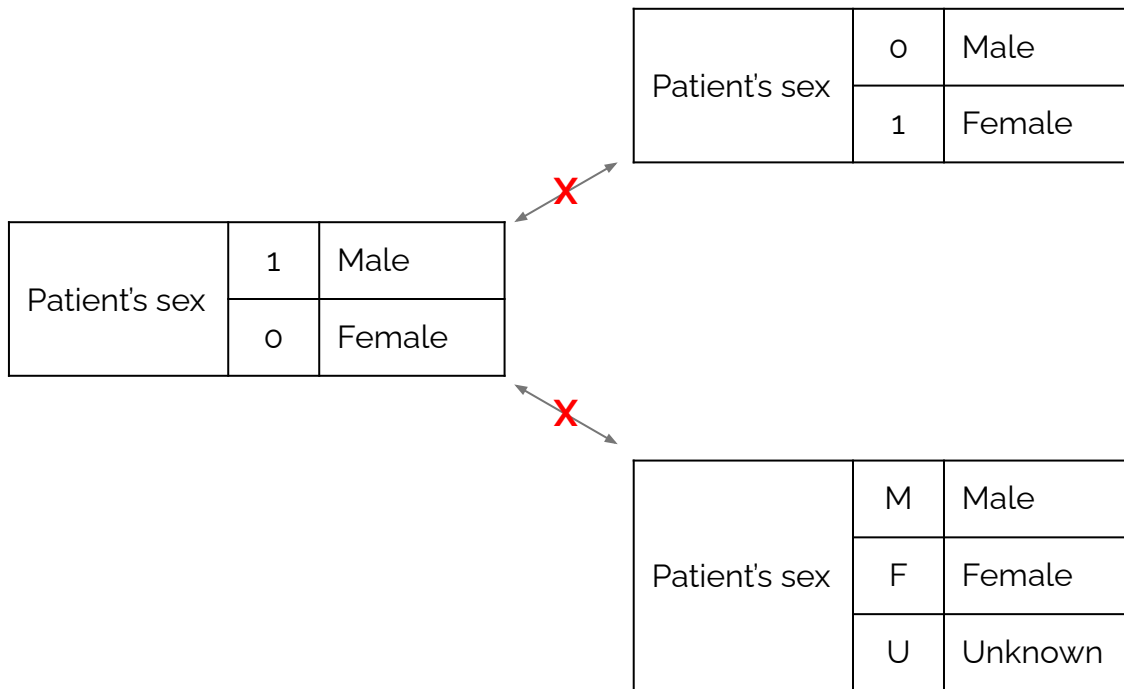
We employ a “big tent” philosophy.



The importance of data standards



The importance of data standards



Sex	
Description	NCIt Code
Male	C20197
Female	C16576
Unknown	C17998

The importance of data standards

Patient's sex	1	Male
	0	Female

Patient's sex	F	Female
	U	Unknown

Sex	
Description	NCIt Code
Male	C20197
Female	C16576
Unknown	C17998

Building a consensus data dictionary

Off Protocol Therapy/Study: one row per subject per off protocol therapy/study per reason off						
AGE_OFF	Number	Age in Days When Off Protocol Therapy or Study	C172678	Age of subject (in days) when		
DISEASE_PHASE	Code	Disease Phase	C168878	The stage or period of an in	Initial Diagnosis	C156813
					Relapse	C38155
DISEASE_PHASE_NUMBER	Number	Disease Phase Number	C173258	The number of the disease		
COURSE	Code	Protocol Treatment Course	C168807	The type of protocol treatm	Prephase	C168826
					Induction	C158876
					Intensification	C173105
					Consolidation	C15679
					Stem Cell Transplant Conditioning	C168794
					Maintenance	C15688
					Palliative Treatment	C15292
					Other	C17649
COURSE_NUMBER	Number	Course Number	C166235	The number assigned to a c		
OFF_TYPE	Code	Off Protocol Therapy or Study	C173256	The code used to designate	Protocol Therapy	C173257
					Study	C29851
REASON_OFF	Code	Off Protocol Therapy or Study Reason	C173519	The reason a subject went	Death	C93546
					Lost to Follow-Up	C70740
					Completion of Planned Therapy	C168935
					Physician Decision	C48250
					Withdrawal of Consent	C48271
					Subject/Guardian Refused Further Treatment	C168934
					Disease Progression	C35571
					Relapse	C38155
					Adverse Event	C41331
					Secondary Malignancy	C4968
					Other	C17649
					Unknown	C17998
					Not Reported	C43234

