

# Pediatric Cancer Data Commons

Samuel Volchenboum, MD, PhD

May 15, 2018

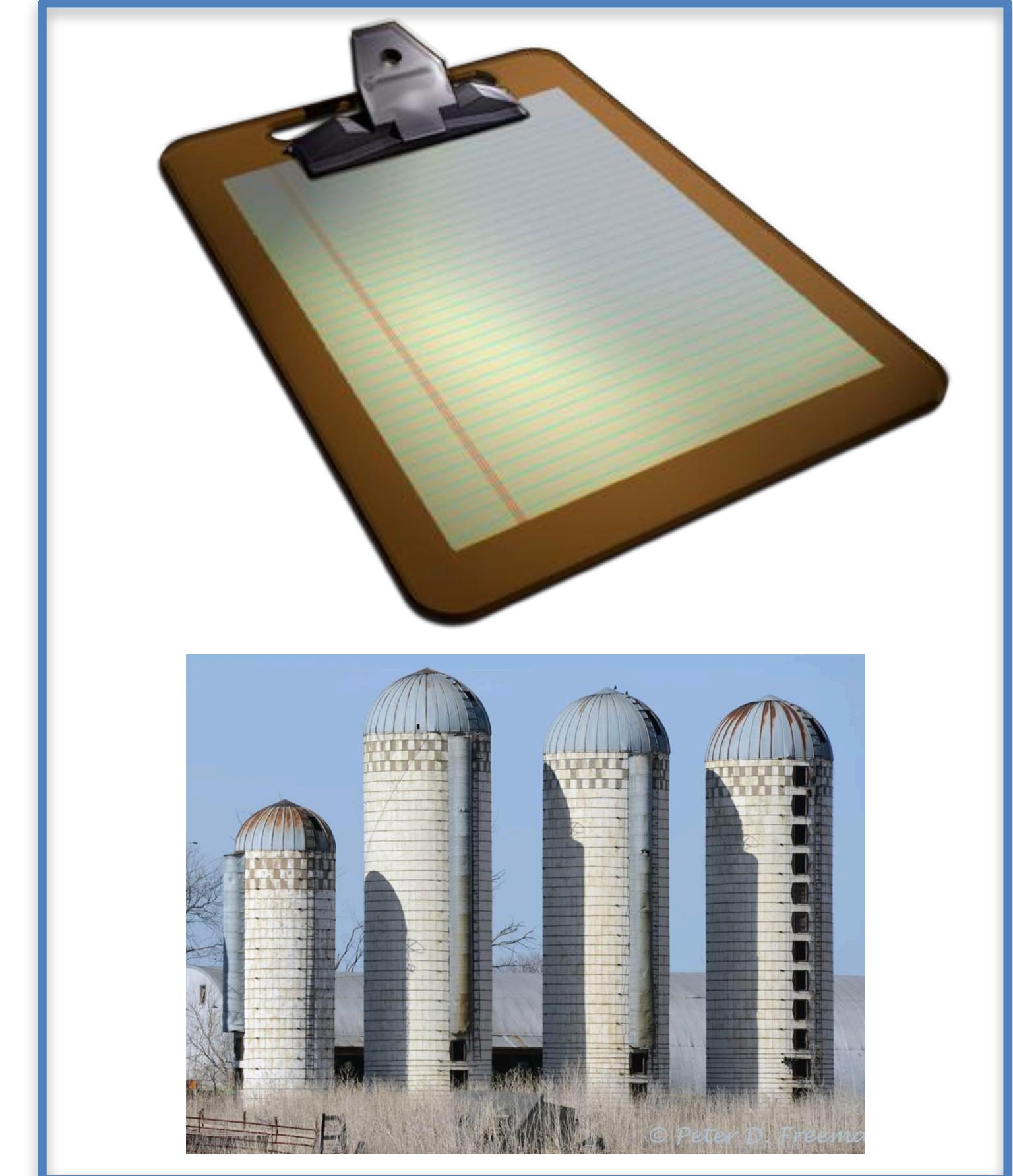


THE UNIVERSITY OF  
CHICAGO MEDICINE &  
BIOLOGICAL SCIENCES



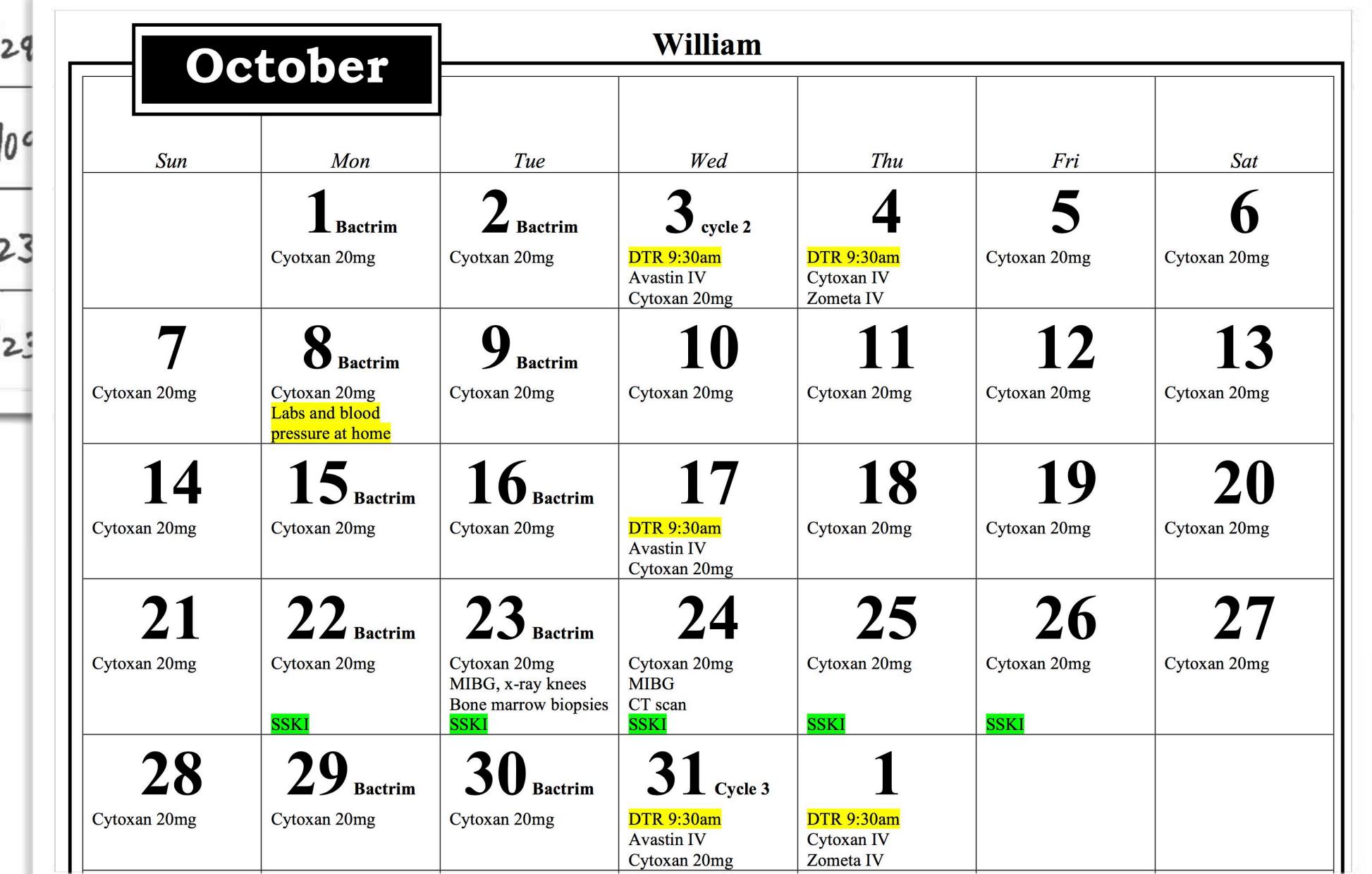
CENTER FOR  
**RESEARCH**  
INFORMATICS

# Manual processes and lack of data standards plague clinical trials



# Manual processes and lack of data standards plague clinical trials

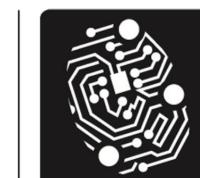
CTCAE CODE (per protocol)	CTCAE SHORT NAME (per protocol)	Current Grade	Maximum Grade This Course* (for occurrence)	Maximum Grade This Occurrence **	Attribu-tion	Date Onset	Resolved Yes / No	Date Resolved	Comments
10020943	Hypo albuminemia	0	1	1	3	05/23	Yes	06/02	
10020943	Hypo albuminemia	0	1	1	3	06/05	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	05/06			
10021059	Hypo phosphatemia	0	1	1	3	05/23			
10016256	Fatigue	0	1	1	3	05/23			



Activation Date:	8/8/01	Version Date:	9-4-10	
<b>Address for shipping:</b>	the frozen GSH pellets is found section 6.3.6. Include NANT specimen transport with the shipment.			
GSH Timepoints:	Day -4: Prior to starting BSO bolus Day -2 (Hour 48): Prior to giving 2 <sup>nd</sup> dose of L-PAM			
6.3.3 Plasma for Melphalan Pharmacokinetics	Collect 3cc peripheral blood in a green top tube or preservative-free heparinized tube for Day -2 dose of melphalan. All specimens must be immediately sent to the laboratory for processing. Arrange in advance for specimen collection with patient's initials, date and time specimen was taken. The specimen is to accompany specimen to the laboratory. The specimen is to be centrifuged at 2500 RPM x 10 minutes and placed in polyethylene tubes, labeled "L-PAM" until analysis.			
Timepoints:	Day -4: Prior to starting BSO bolus Day -2 (Hour 48): Prior to giving 2 <sup>nd</sup> dose of L-PAM			
5.0				
5.1				
All pa (septic)				
5.2				
Patient				
<b>6.0 REQUIRED OBSERVATIONS</b>				
6.1 Clinical and Laboratory Studies	All entry/eligibility laboratory studies on blood and urine and baseline CXR must be performed weeks prior to study entry. CT/MRI for intracranial disease, tumor disease evaluation including MIBG scan, bilateral bone marrow biopsies and FET's (only if indicated) must be performed within 4 weeks of study entry. STMRH for intracranial disease, tumor disease evaluation including GFR, ICG-chloramphenicol or MUGA, and FET's (only if indicated) must be performed within 4 weeks of study entry (and after the most recent disease surveillance per your institutional guidelines). Please also provide any other relevant laboratory results. Please also provide any other relevant laboratory results. Please also provide any other relevant laboratory results.			
