

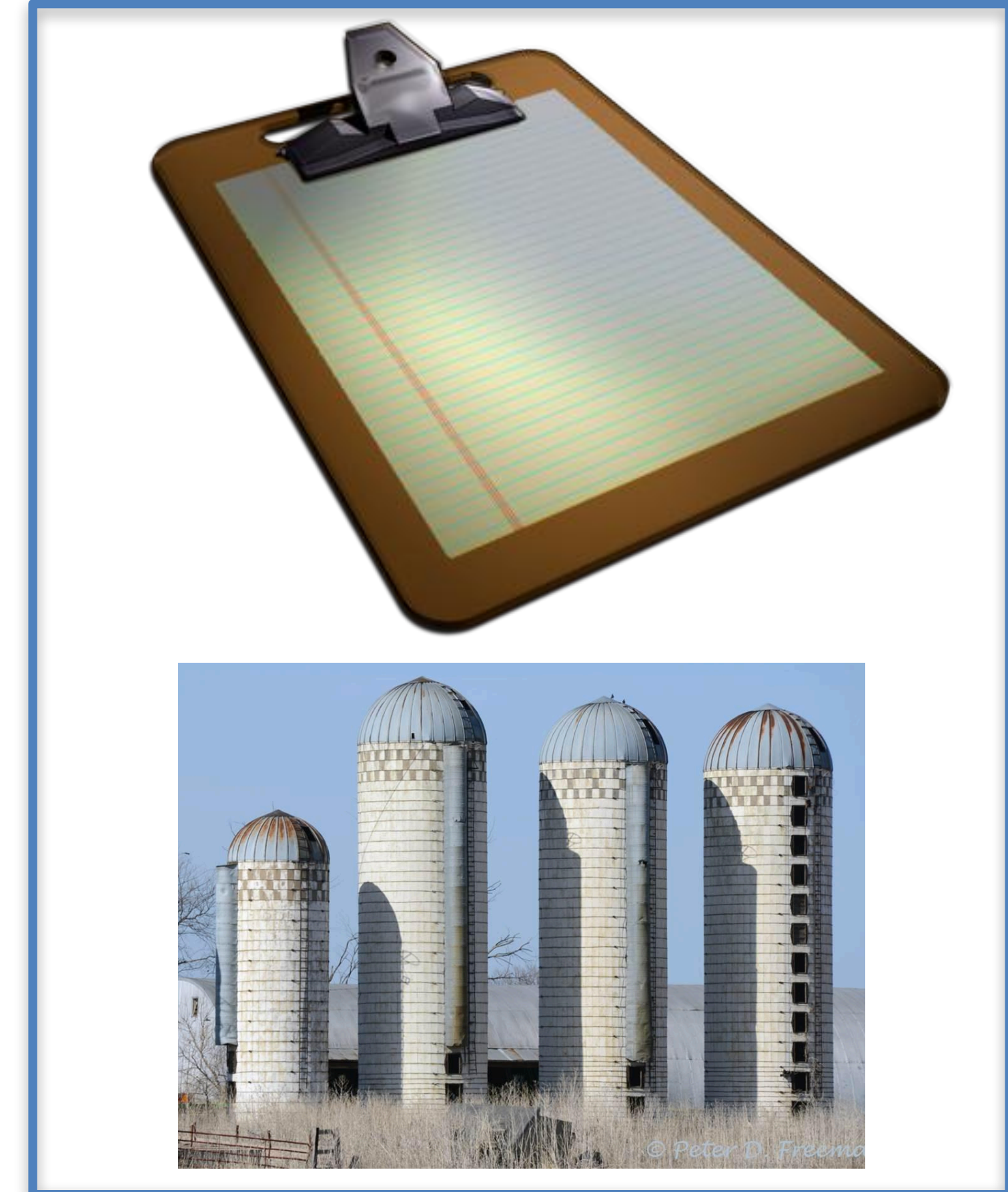
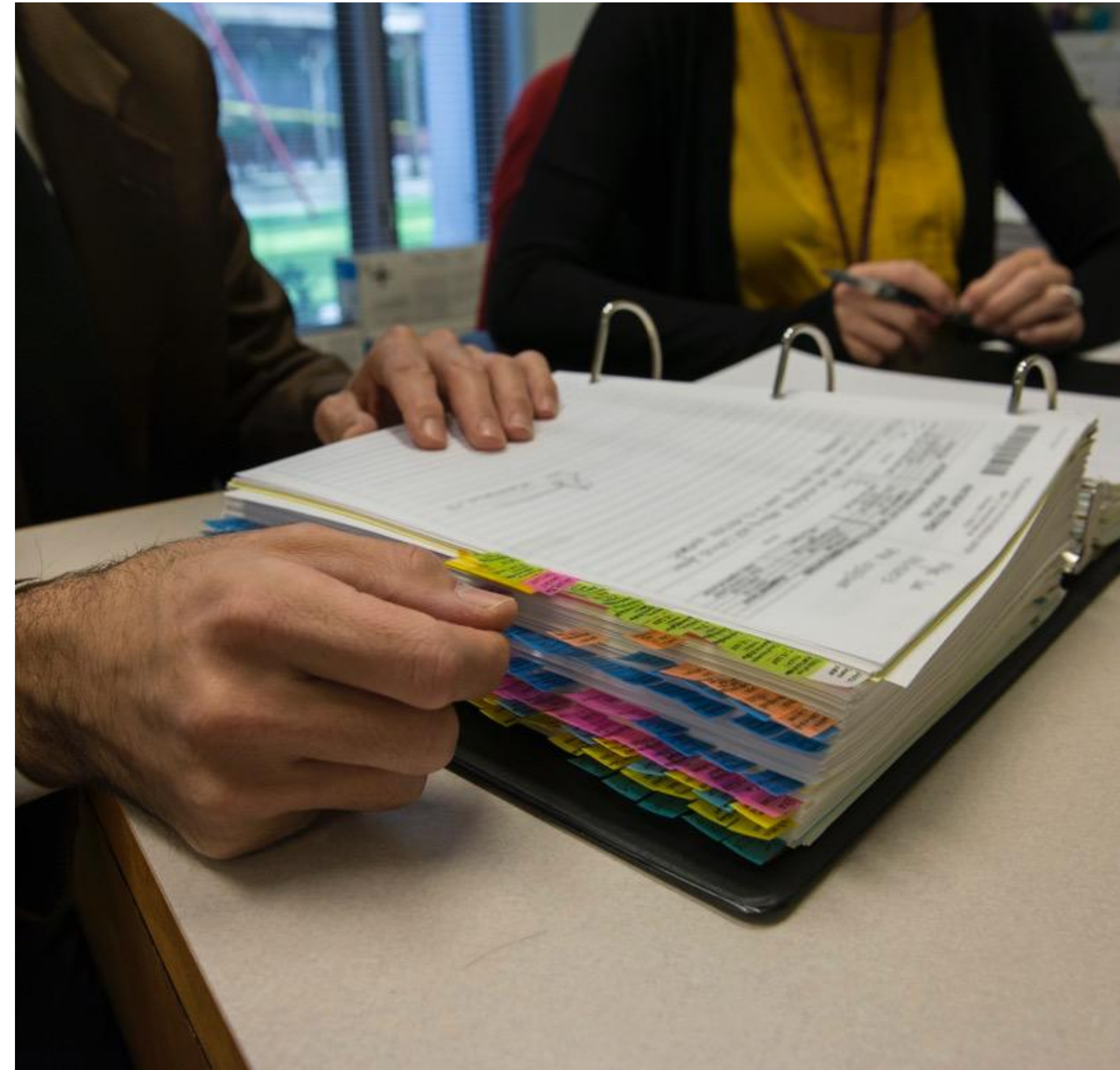
Pediatric Cancer Data Commons

Samuel Volchenboun, MD, PhD

May 15, 2018



Manual processes and lack of data standards plague clinical trials



© Peter D. Freeman



Manual processes and lack of data standards plague clinical trials

CTCAE CODE (per protocol)	CTCAE SHORT NAME (per protocol)	Current Grade	Maximum Grade This Course* (for occurrence)	Maximum Grade This Occurrence**	Attribution	Date Onset	Resolved Yes / No	Date Resolved	Comments
10020943	Hypo albuminemia	0	1	1	3	05/23	Yes	06/02	
10020943	Hypo albuminemia	0	1	1	3	06/05	Yes	06/09	
10021038	Hypo natremia	0	1	1	3	05/26	Yes	05/27	
10021038	Hypo natremia	0	1	1	3	05/29			
10021059	Hypo phosphatemia	0	1	1	3	05/10			
10021059	Hypo phosphatemia	0	1	1	3	05/23			
10016256	Fatigue	0	1	1	3	05/23			

Address for shipping the frozen GSH pellets is found section 6.3.6. Include NANT specimen train form with the shipment.

Phase 1 Study
TITLE: MODULATION OF INTENSIVE MELPHALAN (L-PAM) BY BUTHIONINE SULFOXIMINE (BSO) WITH AUTOLOGOUS STEM CELL SUPPORT FOR RESISTANT/RECURRENT HIGH-RISK NEUROBLASTOMA (IND # 69,112)

Phase 1 Study
COORDINATING CENTER
 Children's Hospital Los Angeles / NANT Operations Center

PARTICIPATING INSTITUTIONS AND CONTACT PHYSICIANS

Children's Hospital Los Angeles, 4650 Sunset Blvd., MS-54, Los Angeles, CA 90027.
 Judy Villablanca, M.D., Phone: 323-361-5654; Email: jvillablanca@chla.usc.edu

USCF School of Medicine, 505 Parnassus Avenue, M547, San Francisco, CA 94143.
 Katherine K. Mathay, M.D. Phone: 415-476-3831; Email: kmathay@ucsf.edu

Lucille Salter Packard Children's Hospital, 300 Pasteur Drive, Stanford, CA 94305.
 Clare Twiss, M.D. Phone: 650-723-6535; Email: clare.twiss@stanford.edu

Cincinnati Children's Hospital, 3333 Burnett Avenue, MLC 7015, Cincinnati, OH 45229.
 Brian Weiss, M.D., Phone: 513-636-9883; Email: brian.weiss@ccmc.com

Children's Hospital of Philadelphia, 3615 Civic Center Boulevard, ARC 907C, Philadelphia, PA 19104.
 Yael Moses, M.D., Phone: (215) 990-0965; Email: ymoses@chop.edu

C.S. Mott Children's Hospital, 1500 East Medical Center Drive, B1-208, Ann Arbor, MI 48109.
 Gregory Yank, M.D., Phone: 734-636-6795; Email: gyan@umich.edu

Texas Children's Hospital, Baylor College of Medicine, 6701 Fannin, 14th Floor, MC CC1410, Houston, TX 77030.
 Susan Blaney, MD, Phone: (832) 822-4586; Email: SBlaney@bcm.tmc.edu