<http://sam.am/datadictionaries>

Building a consensus data dictionary

	Age of subject (in days) when		
A	The stage or period of an in	Initial Diagnosis	C156813
D		Relapse	C38155
D	The number of the disease		
C	The type of protocol treatm	Prephase	C168826
		Induction	C158876
		Intensification	C173105
		Consolidation	C15679
		Stem Cell Transplant Conditioning	C168794
		Maintenance	C15688
		Palliative Treatment	C15292
		Other	C17649
C	The number assigned to a c		
O	The code used to designat	Protocol Therapy	C173257
		Study	C29851
R	The reason a subject went	Death	C93546
		Lost to Follow-Up	C70740
		Completion of Planned Therapy	C168935
		Physician Decision	C48250
		Withdrawal of Consent	C48271
		Subject/Guardian Refused Further Treatment	C168934
		Disease Progression	C35571
		Relapse	C38155
		Adverse Event	C41331
		Secondary Malignancy	C4968
		Other	C17649
		Unknown	C17998
		Not Reported	C43234

<http://sam.am/datadictionaries>



Data collection is highly localized

<input type="checkbox"/> Male <input type="checkbox"/> Female <input type="checkbox"/> Unknown <i>English</i>	<input type="checkbox"/> 男 <input type="checkbox"/> 女性 <input type="checkbox"/> わからない <i>Japanese</i>
<input type="checkbox"/> ذكر <input type="checkbox"/> أنثى <input type="checkbox"/> مجهول <i>Arabic</i>	<input type="checkbox"/> Homme <input type="checkbox"/> Femelle <input type="checkbox"/> Inconnue <i>French</i>
<input type="checkbox"/> Männlich <input type="checkbox"/> Weiblich <input type="checkbox"/> Unbekannt <i>German</i>	<input type="checkbox"/> 男性 <input type="checkbox"/> 女性 <input type="checkbox"/> 未知 <i>Chinese</i>

CDE mapping solves the localization problem

Male
 Female
 Unknown
English

男
 女性
 わからない
Japanese

ذكر
 أنثى
 مجهول
Arabic

Homme
 Femelle
 Inconnue
French

Männlich
 Weiblich
 Unbekannt
German

男性
 女性
 未知

C20197
C16576
C17998

The screenshot shows the NCI Thesaurus web interface. At the top, it says 'NIH NATIONAL CANCER INSTITUTE' and 'www.cancer.gov'. Below that, it says 'NCI Term Browser' and 'EVS Enterprise Vocabulary Services'. There are tabs for 'Terminologies', 'Value Sets', and 'Mappings'. A search bar is present with a search button. Below the search bar, there are radio buttons for 'Contains', 'Exact Match', and 'Begins With', and checkboxes for 'Name', 'Code', 'Property', and 'Relationship'. There are also links for 'Back to search results' and 'Advanced Search'. Below the search bar, there are links for 'Hierarchy', 'Value Sets', 'Maps', and 'Help'. A 'Quick Links' dropdown menu is also visible. The main content area shows the entry for 'Male (Code C20197)'. It has tabs for 'Terms & Properties', 'Synonym Details', 'Relationships', 'Mappings', and 'View All'. The 'Terms & Properties' tab is selected. The entry shows the 'Preferred Name' as 'Male', the 'Definition' as 'A person who belongs to the sex that normally produces sperm. The term is used to indicate biological sex distinctions, cultural gender role distinctions, or both.', the 'DISC Definition' as 'A person who belongs to the sex that normally produces sperm. The term is used to indicate biological sex distinctions, cultural gender role distinctions, or both. (NCI)', the 'Label' as 'Male', the 'NCI Thesaurus Code' as 'C20197', the 'NCI Metathesaurus Link' as 'C0086582', and 'Synonyms & Abbreviations' as 'Human, Male', 'M', 'male', 'Male', and 'MALE'.

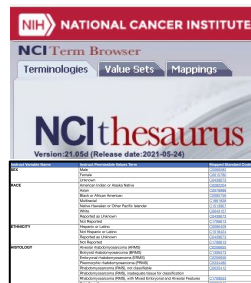
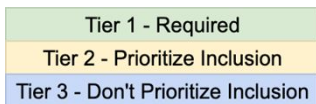
PCDC data dictionaries