OBTAIN OTHER STUDIES AS NEEDED FOR GOOD PATIENT CARE	CXR must be performed within 4 weeks prior to study entry. CT/MRI for intracranial disease, tumor disease evaluation including MIBG scan, bilateral bone marrow biopsies and FET's (only if indicated) must be performed within 4 weeks of study entry. STMRH for intracranial disease, tumor disease evaluation including GFR, ICG-chloramphenicol or MUGA, and FET's (only if indicated) must be performed within 4 weeks of study entry (and after the most recent disease surveillance per your institutional guidelines). Please also provide any other relevant laboratory results. Please also provide any other relevant laboratory results. Please also provide any other relevant laboratory results.			
Observation	Pre-Study	Day -4 through Day 0	Day 0 through Day +28	
		QD	Weekly	
		2x/ week until ANC > 500 & PR count > 20,000 x 3 days	2x/ week	
5.2.1 E BSO will				
University of Chicago; Comer Children's Hospital, 5841 South Maryland Avenue, Chicago, IL 60637 Susan Cohn, MD, Phone: (773) 702-2571; Email: <a href="mailto:scohn@pediatrics.bchicago.edu">scohn@pediatrics.bchicago.edu</a>				
Children's Healthcare of Atlanta, 2015 Up逮ate Drive, Atlanta, GA 30322 MD, Phone: (404) 785-0853; Email: <a href="mailto:Howard.katzenstein@choa.org">Howard.katzenstein@choa.org</a>				
Capital for Sick Children, 555 University Avenue, Toronto Ontario, Canada M5G 1X8 Baruchel, MD, Phone: (416) 813-7795; Email: <a href="mailto:sylvain.baruchel@ Sickkids.ca">sylvain.baruchel@ Sickkids.ca</a>				
Children's 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: <a href="mailto:jgl59@columbia.edu">jgl59@columbia.edu</a>				
Cook Children's Medical Center, 901 Seventh Avenue, Fort Worth, TX 76104 Granger, MD, Phone: (682) 885-2580; Email: <a href="mailto:mgranger@cookchildrens.org">mgranger@cookchildrens.org</a>				
<b>Protocol Chairperson</b>				
Judith G. Villalba, M.D. Childrens Hospital Los Angeles 4650 Sunset Blvd., MS-54, Los Angeles, CA 90027 Phone (323) 361-5654 or (323) 361-5687 Email: <a href="mailto:jvillalba@chla.usc.edu">jvillalba@chla.usc.edu</a>				
2010	1			
4.1.7 Performance Status	Patients must have a performance status of 0 or 1 (Appendix I) and a life expectancy of ≥ 2 months.			
4.1.8 Organ Function				