Children's Hospital and Regional Medical Center, 4800 Sand Point Way NE, MS-401, Seattle, WA 98105.
 Julie Park, MD, Phone: 206-987-1947; Email: julia.park@seattlechildrens.org

Children's Hospital Boston, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02215.
 Suzanne Shusterman, MD, Phone: (617) 632-3725; Email: Suzanne_shusterman@dfci.harvard.edu

University of Chicago, Comer Children's Hospital, 5841 South Maryland Avenue, Chicago, IL 60637.
 Susan Coates, MD, Phone: (773) 702-5271; Email: scoates@bsd.uchicago.edu

Children's Healthcare of Atlanta, 2015 Uppergate Drive, Atlanta, GA 30322.
 MD, Phone: (404) 795-0853; Email: Hopen@childrens.com

Stoll for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8.
 In Banuchel, MD, Phone: (416) 813-7795; Email: sylvain.banuchel@sickkids.ca

Stem Hospital of New York-Presbyterian, Herbert Irving Division of Child & Adolescent Onc, 161 Ave. IP7, New York, NY 10032, Julia Glade-Bender, MD, Phone: (212) 305-3379; Email: jglade@nyp.org

St. Granger, MD, Phone: (862) 885-2580; Email: mgranger@stgranger.org

Protocol Chairperson
 Judith G. Villablanca, M.D.
 Children's Hospital Los Angeles
 4650 Sunset Blvd., MS-54, Los Angeles, CA 90027
 Phone: (323) 361-5654 or (323) 361-5687
 Email: jvillablanca@chla.usc.edu

1.1.7. BSO
 Patients must have a performance status of 0 or 1 (Appendix I) and a life expectancy of ≥ 2 months.

1.1.8. Organ Function
 1.1.8.1 Kidney Function
 Patients must have adequate renal function defined as a glomerular filtration rate (GFR) or 12 hour urine collection for creatinine clearance ≥ 100 mL/min/1.73m². **AKI** or serum creatinine $< 1.5 \times$ normal for age.

Age Range | Serum Creatinine
 ≤ 5 years | ≤ 0.8 mg/dL

N9902 Amd #7 9-4-2010 17

October		William				
Sun	Mon	Tue	Wed	Thu	Fri	Sat
	1 Bactrim Cytoxin 20mg	2 Bactrim Cytoxin 20mg	3 cycle 2 DTR 9:30am Avastin IV Cytoxin 20mg	4 DTR 9:30am Cytoxin IV Zometa IV	5 Cytoxin 20mg	6 Cytoxin 20mg
7 Cytoxin 20mg	8 Bactrim Cytoxin 20mg Labs and blood pressure at home	9 Bactrim Cytoxin 20mg	10 Cytoxin 20mg	11 Cytoxin 20mg	12 Cytoxin 20mg	13 Cytoxin 20mg
14 Cytoxin 20mg	15 Bactrim Cytoxin 20mg	16 Bactrim Cytoxin 20mg	17 DTR 9:30am Avastin IV Cytoxin 20mg	18 Cytoxin 20mg	19 Cytoxin 20mg	20 Cytoxin 20mg
21 Cytoxin 20mg	22 Bactrim Cytoxin 20mg SSK	23 Bactrim Cytoxin 20mg MIBG, x-ray knees Bone marrow biopsies SSK	24 Cytoxin 20mg MIBG CT scan SSK	25 Cytoxin 20mg	26 Cytoxin 20mg	27 Cytoxin 20mg
28 Cytoxin 20mg	29 Bactrim Cytoxin 20mg	30 Bactrim Cytoxin 20mg	31 Cycle 3 DTR 9:30am Avastin IV Cytoxin 20mg	1 DTR 9:30am Cytoxin IV Zometa IV		

Lack of rich phenotype data hinders progress

- Many samples for genomics lack sufficient clinical information
- Without deep phenotype data, analyses are limited
- Deep phenotype data should be collected at the time of sample acquisition, directly from electronic health record systems



Future* ideal state

* Way in the future

Electronic health record

~PCORI
~All of Us

Automatic
extraction

Sequencing

Genomics → Rich analyses

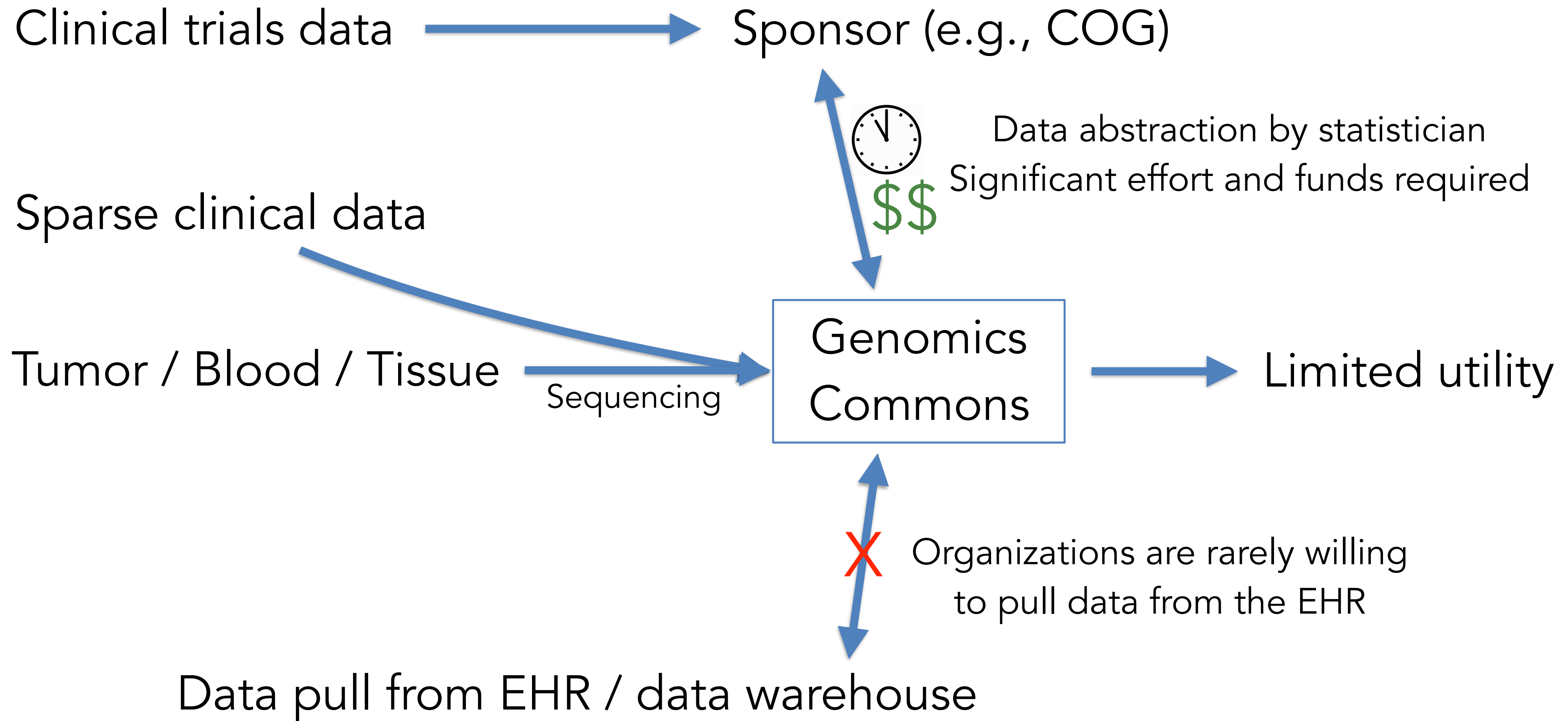
Tumor / Blood / Tissue

Requires:

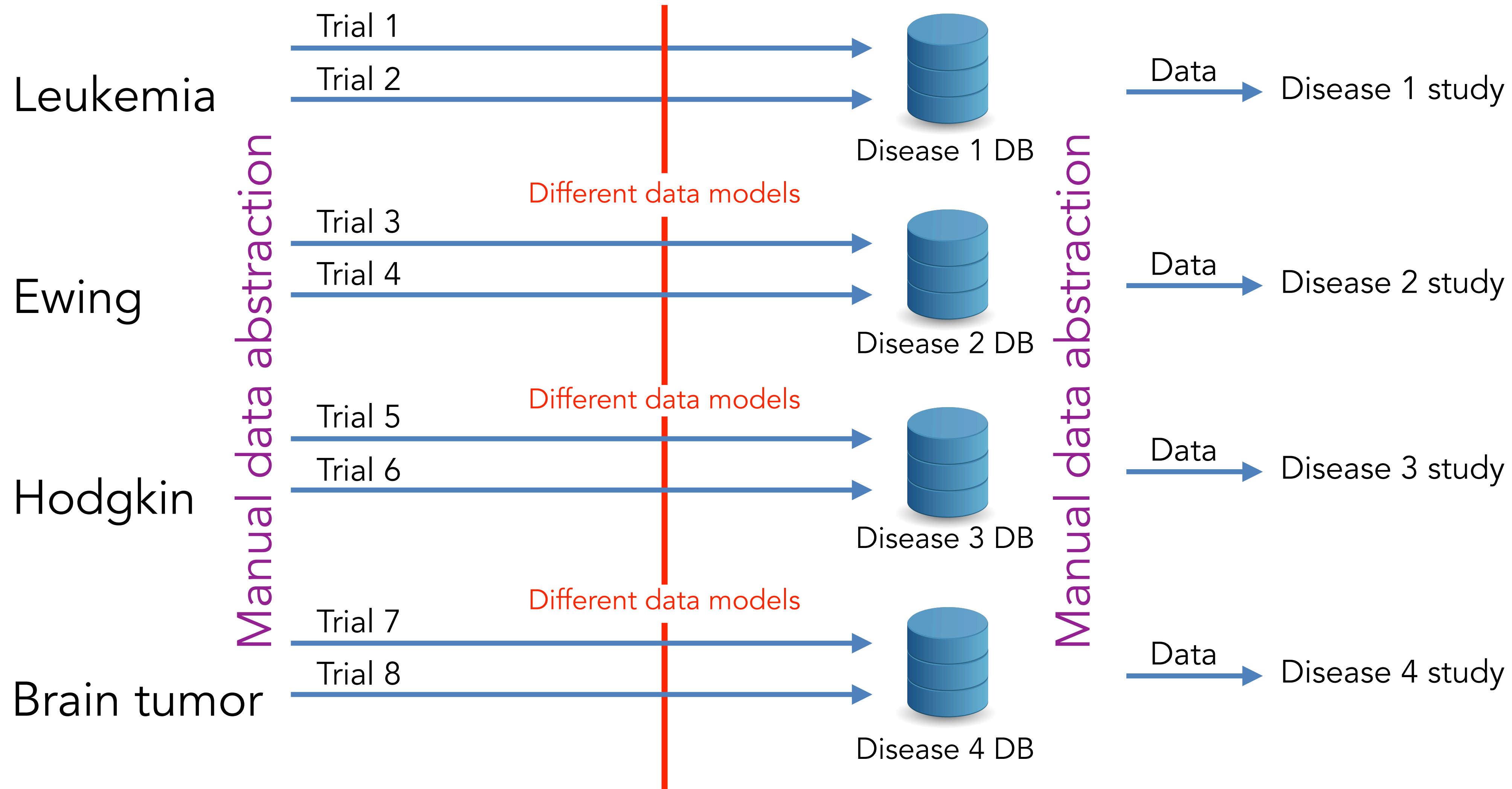
1. Common data model
2. Broad consent
3. State-of-the-art data warehouse
4. Foresight and planning