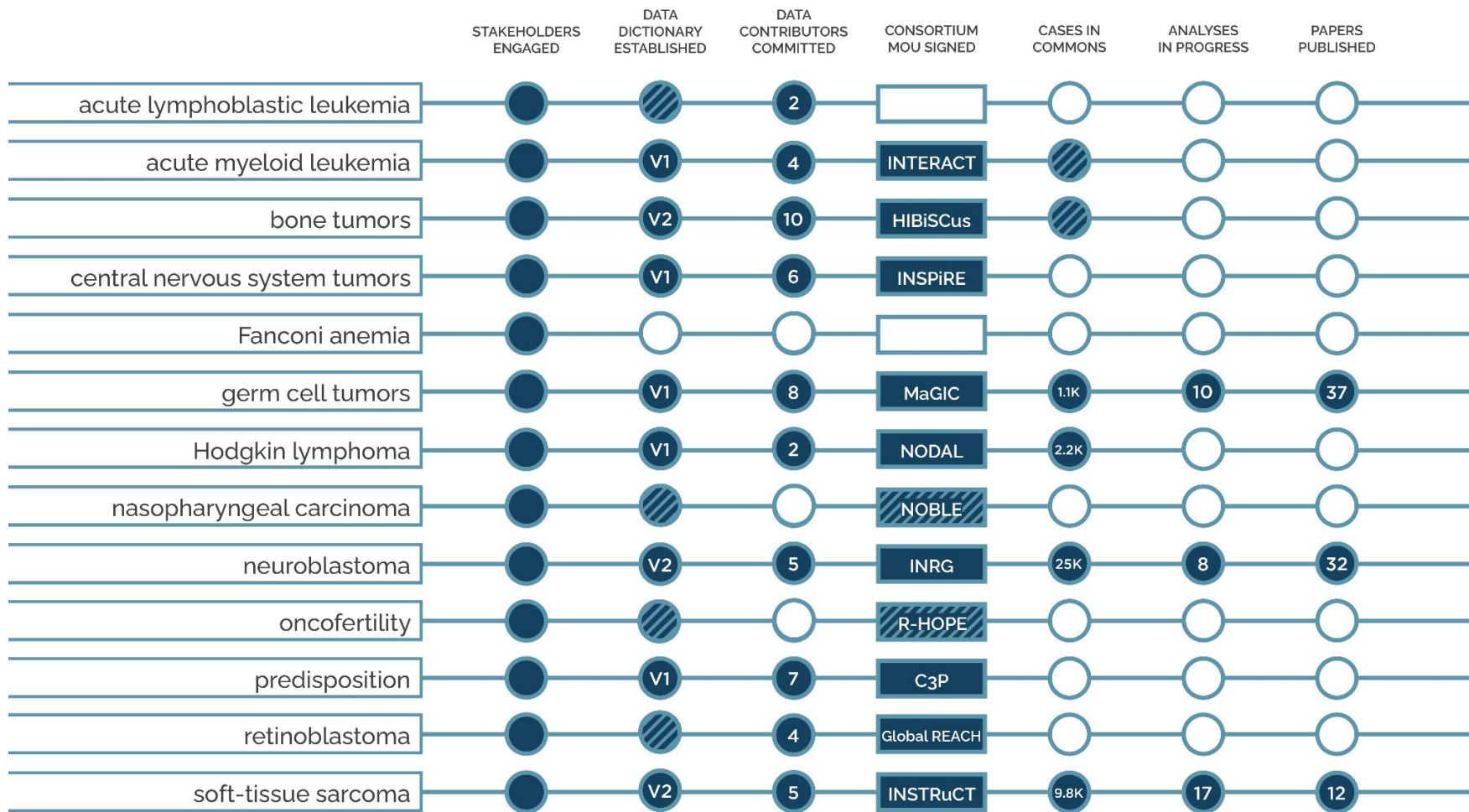
- PCDC Master Data Dictionary
- Acute Lymphoblastic Leukemia (ALL)
- Acute Myeloid Leukemia (AML)
- Central Nervous System Tumors (CNS)
- Ewing Sarcoma (EWS)
- Germ Cell Tumors (GCT)
- Hodgkin Lymphoma (HL)
- Neuroblastoma (NBL)
- Non-Rhabdomyosarcoma Soft Tissue Sarcoma (NRSTS)
- Osteosarcoma (OS)
- Retinoblastoma (RB)
- Rhabdomyosarcoma (RMS)

<https://commons.cri.uchicago.edu/data-dictionaries/>

Getting data into the PCDC

Balloted Dictionary → Tiering → NCI Coding → Harmonization → Quality Control → Ingestion into PCDC





PCDC worldwide participation

Worldwide Data Use Agreements

US - 8 master agreements
(+21 addenda/projects)

Non-US - 6 master agreements
(+5 addenda/projects)

Worldwide Data Contributor Agreements

US - 7 master agreements
(+18 addenda)

Non-US - 13 master agreements
(+12 addenda)

N. America

CBTN
COG
DFCI
IDIPGR
NRG
PNOC
RBTC
St. Jude



Europe

AIEOP
CCLG
COSS
CRCTU
EEC
EpSSG
EuPAL
GPOH
GEIS
GSF-GETO
ISG
MRC
NCRI
SIOPE
SIOPEN
SSG
SFCE
UNICANCER



Asia

JCCG
SIOPEN
EpSSG
COG

S. America

SOBOPE
EpSSG
GALOP
GLATO

Oceania

COG
EpSSG



New cancer-related consortia

post transplant
lymphoproliferative
disease

acute
lymphocytic
leukemia

acute myeloid
leukemia

INTERACT

bone tumors
(OS and EWS)