4.1.8.1 Kidney Function	Patients must have adequate renal function defined as a glomerular filtration rate (GFR) or 12 hour urine collection for creatinine clearance ≥ 100 milliliter/1.73m <sup>2</sup> AND a serum creatinine <= 1.5 x normal for age.			
Age Range	Serum Creatinine			
≤ 5 years	≤ 0.8 mg/dl			
10	11	12	13	17
<b>5.2.2 Melphalan (L-PAM)</b>	The dose level of melphalan will be assigned at study entry onto protocol as outlined in Table 1 and as defined in Section 11.3. The dose for the initial cohort of patients will be defined as Level 1a.			
Table 1.				
Dose Level	L-PAM dose/day x 2 days	Total L-PAM dose		
0a	15 mg/m <sup>2</sup> /day	30 mg/m <sup>2</sup>		
1a	20 mg/m <sup>2</sup> /day	40 mg/m <sup>2</sup>		
2a	25 mg/m <sup>2</sup> /day	50 mg/m <sup>2</sup>		
3a	31 mg/m <sup>2</sup> /day	62 mg/m <sup>2</sup>		
4a	40 mg/m <sup>2</sup> /day	80 mg/m <sup>2</sup>		
a	50 mg/m <sup>2</sup> /day	100 mg/m <sup>2</sup>		
	62.5 mg/m <sup>2</sup> /day	125 mg/m <sup>2</sup>		
is administered at hour 0 on day -3 and day -2. Hour 0 is defined as 24 hours after the				
an IV infusion through a central venous catheter over a minimum of 15 minutes. It is less than 10 mg/min but within one hour of reconstitution on Day -3 and Day -2. A new batch of melphalan must be prepared. The diluent propylene glycol must be hypotension or arrhythmia must be present. The diluent propylene glycol must be administered at least 7 days after completing effective therapy to prevent skin contact or inhalation of aerosolized particles of the drug. Then approximately 30 minutes prior to melphalan injection and again every 4 hours for a minimum of 24 hours after last melphalan dose.				
ystatin or Sulfa methoxazole/Trimethoprim) must be given prior to and following melphalan administration due to bladder irritation due to excretion of tumor lysis products. Minimal urine output immediately after melphalan administration should be more than 90 ml/min/hour. To effect				
IV lines for BSO and melphalan to avoid interruptions of drug infusions, if been infused through same line by stopping BSO only during time of giving BSO. All pharmacokinetics must be drawn through a separate IV line.				
we care shall be given in accordance to Section 8.0, zed until ANC ≥ 500 for 3 consecutive counts				
E CD34+ cells/kg (optimum 5 x 10 <sup>6</sup> /kg) must be infused, and PBSC is required when <1.5 x 10 <sup>6</sup> CD34+ Viable Purged				
γluclear cells/kg (optimum 5 x 10 <sup>6</sup> /kg) must be infused, with immunomagnetic beads.				