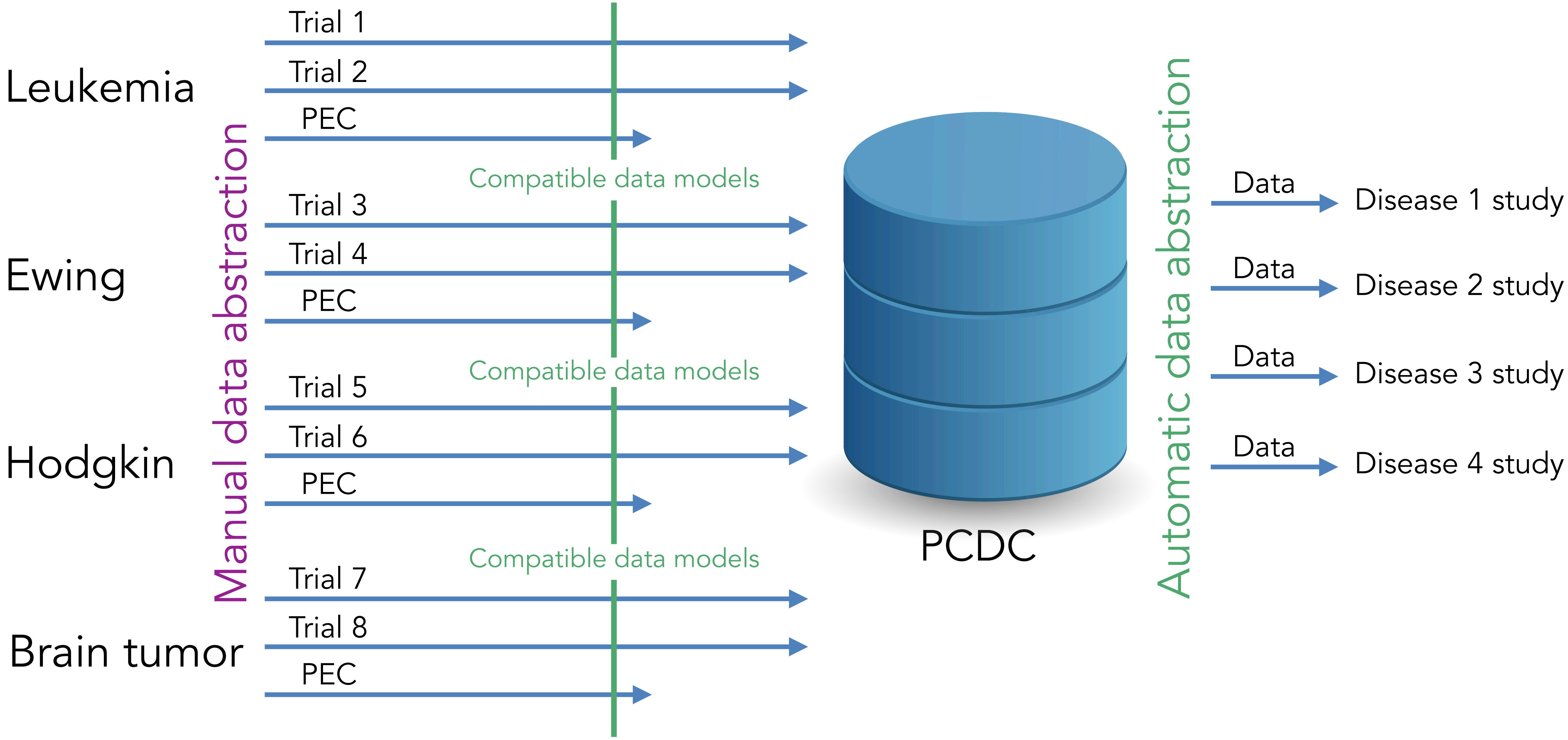
Current state for genomic commons



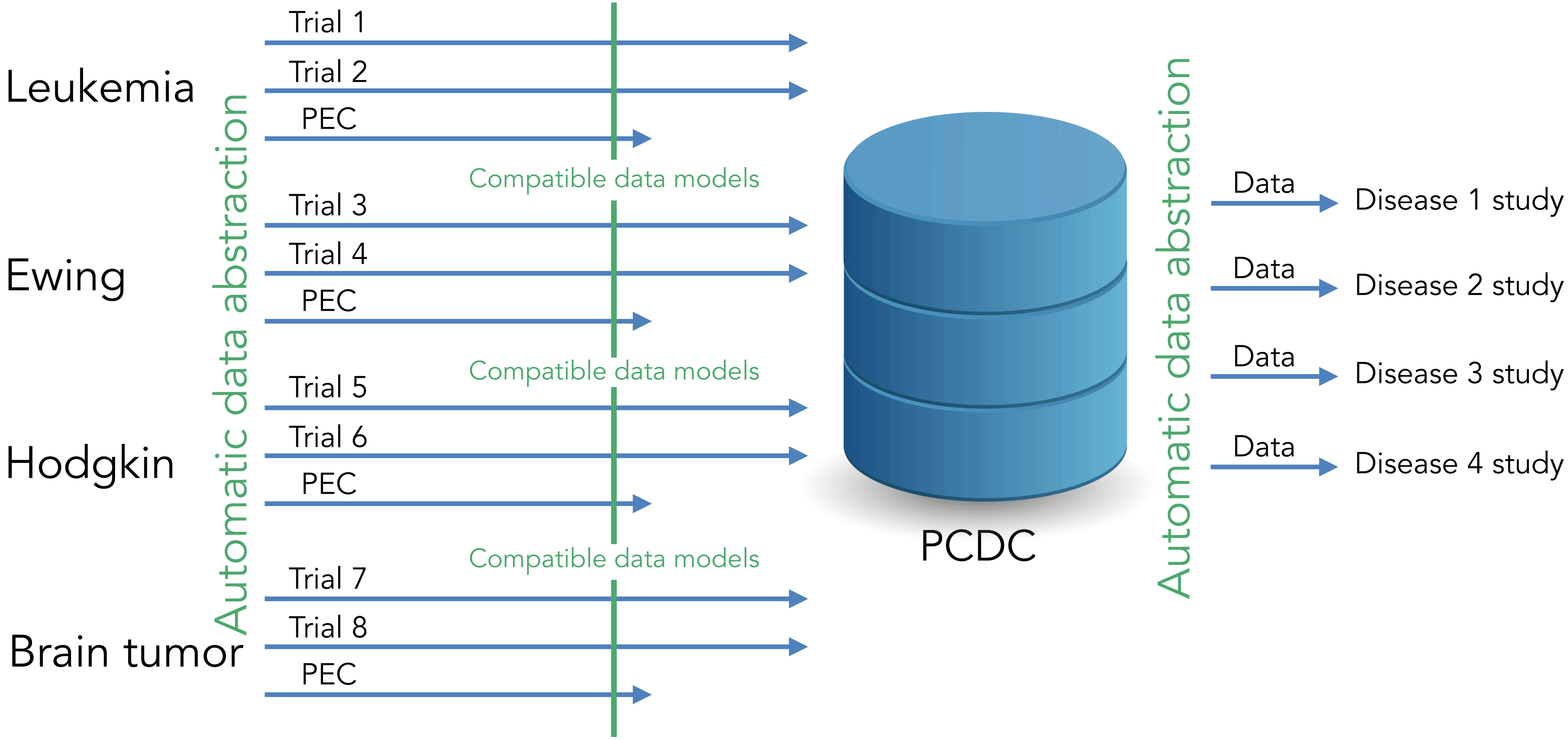
Current paradigm for COG



Goal



Goal



Pediatric cancer is rare - making it difficult to study

Adult cancers annual incidence

All	1,688,780
Oral	49,670
GI	310,440
Lung	222,500
Skin	95,360
Breast	255,180
Ovary	22,440
Prostate	161,360
Urinary	146,650
Lymphoma	80,500
Myeloma	30,280
Leukemia	62,130

Source: cancer.org - 2017

Pediatric cancers annual incidence

All	15,780
ALL	3080
CNS	2780
Hodgkin lymphoma	1180
NHL	1040
AML	730
Neuroblastoma	710
Bone	820
Thyroid	570
Wilms	510
Germ cell	540
Rhabdomyosarcoma	340
Retinoblastoma	280
Melanoma	310
Other	2890

Source: CDC - 2014



Most children are treated on a Children's Oncology Group (COG) study

1955

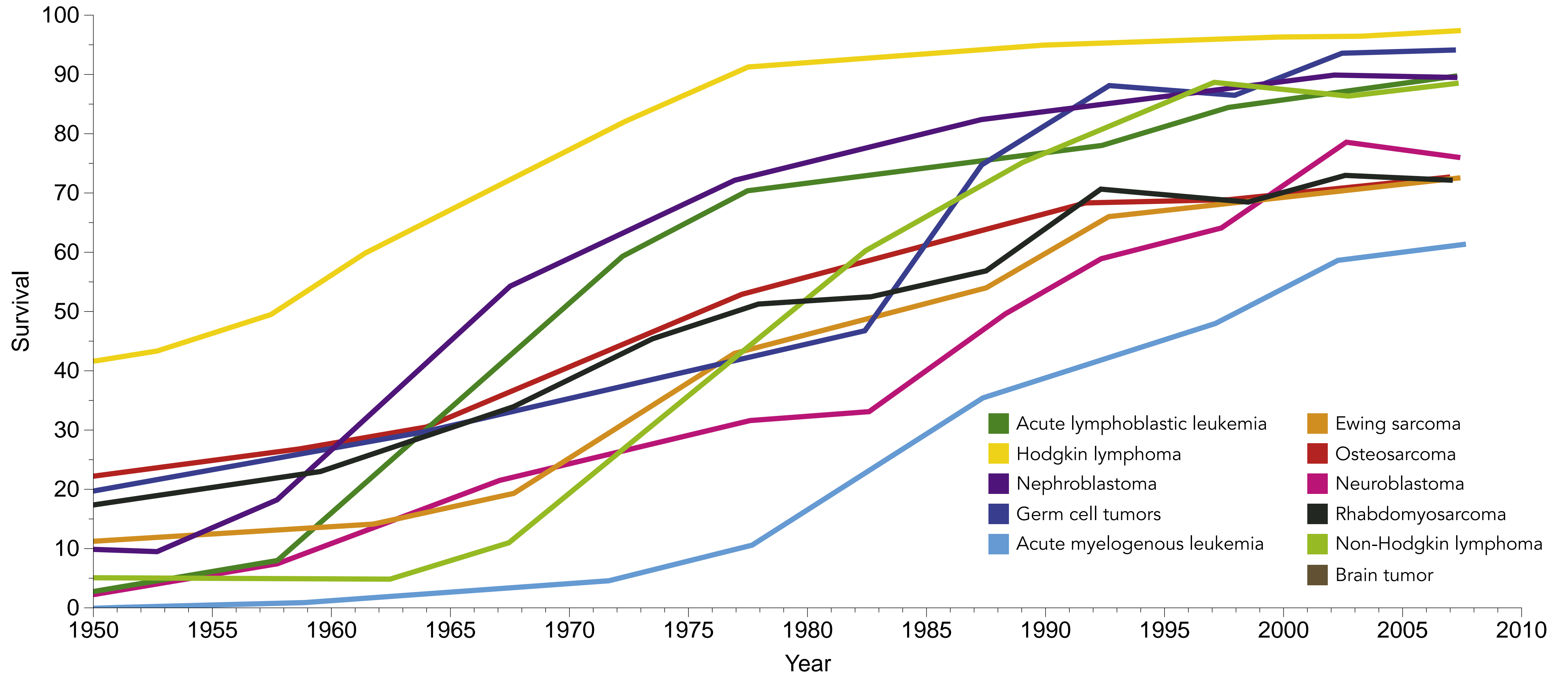
Cooperative group system for clinical research