HIBiScus

Fanconi
anemia

central nervous
system tumors

INSPIRE

germ cell
tumors

MaGIC

Hodgkin
lymphoma

NODAL

neuroblastoma

INRG

nasopharyngeal
carcinoma

NOBLE

oncofertility

**Reproductive
HOPE**

Langerhans cell
histiocytosis

cancer
predisposition

C3P

retinoblastoma

Global REACH

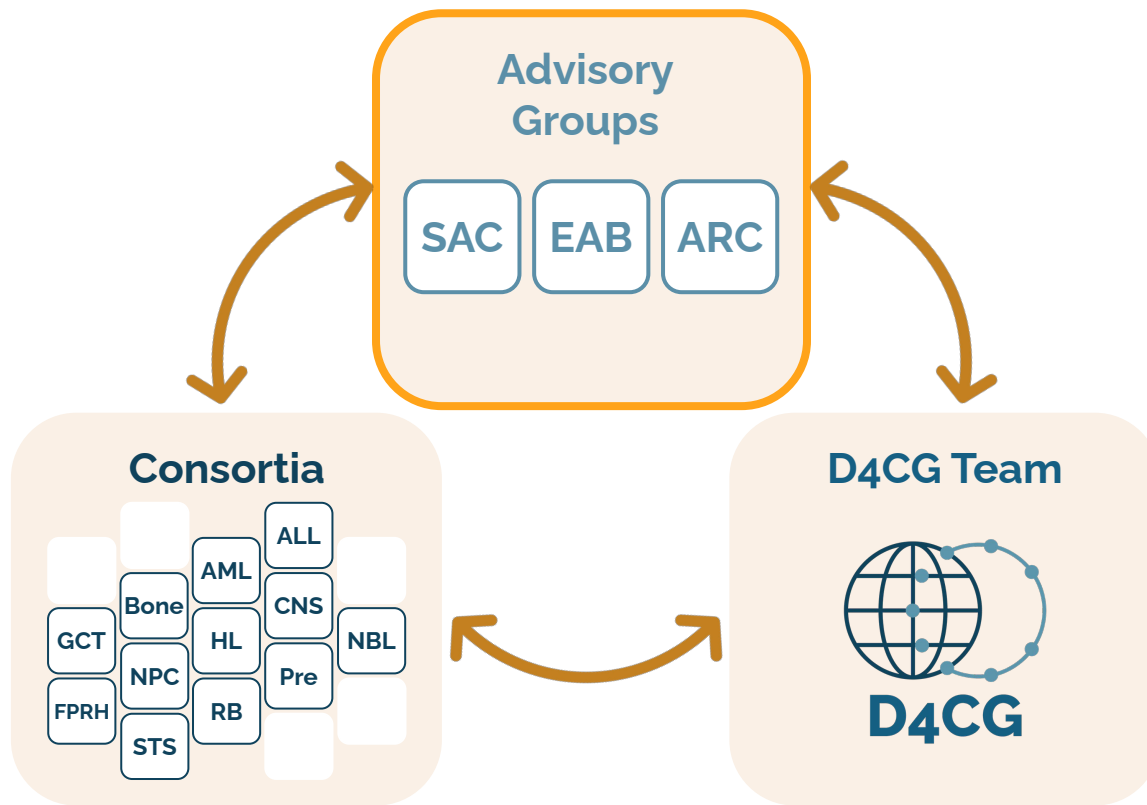
soft-tissue
sarcoma

INSTRuCT

myelodysplastic
syndrome



PCDC advisory groups structure



PCDC advisory groups structure



Daisuke Tomizawa

Head, Division of Leukemia and Lymphoma
National Center for Child Health and Development
Tokyo, Japan