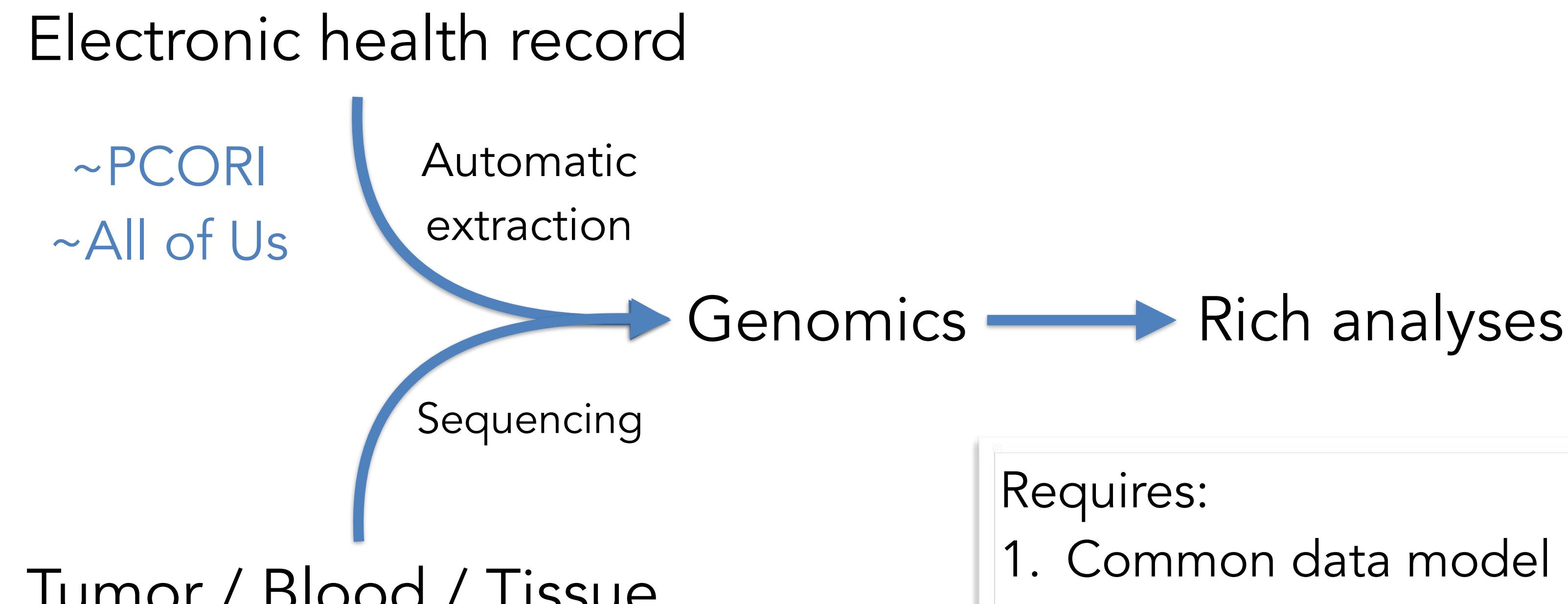
# Lack of rich phenotype data hinders progress