Pediatric Oncology Group (POG)
Children's Cancer Group (CCG)
National Wilms' Tumor Study Group (NWTSG)
Intergroup Rhabdomyosarcoma Study Group (IRSG)

2000

Children's Oncology Group (COG)

Survival - Pediatric cancer - still a long way to go



Neuroblastoma data commons

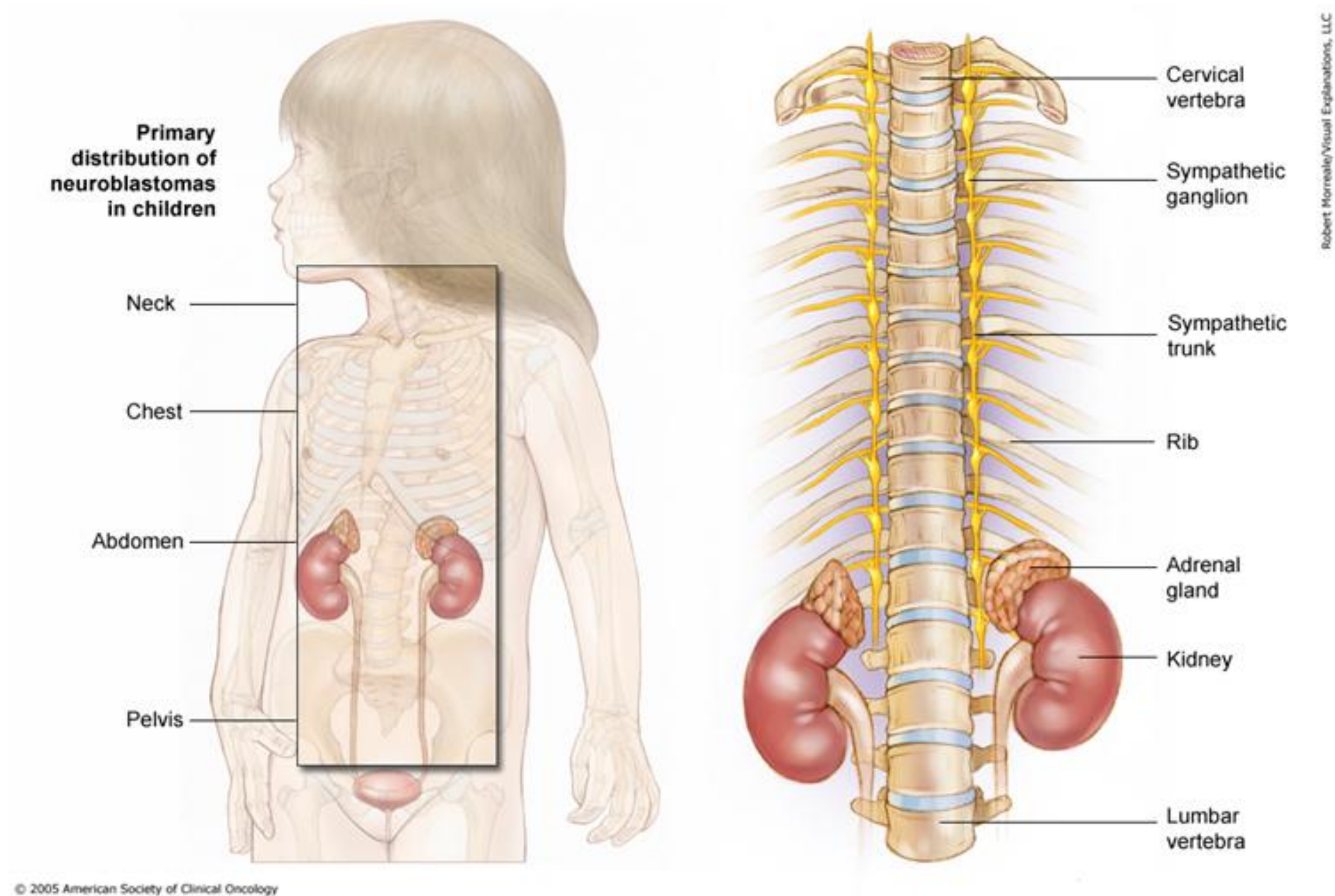


THE UNIVERSITY OF
CHICAGO MEDICINE &
BIOLOGICAL SCIENCES



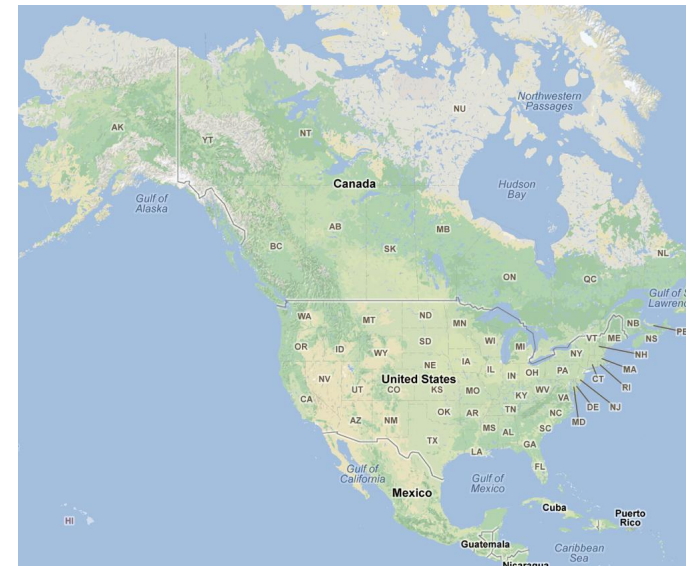
CENTER FOR
RESEARCH
INFORMATICS

Neuroblastoma



The most common solid tumor in children

International Neuroblastoma Research Group (2004)



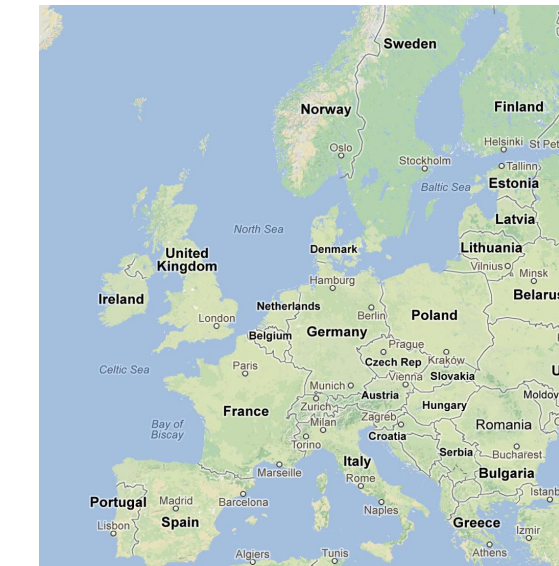
Children's Oncology Group (COG)



Germany



Japan



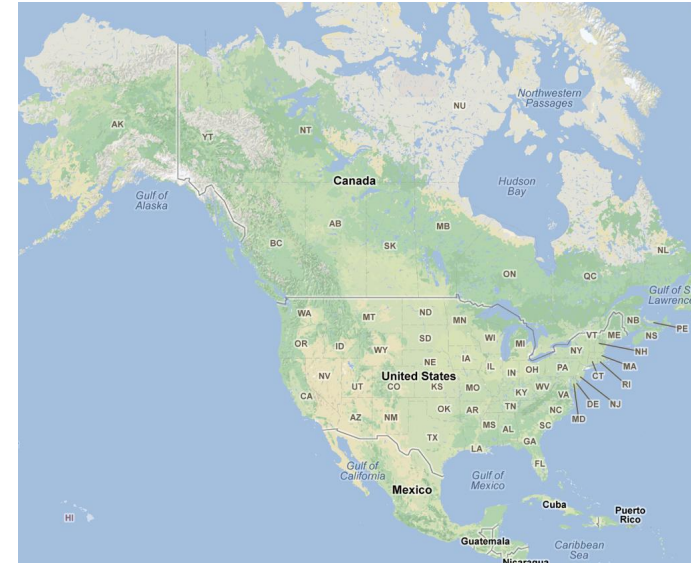
SIOPEN

COG	SIOPEN	Japan	Germany
Study 1	Study 4	Study 7	Study 10
Study 2	Study 5	Study 8	Study 11
Study 3	Study 6	Study 9	Study 12
COG	SIOPEN	Japan	Germany
Consensus	Consensus	Consensus	Consensus

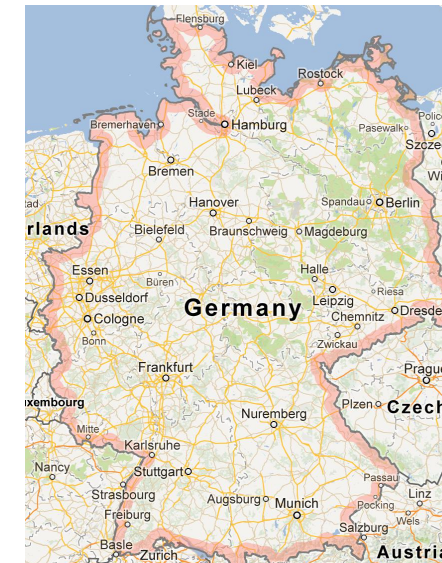


Consensus standard

International Neuroblastoma Research Group (2004)