co-chair

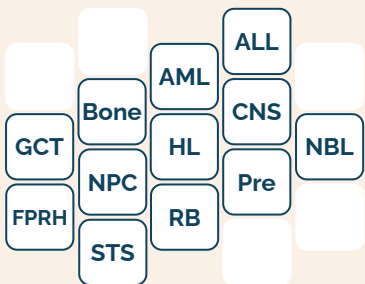
SAC

EAB

ARC

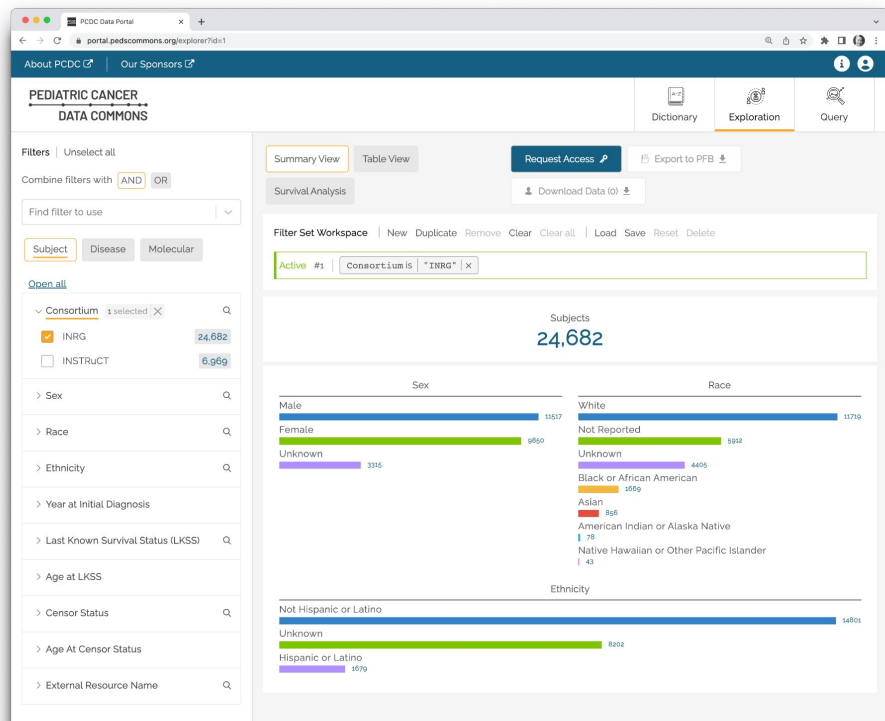
Advisory
Groups

Consortia



D4CG Team





<http://portal.pedscommons.org>

The screenshot displays the PCDC Data Commons interface. On the left, a filter sidebar is visible with categories like Subject, Disease, and Molecular. The main area shows a 'Filter Set Workspace' with an active filter for 'Consortium is "INRG"'. Below this, a summary view indicates 24,682 subjects. Three horizontal bar charts are shown: Sex (Male: 11,917, Female: 9,850, Unknown: 3,315), Race (White: 11,917, Not Reported: 9,850, Unknown: 3,315, Black or African American: 78, Asian: 43, American Indian or Alaska Native: 43), and Ethnicity (Not Hispanic or Latino: 11,917, Unknown: 9,850, Hispanic or Latino: 3,315). A modal window is open over the 'Stage' filter, showing 4 selected options: Stage 1 (4,122), Stage 2a (1,279), Stage 2b (1,777), and Stage 3 (3,700). Stage 4 is unselected with 11,413 subjects.

<http://portal.pedscommons.org>

PCDC Data Commons

PEDIATRIC CANCER DATA COMMONS

Summary View | Table View | Request Access | Export to PFB | Download Data

Filter Set Workspace

Active #1 | Consortium is "INRG" X

Subjects: 24,682

Sex: Male (11,917), Female (9,850), Unknown (3,315)

Race: White (9,741), Not Reported (4,544), Unknown (3,315), Black or African American (2,764), Asian (1,781), American Indian or Alaska Native (43)

Ethnicity: Not Hispanic or Latino (11,917), Unknown (3,315), Hispanic or Latino (2,799)

Filters: Consortium is "INRG" X

PCDC Data Commons

PEDIATRIC CANCER DATA COMMONS

Summary View | Table View | Request Access | Export to PFB | Download Data

Filter Set Workspace

Active #1 | Consortium is "INRG" X | AND | Stage is any of "Stage 1", "2", "3", "4" X | AND | Stage System is "INSS" X

Subjects: 10,878

Sex: Male (4,122), Female (4,544), Unknown (2,197)

Race: White (4,122), Not Reported (4,544), Unknown (2,197), Black or African American (2,764), Asian (1,781), American Indian or Alaska Native (43)

Save changes to the current Filter Set

Name: INSS Stage 1, 2a, 2b, 3

Description: Describe the Filter Set (optional)

Filters: Consortium is "INRG" AND Stage System is "INSS" AND Stage is any of "Stage 3", ...

Back to page | Save changes

Stage 4 selected

- Stage 1: 4,122
- Stage 2a: 1,279
- Stage 2b: 1,777
- Stage 3: 3,700
- Stage 4: 11,413