- 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

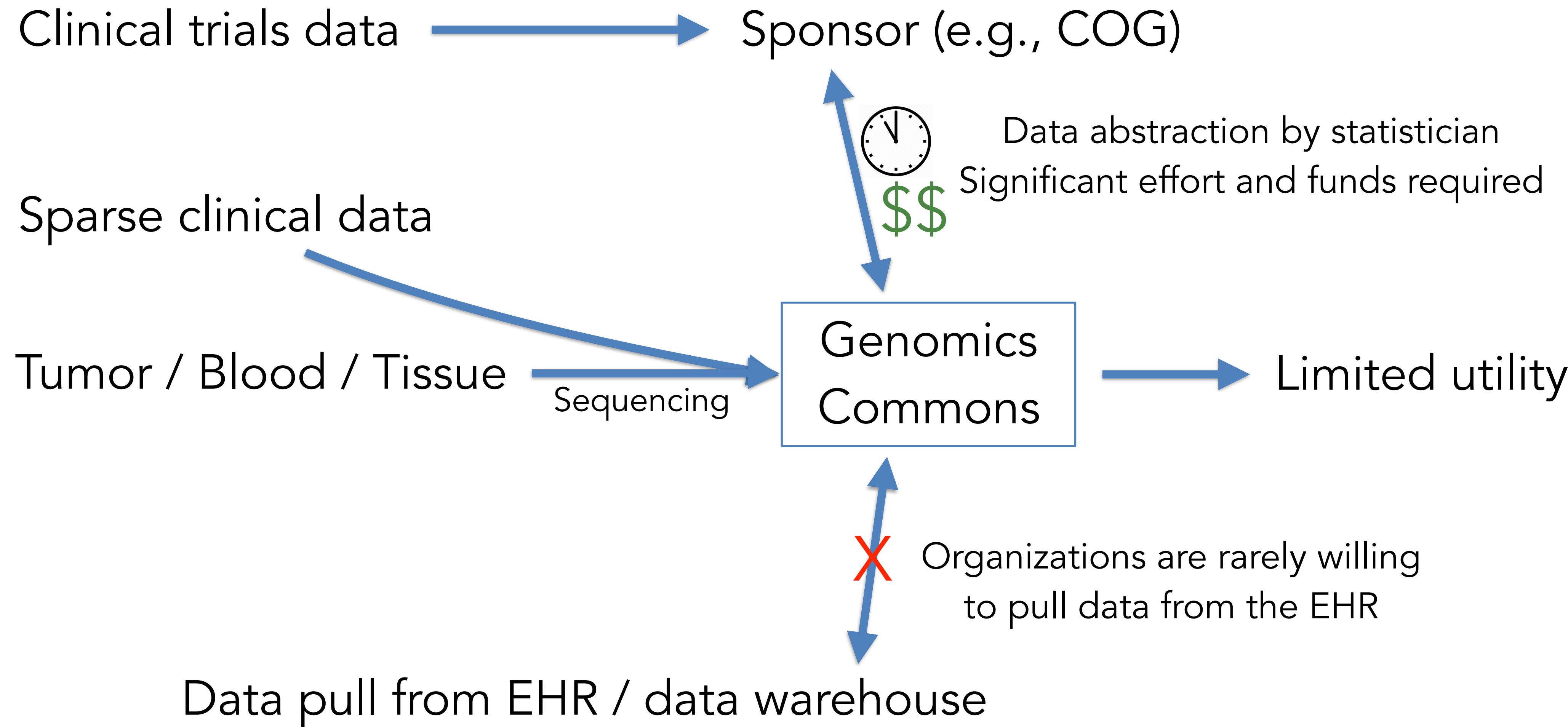


Requires:

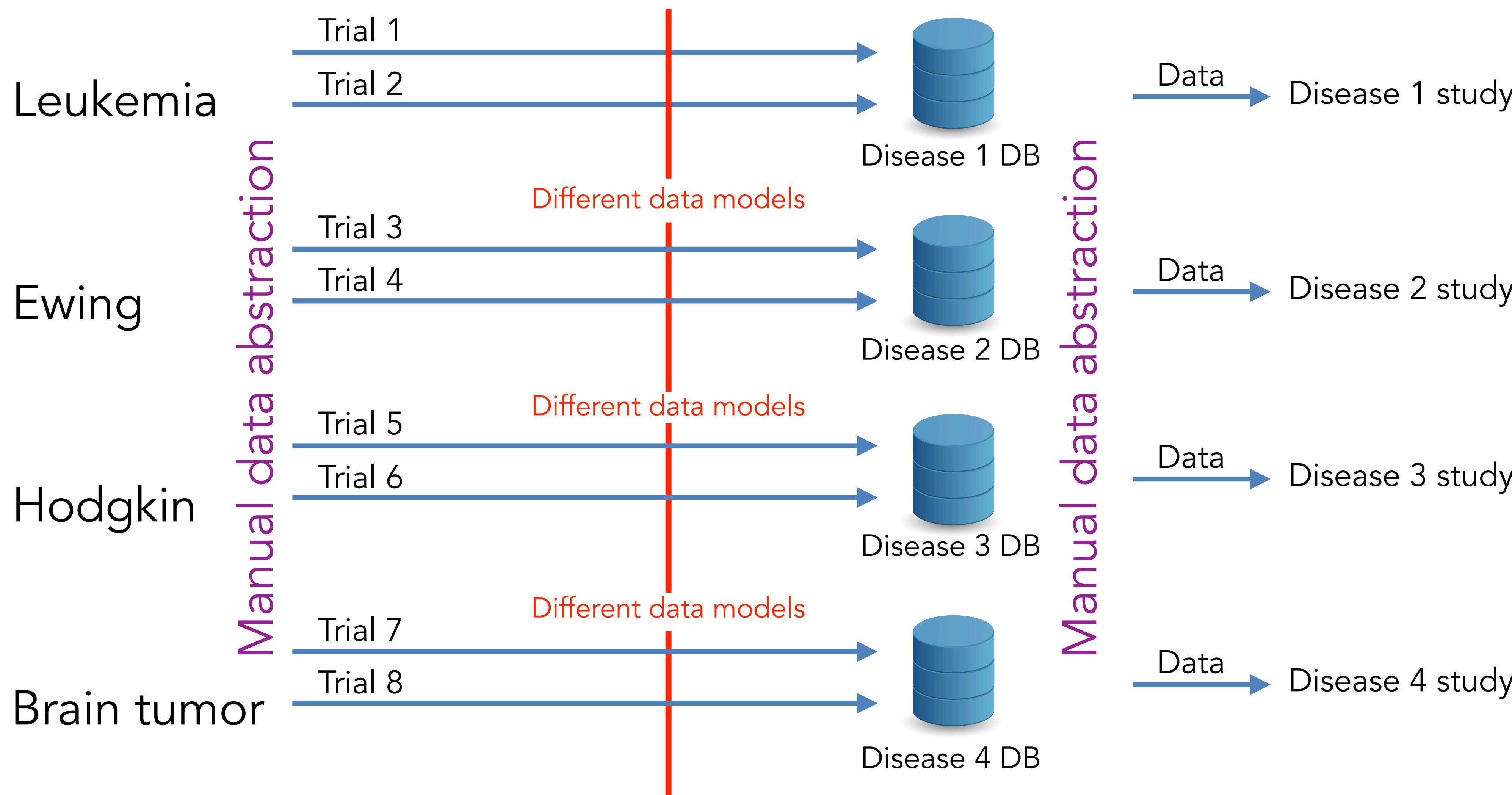
1. Common data model
2. Broad consent
3. State-of-the-art data warehouse
4. Foresight and planning