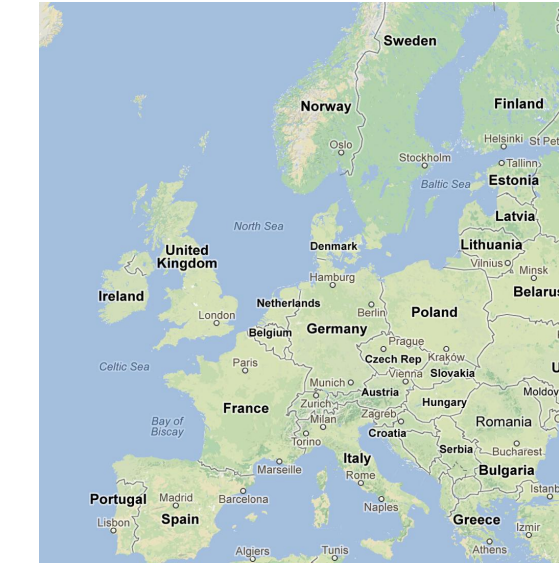
Children's Oncology Group (COG)



Germany



Japan



SIOPEN

Group	Number
COG	4235
Germany	1938
Japan	470
SIOPEN	936
Total	8800

The good news: 8800 patients

The bad news:
Data in Excel

Age	Year	Ferritin	LDH	EFtime	Stime	EFScns	Scns	Initial_Treat	INSS_Sig	Evans_Sig	Ip_tool	Lq_ubi	mycn	plody	1q_gaprr	adrrr	abdrpr	neckpr	thorpr	pelvpr	othmet	brmet	bonmet	dirmet	livmet	skmet	lunmet	crmet	oth	hist	grade	mki	diag	img_1					
691	1992	238	1024	2284	2284	0	0	2	2	2	9	9	9	1	9	1	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	101							
2272	1995	139	263	2345	2345	0	0	2	4	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	102							
839	1990			3389	3389	0	0	2	4	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	103							
787	1994	84	1849	925	1111	1	1	3	5	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	1	9	9	104								
792	1995	36	347	2273	2273	0	0	1	1	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	105							
9	1995			737	737	1	0	1	1	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	106							
393	2001			1	1	0	0	9	5	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	107							
9465	1994			528	878	1	1	2	5	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	108							
168	1998		420	1837	1837	0	0	1	1	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	109							
8	1982			8067	8067	0	0	2	9	5	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	110							
1225	1993	50	1744	176	176	1	1	9	5	4	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	111							
2122	1995			592.8	3349	3349	0	0	2	6	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	1	9	9	9	112						
528	1995			809	1834	1834	0	0	1	1	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	113							
1630	1990			676	975	1141	1	1	2	9	4	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	9	9	9	114						
5096	1989			375	531	1	1	2	9	4	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	115							
38	1999			1250	1250	0	0	2	4	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	116							
653	1994			2503	2503	0	0	5	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	1	0	3	1	117							
809	2002			812	812	0	0	1	3	2	9	9	9	9	1	9	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	118						
772	1990	17.9	266	4719	4719	0	0	2	4	9	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	119							
440	1998	13	260	1987	1987	0	0	2	2	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	120							
482	2000	71	672	32	32	1	1	3	3	9	1	0	1	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	121							
1045	1997	200	200	200	200	1	1	5	5	9	0	9	1	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	1	9	3	1	122						
2204	1989			3196	3196	0	0	2	5	9	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	1	9	9	9	123						
323	1991			2878	2878	0	0	2	3	9	9	9	9	9	1	9	9	9	9	9	9	9	9	9	9	9	9	0	0	0	0	1	9	9	9	124			
401	1999			314	314	0	0	9	4	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	1	9	9	9	125						
19	2001	89	242	985	985	0	0	1	1	9	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	126						
14	1987			4495	4495	0	0	1	9	3	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	127						
1988	1998			2193	2193	0	0	1	1	1	0	9	0	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	128						
879	1988			314	345	1	1	3	9	4	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	129						
211	1986			3620	3620	0	0	9	9	2	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	130							
436	1998			30	30	1	1	3	5	4	1	9	1	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	131					
754	2001			853	853	0	0	9	4	9	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	132					
2118	1998			1832	1832	0	0	1	1	9	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	133						
454	1999	39	338	1575	1575	0	0	1	1	9	0	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	134					
9	1994	211	1464	1207	1207	0	0	2	5	4	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	135				
191	1994			2380	2380	0	0	2	5	9	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	136			
270	1984			4660	4660	0	0	2	9	2	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	137						
604	1987			360	360	1	1	9	9	4	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	9	9	9	138			
22	1983			1368	1368	1	1	9	9	4	9	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	139			
1051	1992	124	180	684	708	1	1	3	5	9	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	1	140					
450	1980			497	618	1	1	4	9	4	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	9	141					
712	1995			424	424	1	1	5	9	9	1	0	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	142			
1712	1997	231	2410	39	39	1	1	3	5	9	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	9	9	9	9	143				
302	1982			6077	6077	0	0	9	9	3	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	9	1	144				
1381	2001	15		1125	1125	0	0	1	5	2	9	9	9	9	1	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	145		
126	2001			849	849	0	0	9	4	9	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	146	
444	1997	210	5994	2584	2584	0	0	3	5	4	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	147
4599	1999			1407	1407	0	0	2	4	9	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	9	9	9	148		
1176	1993	155	488	505	582	1	1	5	9	1	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	9	9	9	9	149		
486	1996	9	1598	1894	1894	0	0	9	1	1	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	150
1734	1990			2052	2052	0	0	9	3	9	9	9</																											

Neuroblastoma Commons Cohort Discovery

Example: Favorable tumor biology, tissue available

Cohort Search

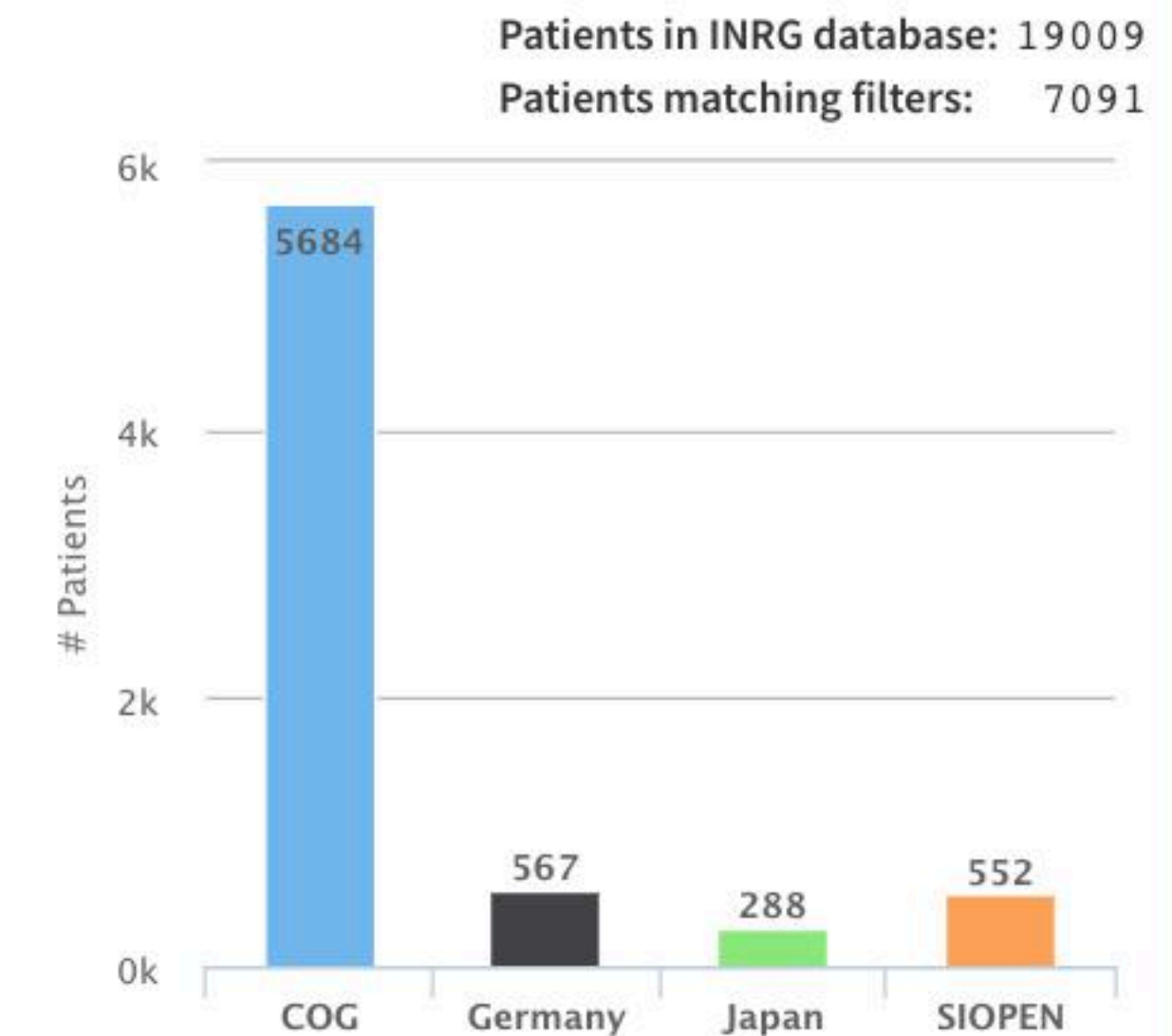
Search: New Save Search

Add a filter: - Please select - Clear Filters

Revised INPC Prognostic Group/ Shimada Diagnostic Category ×

- Favorable
- Unfavorable
- Unknown/Slides not received/Inadequate

Results



Neuroblastoma Commons Cohort Discovery

Cohort Search

Search: New Save Search

Add a filter: - Please select - Clear Filters

- Please select -
- Primary Tumor-Adrenail
- Primary Tumor-Neck
- Primary Tumor-Other
- Primary Tumor-Pelvis
- Primary Tumor-Thorax
- Race
- Revised INPC Prognostic Group/ Shimada Diagnostic Category
- Site of Relapse
- Time from Dx to Death or Last Contact
- Time from Dx to Event or Last Contact
- Year of Diagnosis
- GEO Data - **Note! External Data**
- GWAS Data - **Note! External Data**
- Nationwide Tissue Bank - **Note! External Data**
- Nucleic Acid Data - **Note! External Data**
- TARGET Data - **Note! External Data**