<http://portal.pedscommons.org>

Survival type: Overall Survival

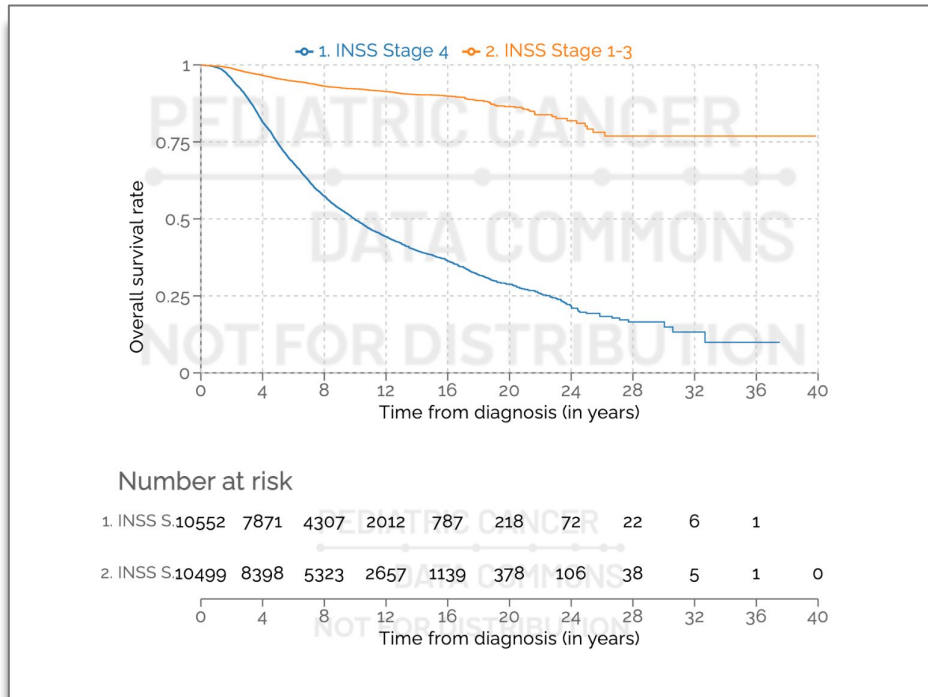
Start time (year): 0

End time (year): Optional; max value if left blank

Time interval (year): 4

Select Filter Set to analyze: Add

- > 1. INSS Stage 4 10552/11413 X
- > 2. INSS Stage 1-3 10499/10878 X



<http://portal.pedscommons.org>

Cohort discovery - linking to Genomic Data Commons

The screenshot displays the Pediatric Cancer Data Commons interface. A modal window titled "Find Cohort in An External Data Commons" is open, showing options to select data commons and genomic data commons. The background shows a filter set workspace with 38,277 subjects and various demographic filters.

The screenshot displays the National Cancer Institute GDC Data Portal interface. It shows a list of projects with columns for Case ID, Project, Primary Site, Gender, Files, and Available Files per Data Category. A table of projects is visible, including TARGET-30-PANBIOK, TARGET-30-PANBIV, and others.

Direct connection to other data commons
(Genomic Data Commons, Gabriella Miller Kids First)

<http://portal.pedscommons.org>

Engagement / Education

Education

- Promoting data literacy through training
- Young / early stage investigator seminars

Engage the community

- Webinars
- Thought leadership
- Live events
- Patient and family involvement
- International advisory board



sam.am/pcdcresearch

A	B		C	D			
1	INRG Research		Click on a description to view the original project proposal.				
2	INRG #	Principal Investigator	Description	Project Type	Status	Publication	Presentation
3	2022-04	Mallory Taylor Thomas Cash Wendy London Julie Park Meredith Irwin	Outcomes for patients aged 12-18 months with Stage M MYCN non-amplified neuroblastoma and unfavorable biologic features (Mixed Phenotype Toddlers)	Investigator	In Progress		Presented at ANR 2022
4	2022-03	Hanxiao Yu Xingda Zhan Mark Appelbaum Gudrun Schleiermacher statistician to be named	Prognostic impact of segmental chromosome alterations in high-risk neuroblastoma patients on immunotherapy: A report from the International Neuroblastoma Risk Group (INRG) project	Investigator	In Progress		
5	2022-02	Boris Decarolis Wendy London Susan Cohn Andrew Pearson	Survival of patients with low-, intermediate-, or high-risk neuroblastoma over a 35 year period	Investigator	In Progress		
6	2022-01	Wendy London Ramya Ramanujachar Kavitha Srivatsa Paola Angelini	Neuroblastoma in adolescents and adults- a study of clinical and biological features and outcomes	Investigator	In Progress		
7	2021-01	Kevin Campbell Pei-Chi Kao Arlene Naranjo Takehiko Kamijo Ramya Ramanujachar Wendy London Steven DuBois	Clinical and Biological Features Predictive of Survival After Relapse of Stage MS Neuroblastoma: A Report From the International Neuroblastoma Risk Group Project	Investigator	Published	Clinical and biological features prognostic of survival after relapse or progression of INRGSS stage MS pattern neuroblastoma: A report from the International Neuroblastoma Risk Group (INRG) project. <i>Pediatr Blood Cancer.</i> 2023 Feb;70(2):e30054. Epub 2022 Oct 31. doi: 10.1002/psc.30054.	
8	2020-03	Riyue Bao Stefani Spranger Kyle Hernandez Yuanyuan Zha Peter Pytel Jason Luke Thomas Gajewski Samuel Volchenbom Susan Cohn	Validation of a T-cell inflammatory signature and outcomes in patients with neuroblastoma	Investigator	Published	Immunogenomic determinants of tumor microenvironment correlate with superior survival in high-risk neuroblastoma. <i>J Immunother Cancer.</i> 2021 Jul;9(7):e002417. doi: 10.1136/jitc-2021-002417. PMID: 34272305; PMCID: PMC8287618.	

GEARBOX

Clinical trials matching

GEARBOX ABOUT GEARBOX

PATIENT INFORMATION

Demographics

What is the patient's current age (in years)?

What is the patient's biological sex?

Male Female

Disease

What is the patient's current diagnosis?

Does the patient currently have, or have they in the past had, refractory disease?

Yes No Not sure

Is the patient's disease currently refractory?

Yes No Not sure

Does the patient currently have, or have they in the past had, confirmed or suspected relapse disease?

Yes No Not sure

What is the patient's ECOG score?

OPEN TRIALS

Matched (1)

20-489

Title
A Study of Codrituzumab in Children and Young Adul...

Undetermined (1)

H-47757 AGAR

Title
Interleukin-15 Armored Glypican 3-specific Chimeric...