# Current state for genomic commons



# Current paradigm for COG



# Goal

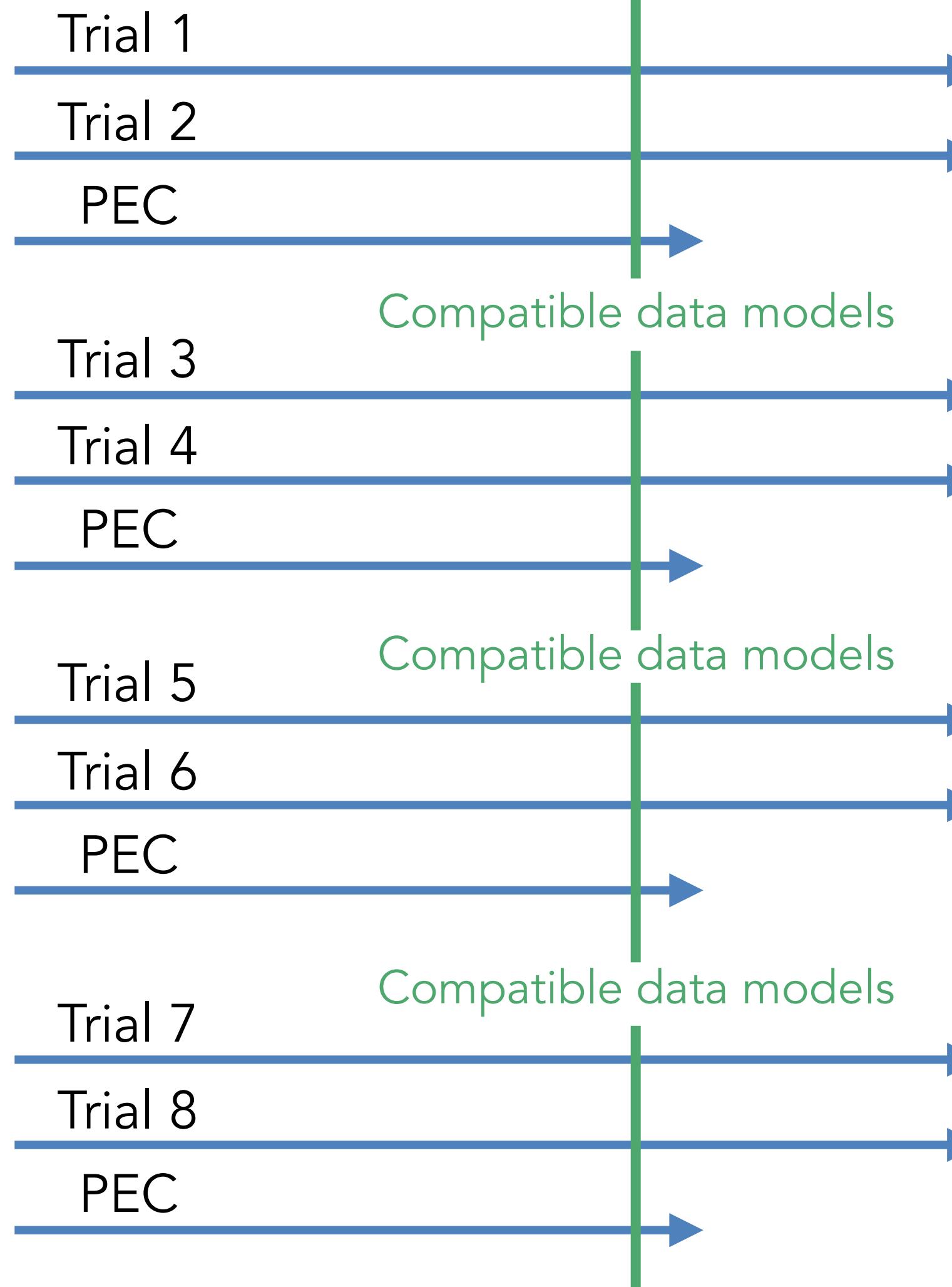
Leukemia

Ewing

Hodgkin

Brain tumor

**Manual data abstraction**

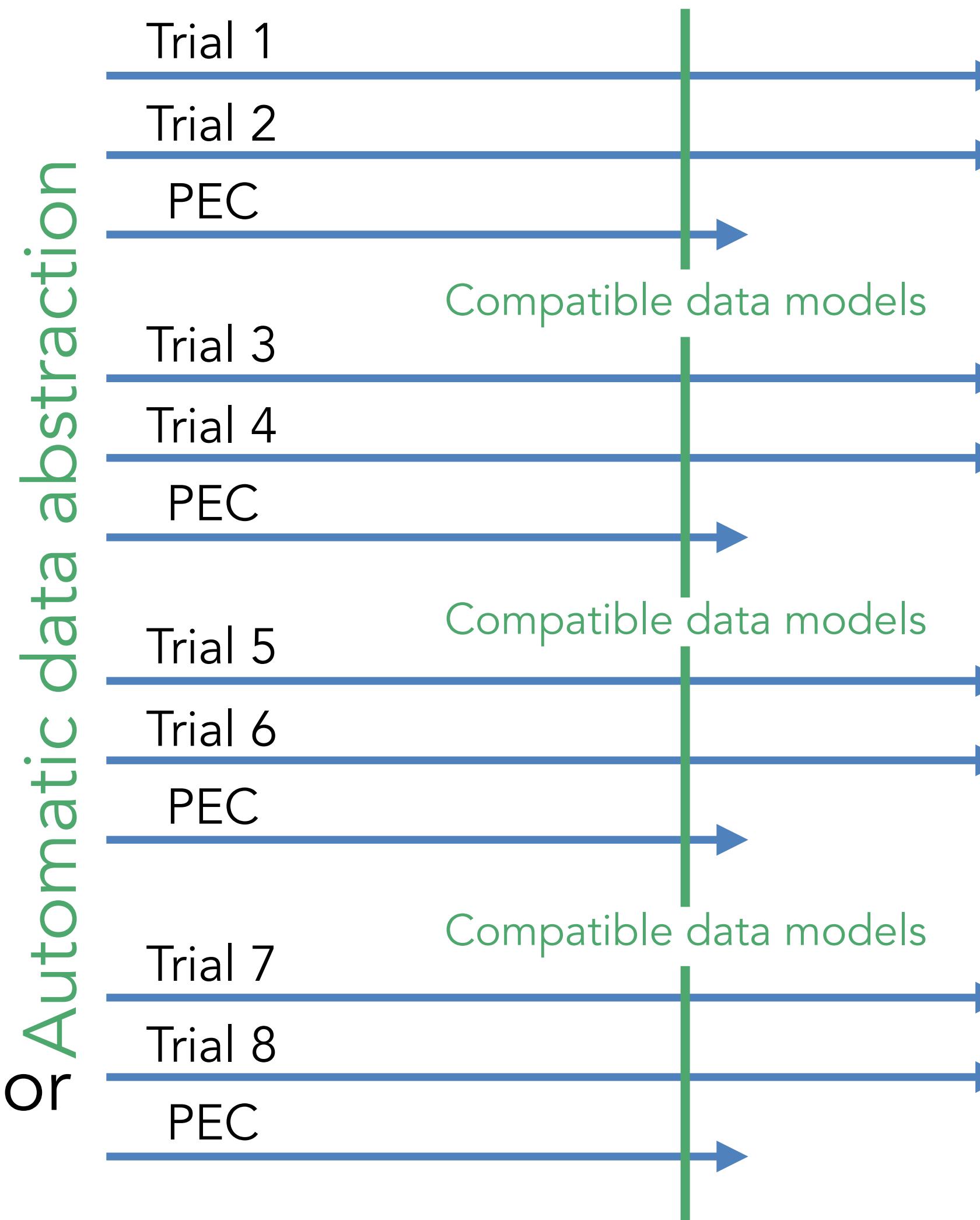


**Automatic data abstraction**



# Goal

Leukemia



Automatic data abstraction

- Data → Disease 1 study
- Data → Disease 2 study
- Data → Disease 3 study
- Data → Disease 4 study

Hodgkin

Brain tumor



THE UNIVERSITY OF  
CHICAGO MEDICINE &  
BIOLOGICAL SCIENCES



CENTER FOR  
**RESEARCH**  
INFORMATICS

# 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 (NWTS)

Intergroup Rhabdomyosarcoma Study Group (IRSG)

2000

Children's Oncology Group (COG)