Revised INP
 Favorable
 Unfavorable
 Unknown

Results

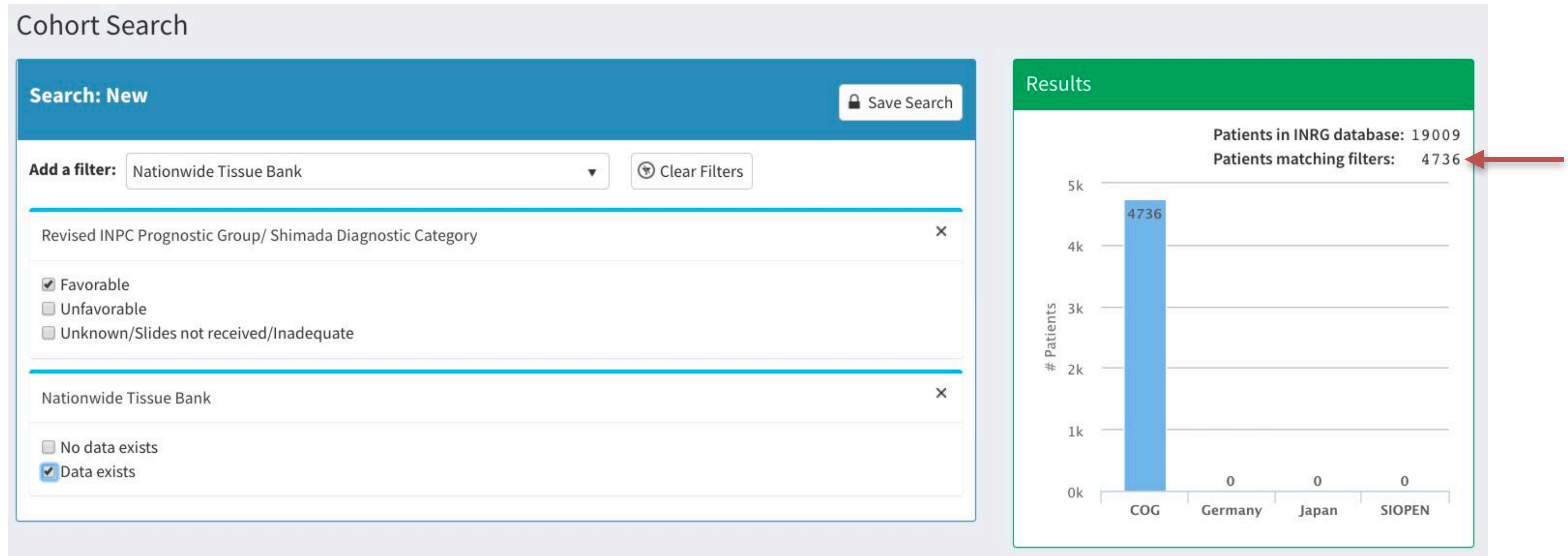
Patients in INRG database: 19009
Patients matching filters: 7091

Group	# Patients
COG	5684
Germany	567
Japan	288
SIOPEN	552

Immediately see cohort counts

Links to external data sets

Neuroblastoma Commons Cohort Discovery



This used to take weeks.
Now it can be done in seconds.

Example: Favorable biology, tissue available



COG Unique Specimen Identifier (USI)



COG institution

COG Identifier (PHI)



Honest broker

COG USI number

<u>COG ID</u>	<u>USI</u>
34574765	RDASXK
65780454	UFUMZQ
40508620	LKNQHD
62646200	JVMECO

Specimens Data

NATIONAL CANCER INSTITUTE
GENOMIC DATA COMMONS



Commons

Neuroblastoma data commons growth

Year	COG	SIOPEN	GPOH	Japan	Total
2004	4235	2157	1938	470	8800
2012	6127	2504	1938	470	11039
2013	11642	2504	1938	470	16554
2015	13060	2504	1938	470	17972
2016	13937	2664	1938	470	19009
2018	14425	3397	2154	470	20446

Data upload can be automated using a standardized data dictionary with error and consistency checking.

Paradigm for building a pediatric cancer commons

1. Engage cooperative group(s)
2. Define scope
3. Identify funding source
4. Identify infrastructure
5. Engage project team
6. Identify data sources
7. Establish governance, create policies and procedures
8. Create contributor / use agreements
9. Create standards working group to create data dictionary, map elements
10. Create database
11. Build front-end query engine
12. Create and execute communication and education plans
13. Create sustainability model

Volchenbom SL, Cox SM, Heath A, Resnick A, Cohn SL, Grossman R
"Data Commons to Support Pediatric Cancer Research"



Remapping NBL to a standard ontology

	Tumor type	Code
Current	Neuroblastoma (Schwannian stroma-poor)	1
	Ganglioneuroblastoma, intermixed (Schwannian stroma-rich)	2
	Ganglioneuroma (Schwannian stroma-dominant), maturing subtype OR Ganglioneuroblastoma, well differentiated (Schwannian stroma rich)	3
	Ganglioneuroblastoma, nodular (composite)	4
	Unknown	9
	Proposed	Neuroblastoma
Ganglioneuroblastoma - intermixed		C1517444
Ganglioneuroma		C0017075
Ganglioneuroblastoma - nodular		C1517445
Unknown		C0087135



Ganglioneuroblastoma

NIH NATIONAL CANCER INSTITUTE www.cancer.gov

EVS Enterprise Vocabulary Services

NCImetathesaurus
NCIm Version: 201706 Version 2.8 (using LexEVS 6.5)

unknown
 Contains Exact Match Begins With
 Name Code Property Relationship
Source: ALL

[Home](#) | [NCIt Hierarchy](#) | [Sources](#) | [Help](#) | [Visited Concepts](#)

Quick Links

Ganglioneuroblastoma, Intermixed (CUI C1517444) [Suggest changes to this concept](#)
[Add to Cart](#)

[Terms & Properties](#) | [Synonym Details](#) | [Relationships](#) | [By Source](#) | [View All](#)

Terms & Properties

Concept Unique Identifier (CUI): C1517444

NCI Thesaurus Code: C42057 ([see NCI Thesaurus info](#))

Semantic Type: Neoplastic Process

NCIt Definition: A ganglioneuroblastoma characterized by the presence of neuroblastic cells in a Schwannian stroma, without the presence of hemorrhagic neuroblastic nodules.

Synonyms & Abbreviations: ([see Synonym Details](#))

- Ganglioneuroblastoma, Intermixed
- Ganglioneuroblastoma, Intermixed (Schwannian Stroma-Rich)
- gemischt Ganglioneuroblastom
- intermixed ganglioneuroblastoma
- Schwannian stroma-rich peripheral neuroblastic neoplasm

External Source Codes:

NCI Thesaurus Code	C42057 (see NCI Thesaurus info)
--------------------	---

Other Properties:

Name	Value	Source
Neoplastic_Status	Malignant	NCI

Additional Concept Data: (none)

URL to Bookmark: [https://ncimeta.nci.nih.gov/ncimbrowser/ConceptReport.jsp?dictionary=NCI Metathesaurus&code=C1517444](https://ncimeta.nci.nih.gov/ncimbrowser/ConceptReport.jsp?dictionary=NCI%20Metathesaurus&code=C1517444)



Remapping NBL primary site

Current

PRI_ADRE	INTEGER	Primary site of tumor is adrenal	0=No 1=Yes 9=Unknown
PRI_ABDRET	INTEGER	Primary site of tumor is Abdominal/retroperitoneal (non-adrenal)	0=No 1=Yes 9=Unknown
PRI_NECK	INTEGER	Primary site of tumor is neck	0=No 1=Yes 9=Unknown
PRI_THOR	INTEGER	Primary site of tumor is thorax	0=No 1=Yes 9=Unknown
PRI_PELV	INTEGER	Primary site of tumor is pelvic	0=No 1=Yes 9=Unknown
PRI_OTH	INTEGER	Primary site of tumor is "other"	0=No 1=Yes 9=Unknown

Proposed

Adrenal gland	C0001625
Retroperitoneal	CL318004
Neck	C0027530
Thorax	C0817096
Pelvis	C0030797
Not otherwise specified	C1518425
Unknown	C0439673



Data commons in planning and development

Soft-tissue sarcoma

AYA lymphoma

Sickle cell

STS data commons

- Discussions began 1/2017
- First meeting in May, 2017 (Copenhagen)
- Second meeting in October 2017 (Chicago)
- Third meeting in March 2018 (Amsterdam)
- Fourth meeting planned September 2018 (Tübingen)
- Executive and informatics calls every 1-2 months



INSTRuCT meetings

October 2017 - Chicago



March 2018 - Amsterdam



Rhabdomyosarcoma data standardization

	Current	
	Code	Description
SEX	1	Male
	2	Female
HISTOLOGY	1	Alveolar rhabdomyosarcoma
	2	Embryonal rhabdomyosarcoma
	3	Botryoid rhabdomyosarcoma
	4	Not otherwise specified
	5	Undifferentiated sarcoma
	6	Sarcoma, not classifiable
	7	Spindle cell sarcoma
	8	Ectomesenchymoma
	9	Other
	10	Mixed rhabdomyosarcoma
99	Unknown	

Male	C0086582
Female	C0015780
Unknown	C0439673

Alveolar rhabdomyosarcoma (ARMS)	C0206655
Botryoid rhabdomyosarcoma (BRMS)	C1306573
Embryonal rhabdomyosarcoma (ERMS)	C0206656
Pleomorphic rhabdomyosarcoma (PRMS)	C0334480
Rhabdomyosarcoma (RMS), not classifiable	C0035412
Rhabdomyosarcoma (RMS), inadequate tissue for classification	
Rhabdomyosarcoma (RMS), w Mixed Embryonal and Alveolar Features	C1709053
Spindle cell	C0205945



Consensus example: Maximum tumor diameter

Old

Maximum diameter

or

x, y, z

or

$>5\text{cm}, <5\text{cm}$

INSTRuCT consensus

- Discrete measurement (in cm)
 - x (or max diameter if single)
 - y
 - z
- Category (if no discrete meas.)
 - ≤ 5 cm
 - >5 cm
 - Unknown



Consensus harmonization of primary site

CWS

COG

EpSSG

Consensus

Major Primary Site	CWS	COG	EpSSG/MMT Name	
ORBIT	Eyelid	1=Eye	Eyelid	
	Orbit	2=Orbit	Orbit	
HEAD & NECK (non PM)	Scalp	10=Scalp	Soft tissue of scalp	
			External auricular canal	
			Ear soft tissue, external ear	
	Parotid	9=Paratoid	7=Oral cavity	Temporal muscle
				Parotid, soft tissue
				Gum
				Base of tongue
				Lip
	Oral Cavity			Lower lip
				Upper lip
				Tongue
	Larynx	5=Larynx	8=Orophaynx	Larynx
				Oropharynx
	Oropharynx			Oropharynx
				Lingual tonsil
				Mandible soft tissue
				Bone of face (Maxillar)
				Masseter
				Oral cavity
Cheek				
Hypopharynx				
Thyroid				
Thyroid & Parathyroid				11=Thyroid & Parathyroid
Neck	6=Neck		Neck	
			Neck supra-clavicular soft tissues	
			Neck, nodes Nos	
			Chin	
			Soft tissue face (non specified region)	
			Face specified region	
		12=Other Head & Neck	Nasolabial fold (skin)	
			Nostril	

Orbit

Eyelid	C0015426
Orbit	C0700042
Other orbit	C0700042

Head and neck

Cheek	C0007966
Hypopharynx	C0020629
Larynx	C0023078
Neck	C0027530
Oral cavity	C1711367
Oropharynx	C0521367
Parotid	C3272625
Scalp	C0036270
Thyroid and parathyroid	C0574117
Other face	C0015450
Other head and neck	C0460004

INSTRuCT consensus data dictionary

- ID
- Sex
- Race
- Ethnicity
- Year of diagnosis
- Age at diagnosis
- Histology
- Stage
- Surgery extent
- Group
- Tumor site
- Metastatic site(s)
- Tumor size
- Invasiveness
- TNM
- Anaplasia
- Fusion status
- Survival
- Age at censor
- Age at death
- Cause of death
- Events
- Secondary metastatic site



Developing a paradigm for treatment data

1.03 ChemoType

- (a) 1
C-PEb: cytoxan added to PEb.
- (b) 2
CTX: cyclophosphamide.
- (c) 3
EP: etoposide and cisplatin.
- (d) 4
HD-PEb: PEb with double dose of cisplatin.
- (e) 5
PEb: bleomycin, etoposide, cisplatin.
- (f) 6
PEj: pharyngo-esophageal junction?
- (g) 7
PVB: vinblastine, bleomycin, cisplatin.