Unmatched (5)

RHM CHI0811

Title
Phase I Study of 131-I mIBG Followed by Nivolumab ...

NCI-2021-00913

Title
Testing the Combination of Two Immunotherapy Dru...

PATIENT INFORMATION



Demographics



What is the patient's current age (in years)?

13

What is the patient's biological sex?

 Male Female

Disease



What is the patient's current diagnosis?

Embryonal Carcinoma



Does the patient currently have, or have they in the past had, refractory disease?

 Yes No Not sure

Is the patient's disease currently refractory?

 Yes No Not sure

OPEN TRIALS

Matched (1)



20-489



Title

A Study of Codrituzumab in Children and Young Adults With Solid Tumors and Have Not Responded to Treatment or Have Come Back After Treatment

Description

The purpose of this study to find out whether codrituzumab is a safe treatment that causes few or mild side effects in children and young adults who have solid tumors that express the protein GPC3. The researchers also want to study the way codrituzumab is absorbed, distributed, and cleared from the body.

Locations

- Cincinnati Children's Hospital Medical Center
- Memorial Sloan Kettering Cancer Center

Link

- [ClinicalTrials.gov](https://clinicaltrials.gov)

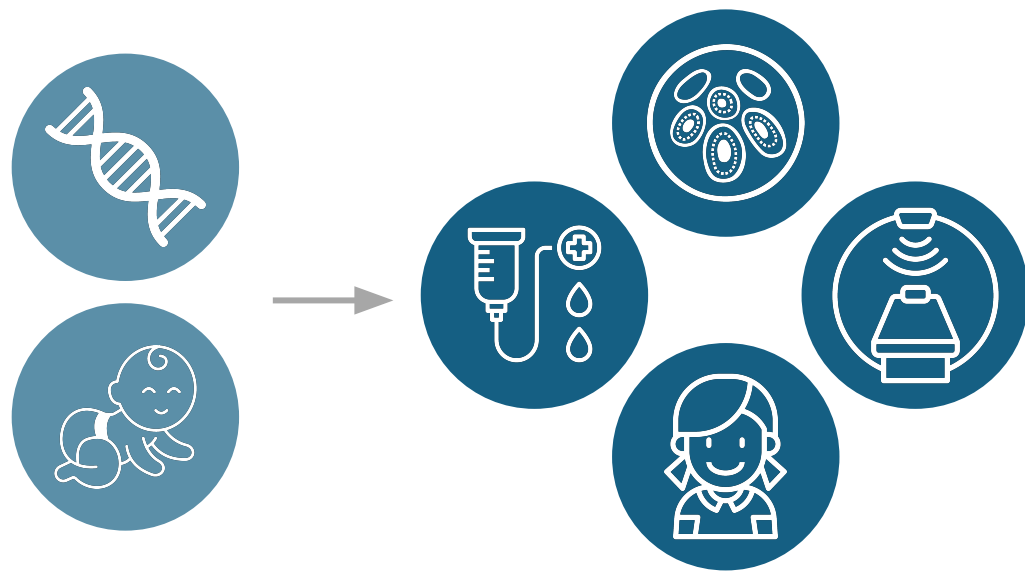
Connecting the dots across the patient's lifetime