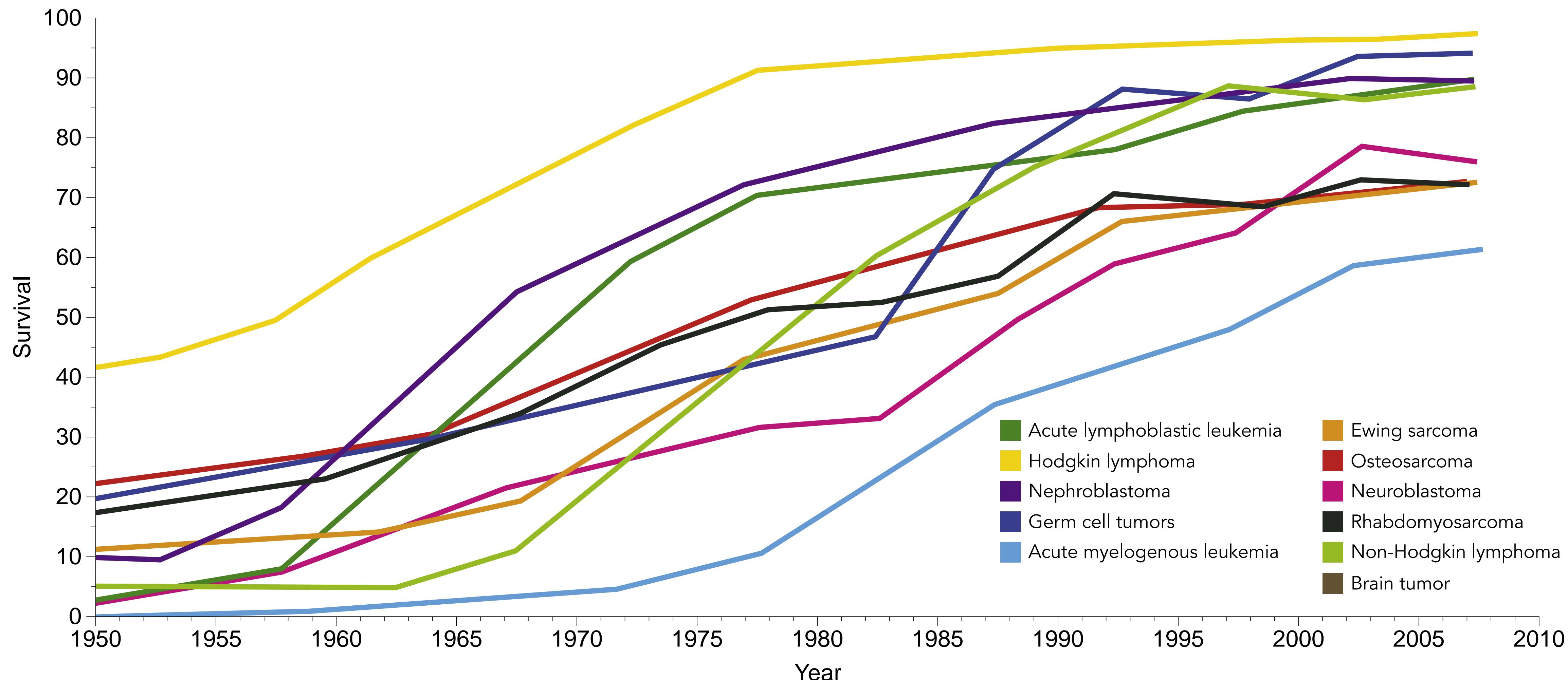
THE UNIVERSITY OF  
CHICAGO MEDICINE &  
BIOLOGICAL SCIENCES



CENTER FOR  
**RESEARCH**  
INFORMATICS



# Survival - Pediatric cancer - still a long way to go



# Neuroblastoma data commons

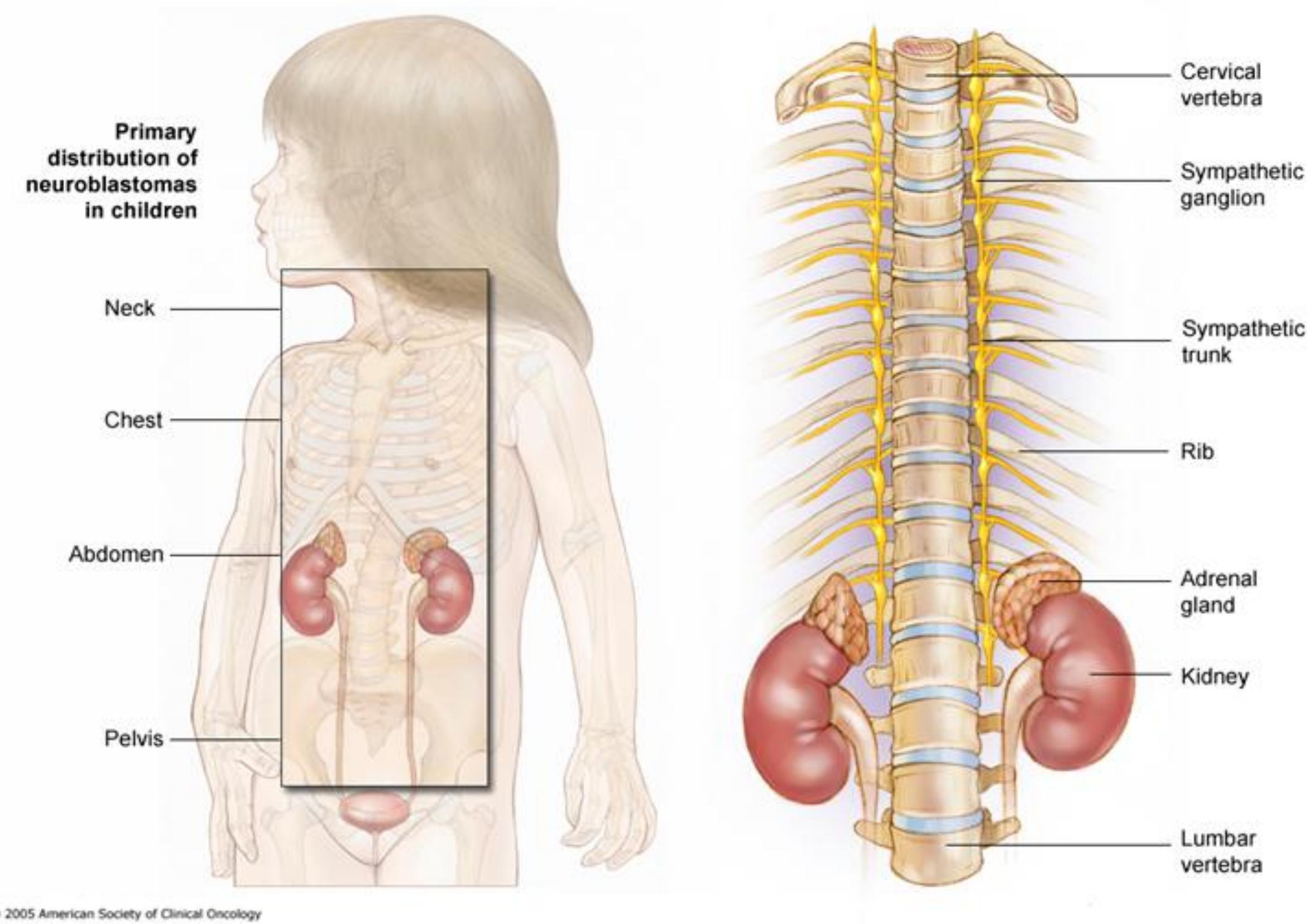


THE UNIVERSITY OF  
CHICAGO MEDICINE &  
BIOLOGICAL SCIENCES



CENTER FOR  
**RESEARCH**  
INFORMATICS

# Neuroblastoma



The most common solid tumor in children



THE UNIVERSITY OF  
CHICAGO MEDICINE &  
BIOLOGICAL SCIENCES



CENTER FOR  
**RESEARCH**  
INFORMATICS

# International Neuroblastoma Research Group (2004)



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)