(hint: this is not it)



Chemotherapy reference data is not sufficient

ANBL0531 - *Response and Biology-Based Therapy for Intermediate-risk Neuroblastoma.*

Administration of chemotherapy through a reporting period of 3 weeks. [Duration of 1 cycle]

Cycle # 1

DRUG	ROUTE	DOSAGE	DAYS	IMPORTANT NOTES	OBSERVATIONS
CARBOplatin (CARBO)	IV over 1 hour	Pts ≤ 12kg: 18.6 mg/kg/dose Pts > 12 kg: 560 mg/m ² /dose	1		a. Physical Exam (Ht., Wt., BSA, VS), b. CBC, differential, platelets ¹ , Serum creatinine ²
Etoposide (ETOP)	IV over 1 hour	Pts ≤ 12kg: 4 mg/kg/dose Pts > 12 kg: 120 mg/m ² /dose	1-3	Give following CARBO on Day 1 only	

Carboplatin/Etoposide (CUI C0280617)

[Suggest changes to this concept](#)
[Add to Cart](#)

Terms & Properties | **Synonym Details** | Relationships | By Source | View All

Terms & Properties

Concept Unique Identifier (CUI): C0280617

NCI Thesaurus Code: C10289 ([see NCI Thesaurus info](#))

Semantic Type: Therapeutic or Preventive Procedure

Synonyms & Abbreviations: ([see Synonym Details](#))

Carboplatin/Etoposide
CBDCA/VP-16

External Source Codes:

NCI Thesaurus Code: C10289 ([see NCI Thesaurus info](#))

Other Properties: [?](#)

Name	Value	Source
NCI_THESAURUS_CODE	C10289	PDQ
ORIG_STY	Drug/agent combination	PDQ
PID	4041	PDQ

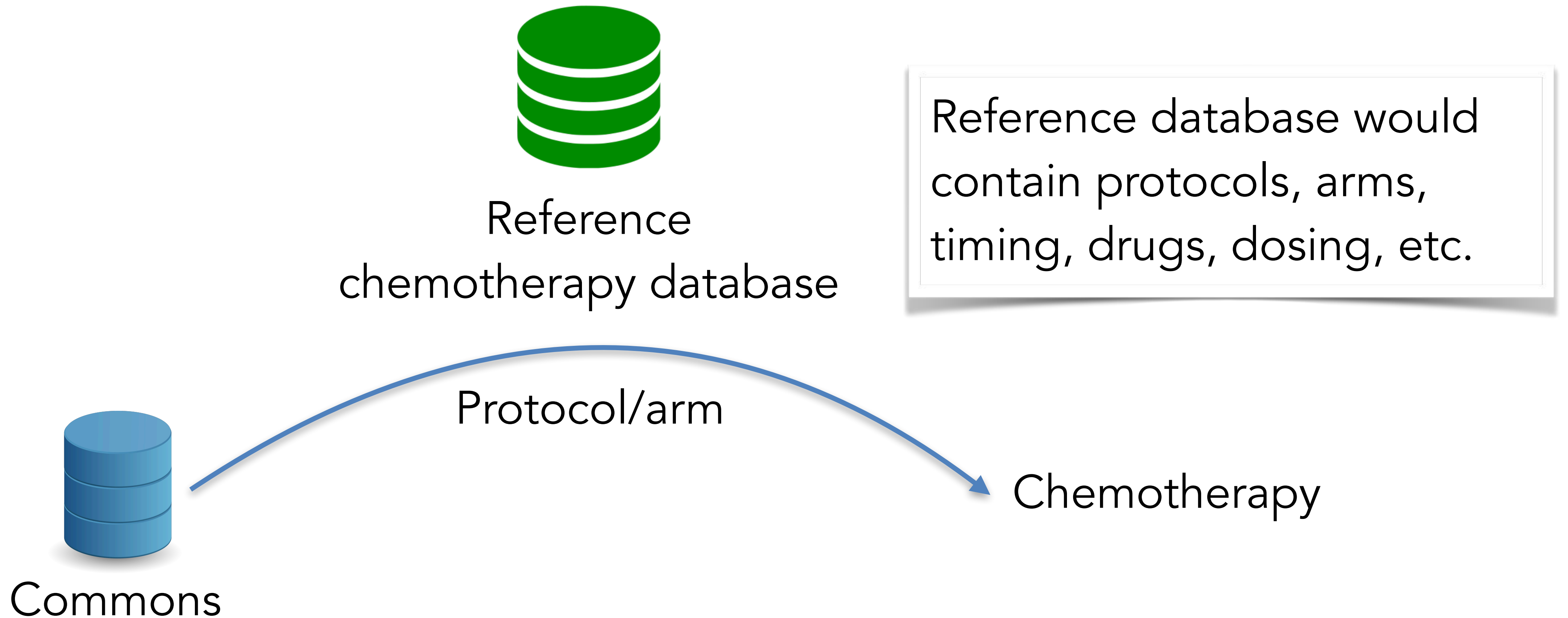
Does not include dose information or a reference

OTHER STUDIES AS
FOR GOOD PATIENT

Protocol Section 7.0 for complete list
ations, including required
at study entry, and Section 15.0
quisition or shipment guidelines



Paradigm for chemotherapy information



COG data standards workshop

April 10, 2018 - St. Louis

- 45 attendees from COG, NCI, CDISC, St. Baldrick's
- Four breakout sessions
 - Data processes
 - Common data platform
 - Data governance
 - Biospecimen curation