pediatric, AYA, or
adult cancer



Connecting the dots across the patient's lifetime

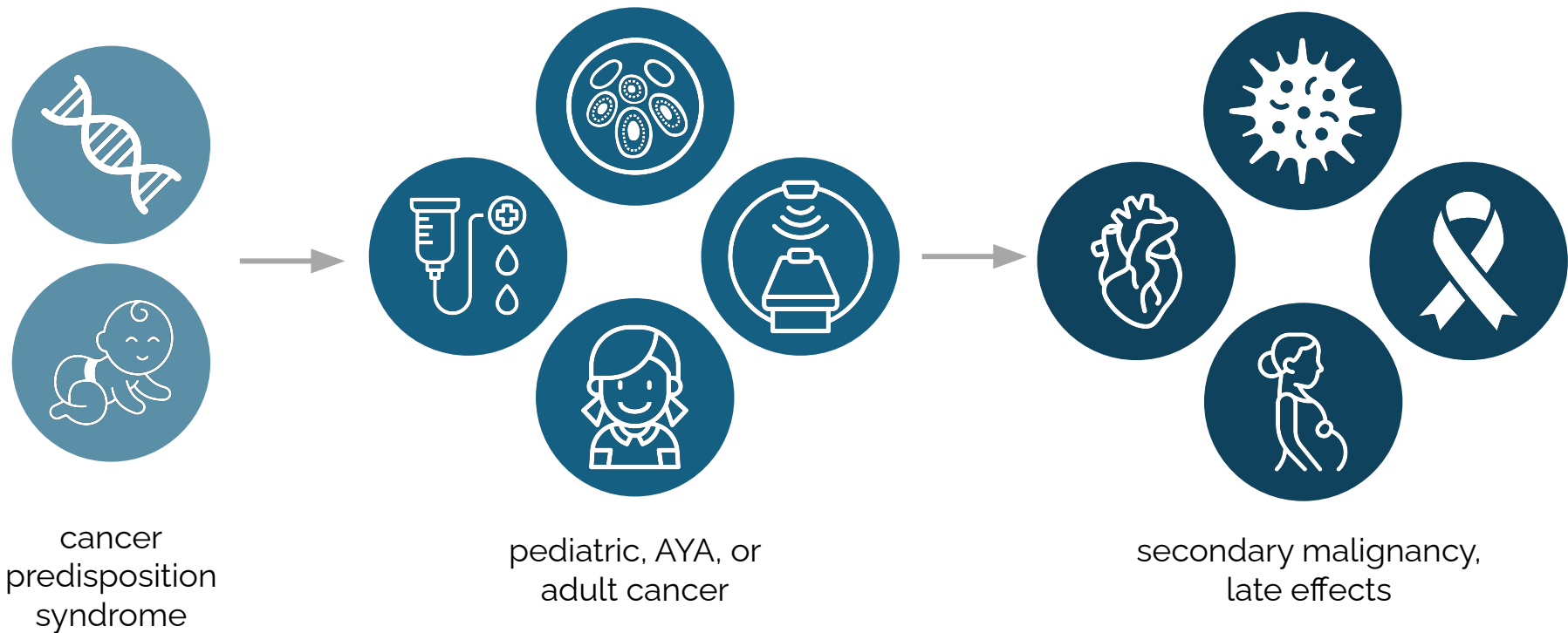


cancer
predisposition
syndrome

pediatric, AYA, or
adult cancer



Connecting the dots across the patient's lifetime



Why use the D4CG model for other rare diseases?



Lack of data hinders research for rare diseases.



Lack of community consensus and **deficient data standards** make data aggregation difficult or impossible.



Researchers cannot **make progress and improve patient outcomes** without high-quality data.

We envision a world where **access to high-quality data** is never a barrier to **improving human health**.



food allergies

PEDIATRIC CANCER
DATA COMMONS

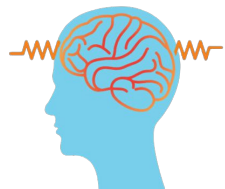
pediatric cancer



monogenic
diabetes



Crohn's disease



epilepsy



Sociome



other rare diseases



...and more



THE UNIVERSITY OF
CHICAGO



DATA FOR THE
COMMON GOOD

cdisc

Japan Academic Workshop
<http://sam.am/Japan2023>

<https://portal.pedscommons.org>
commons.uchicago.edu

A model for transforming human health



- Studying rare diseases requires **collaboration and sharing**
- Data sharing must be built on a foundation of **trust and consensus**
- **Connecting disparate data types** and sources enriches research
- **Consensus data standards** are critical for the success of national and international data ecosystems – allowing aggregation across trials and diseases
- **Early adoption of data standards** and consideration for the **lifecycle of the data** are critical to accelerating progress and discovery



Do research with the PCDC and stay connected

portal.pedscommons.org

Explore the PCDC Data Portal

sam.am/datadictionaries

Access data dictionaries

slv@uchicago.edu

suzi@uchicago.edu

Sign up for our email list!

sam.am/D4CGnews

Sustainability is key

Thank you to these funders for supporting our work in FY23 and/or FY24.



United States
Department of the Interior



Aileen S. Andrew
Foundation



Sarah Jane Adicoff Endowment for
Research in Rhabdomyosarcoma



DATA FOR THE
COMMON GOOD



Japan Academic Workshop
<http://sam.am/Japan2023>

<https://portal.pedscommons.org>
commons.uchicago.edu

Thank you!

consortium members

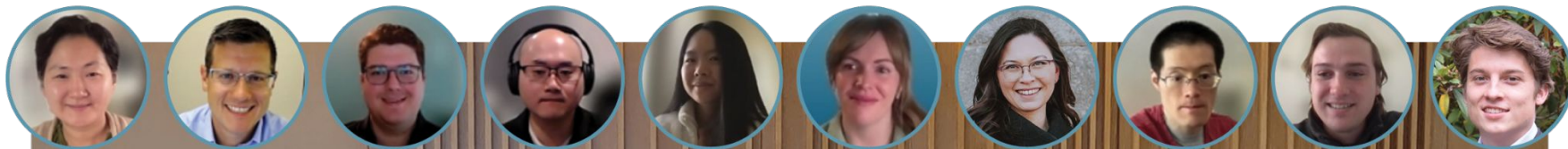
collaborators

data portal users

advisory group members

funders

patients and families



None of this would be possible without YOU.



THE UNIVERSITY OF
CHICAGO



DATA FOR THE
COMMON GOOD

cdisc

Japan Academic Workshop
<http://sam.am/Japan2023>

<https://portal.pedscommons.org>
commons.uchicago.edu



THE UNIVERSITY OF
CHICAGO



DATA FOR THE
COMMON GOOD

cdisc

Japan Academic Workshop
<http://sam.am/Japan2023>

<https://portal.pedscommons.org>
commons.uchicago.edu