April 10 COG data standards workshop



Plan to develop informatics task force for COG

Make recommendations for informatics strategic plan for COG



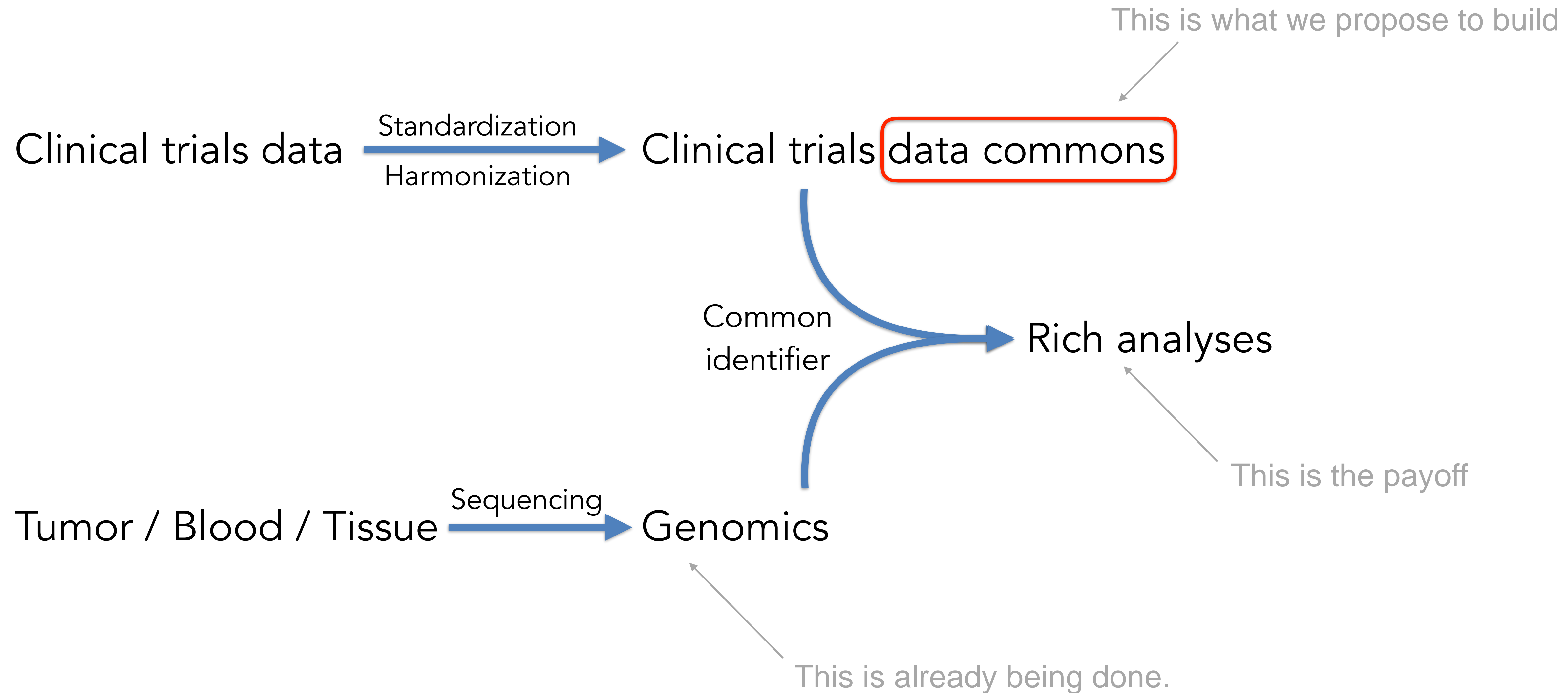
THE UNIVERSITY OF
CHICAGO MEDICINE &
BIOLOGICAL SCIENCES



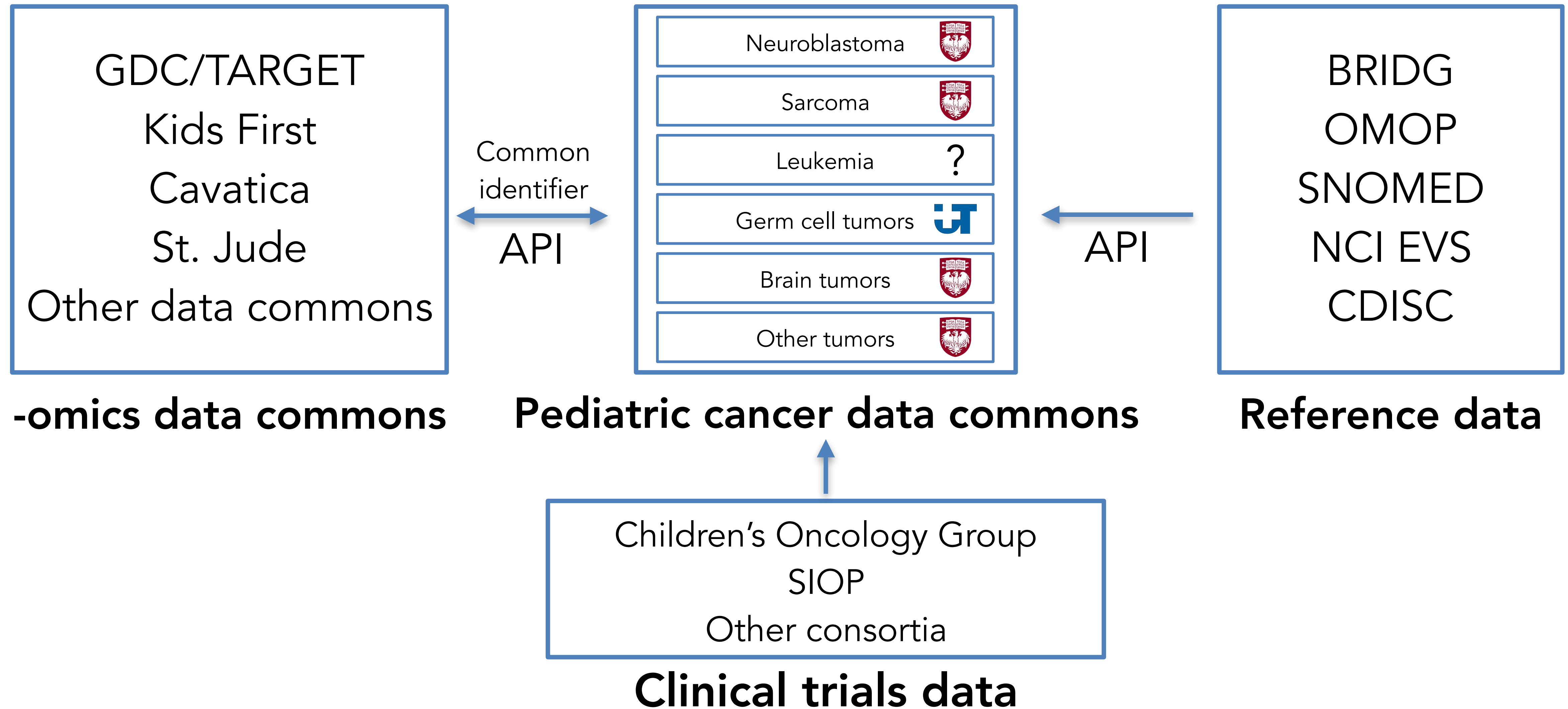
CENTER FOR
RESEARCH
INFORMATICS



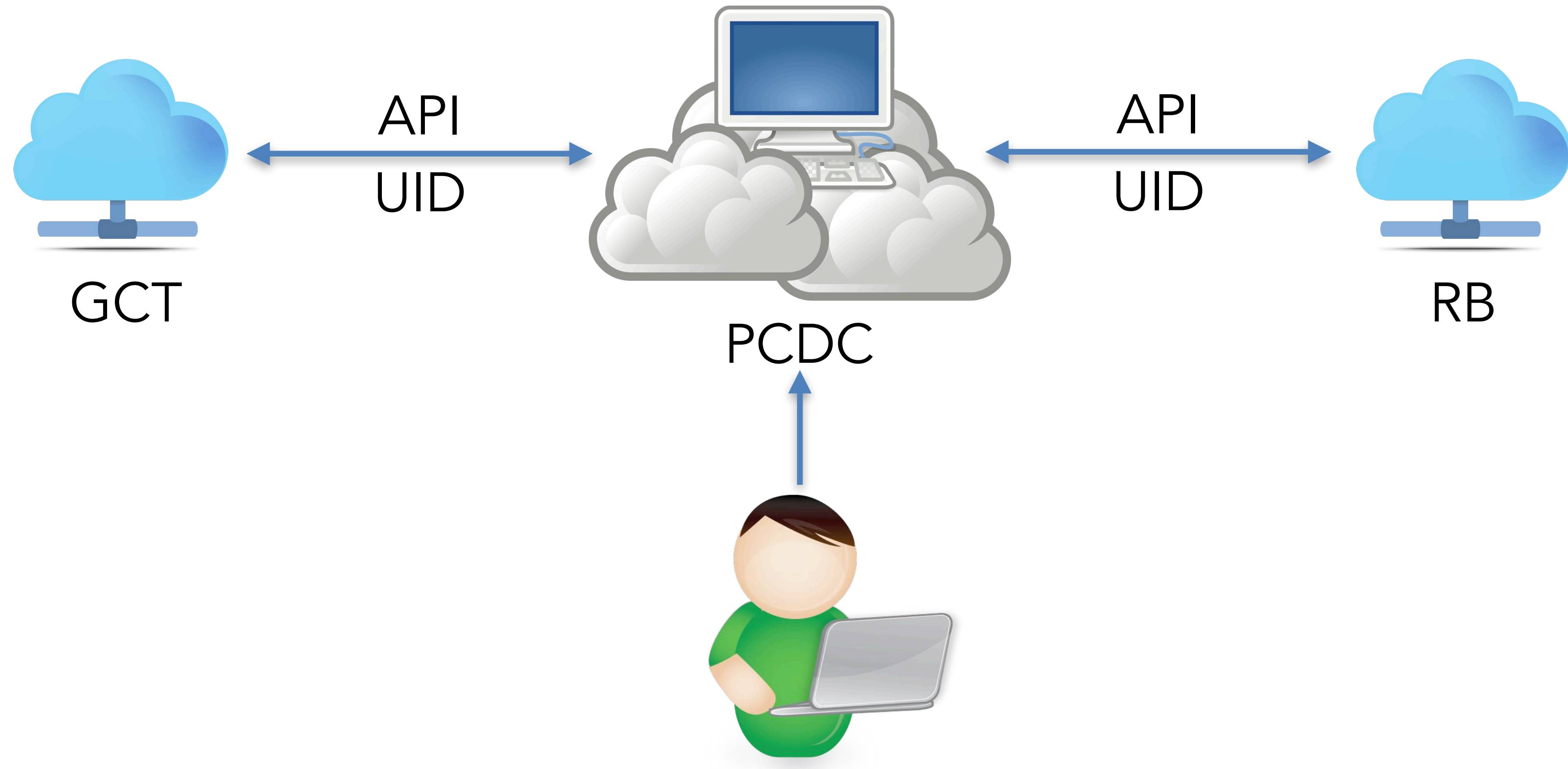
Goal: Build a commons of clinical trials data to enrich the genomic information



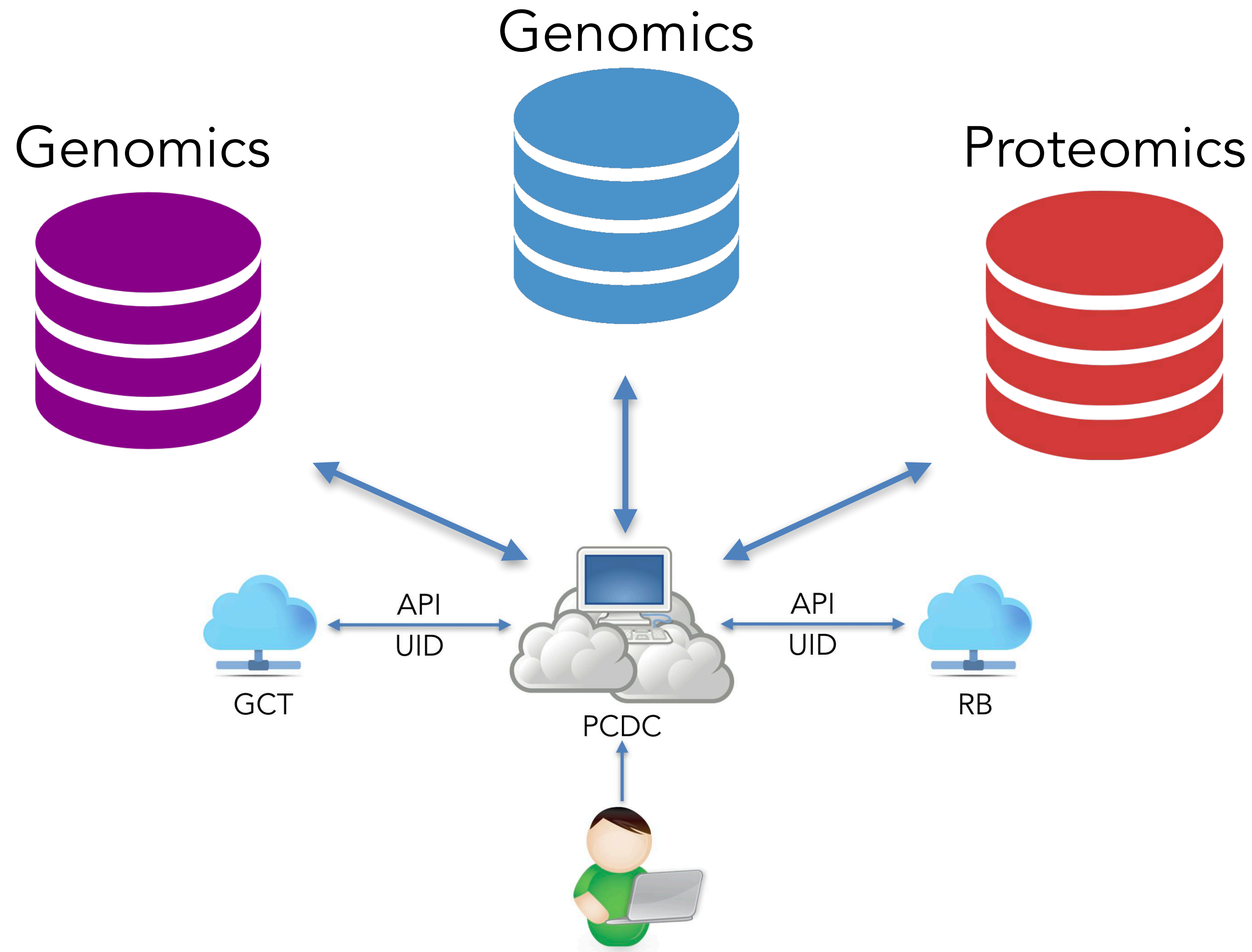
Paradigm for pediatric cancer clinical trials data commons



Interacting commons for clinical data



Interacting commons for clinical data



Summary / Call to action

- Harmonized data leads to shared data
- Data and samples must have universal identifiers
- We must envision data collection and sharing at all stages of care
- The goal is all data from all patients at all times