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

The good news: 8800 patients

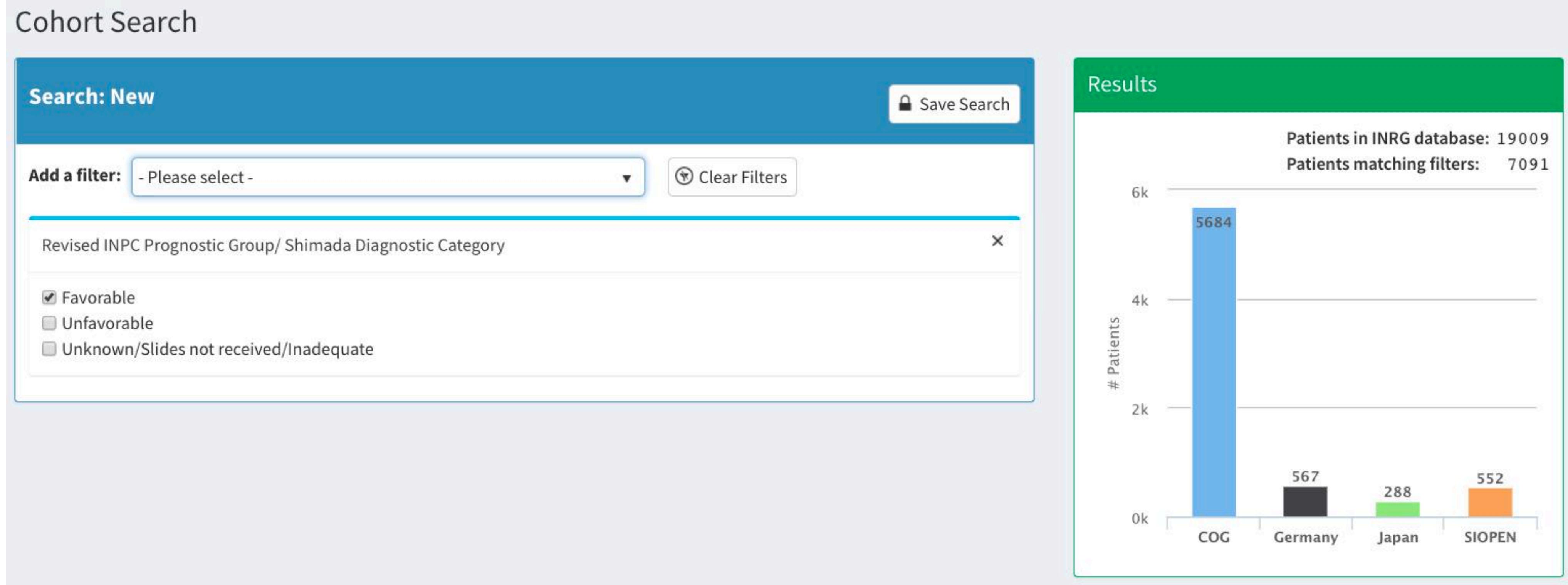
The bad news:  
Data in Excel

>20 high-impact publications  
that changed clinical practice

Age	Year	F	M	Cohort	ExStages	Scans	Initial	Year	INSS\_Stg	Event\_Mtg	Ip\_distr	Incl\_m	Surv	Prob\_1y	Prob\_5y	Prob\_10y	Prob\_15y	Prob\_20y	Prob\_25y	Prob\_30y	Prob\_35y	Prob\_40y	Prob\_45y	Prob\_50y	Prob\_55y	Prob\_60y	Prob\_65y	Prob\_70y	Prob\_75y	Prob\_80y	Prob\_85y	Prob\_90y	Prob\_95y	Prob\_100y	Prob\_105y	Prob\_110y	Prob\_115y	Prob\_120y	Prob\_125y	Prob\_130y	Prob\_135y	Prob\_140y	Prob\_145y	Prob\_150y	Prob\_155y	Prob\_160y	Prob\_165y	Prob\_170y	Prob\_175y	Prob\_180y	Prob\_185y	Prob\_190y	Prob\_195y	Prob\_200y	Prob\_205y	Prob\_210y	Prob\_215y	Prob\_220y	Prob\_225y	Prob\_230y	Prob\_235y	Prob\_240y	Prob\_245y	Prob\_250y	Prob\_255y	Prob\_260y	Prob\_265y	Prob\_270y	Prob\_275y	Prob\_280y	Prob\_285y	Prob\_290y	Prob\_295y	Prob\_300y	Prob\_305y	Prob\_310y	Prob\_315y	Prob\_320y	Prob\_325y	Prob\_330y	Prob\_335y	Prob\_340y	Prob\_345y	Prob\_350y	Prob\_355y	Prob\_360y	Prob\_365y	Prob\_370y	Prob\_375y	Prob\_380y	Prob\_385y	Prob\_390y	Prob\_395y	Prob\_400y	Prob\_405y	Prob\_410y	Prob\_415y	Prob\_420y	Prob\_425y	Prob\_430y	Prob\_435y	Prob\_440y	Prob\_445y	Prob\_450y	Prob\_455y	Prob\_460y	Prob\_465y	Prob\_470y	Prob\_475y	Prob\_480y	Prob\_485y	Prob\_490y	Prob\_495y	Prob\_500y	Prob\_505y	Prob\_510y	Prob\_515y	Prob\_520y	Prob\_525y	Prob\_530y	Prob\_535y	Prob\_540y	Prob\_545y	Prob\_550y	Prob\_555y	Prob\_560y	Prob\_565y	Prob\_570y	Prob\_575y	Prob\_580y	Prob\_585y	Prob\_590y	Prob\_595y	Prob\_600y	Prob\_605y	Prob\_610y	Prob\_615y	Prob\_620y	Prob\_625y	Prob\_630y	Prob\_635y	Prob\_640y	Prob\_645y	Prob\_650y	Prob\_655y	Prob\_660y	Prob\_665y	Prob\_670y	Prob\_675y	Prob\_680y	Prob\_685y	Prob\_690y	Prob\_695y	Prob\_700y	Prob\_705y	Prob\_710y	Prob\_715y	Prob\_720y	Prob\_725y	Prob\_730y	Prob\_735y	Prob\_740y	Prob\_745y	Prob\_750y	Prob\_755y	Prob\_760y	Prob\_765y	Prob\_770y	Prob\_775y	Prob\_780y	Prob\_785y	Prob\_790y	Prob\_795y	Prob\_800y	Prob\_805y	Prob\_810y	Prob\_815y	Prob\_820y	Prob\_825y	Prob\_830y	Prob\_835y	Prob\_840y	Prob\_845y	Prob\_850y	Prob\_855y	Prob\_860y	Prob\_865y	Prob\_870y	Prob\_875y	Prob\_880y	Prob\_885y	Prob\_890y	Prob\_895y	Prob\_900y	Prob\_905y	Prob\_910y	Prob\_915y	Prob\_920y	Prob\_925y	Prob\_930y	Prob\_935y	Prob\_940y	Prob\_945y	Prob\_950y	Prob\_955y	Prob\_960y	Prob\_965y	Prob\_970y	Prob\_975y	Prob\_980y	Prob\_985y	Prob\_990y	Prob\_995y	Prob\_1000y	Prob\_1005y	Prob\_1010y	Prob\_1015y	Prob\_1020y	Prob\_1025y	Prob\_1030y	Prob\_1035y	Prob\_1040y	Prob\_1045y	Prob\_1050y	Prob\_1055y	Prob\_1060y	Prob\_1065y	Prob\_1070y	Prob\_1075y	Prob\_1080y	Prob\_1085y	Prob\_1090y	Prob\_1095y	Prob\_1100y	Prob\_1105y	Prob\_1110y	Prob\_1115y	Prob\_1120y	Prob\_1125y	Prob\_1130y	Prob\_1135y	Prob\_1140y	Prob\_1145y	Prob\_1150y	Prob\_1155y	Prob\_1160y	Prob\_1165y	Prob\_1170y	Prob\_1175y	Prob\_1180y	Prob\_1185y	Prob\_1190y	Prob\_1195y	Prob\_1200y	Prob\_1205y	Prob\_1210y	Prob\_1215y	Prob\_1220y	Prob\_1225y	Prob\_1230y	Prob\_1235y	Prob\_1240y	Prob\_1245y	Prob\_1250y	Prob\_1255y	Prob\_1260y	Prob\_1265y	Prob\_1270y	Prob\_1275y	Prob\_1280y	Prob\_1285y	Prob\_1290y	Prob\_1295y	Prob\_1300y	Prob\_1305y	Prob\_1310y	Prob\_1315y	Prob\_1320y	Prob\_1325y	Prob\_1330y	Prob\_1335y	Prob\_1340y	Prob\_1345y	Prob\_1350y	Prob\_1355y	Prob\_1360y	Prob\_1365y	Prob\_1370y	Prob\_1375y	Prob\_1380y	Prob\_1385y	Prob\_1390y	Prob\_1395y	Prob\_1400y	Prob\_1405y	Prob\_1410y	Prob\_1415y	Prob\_1420y	Prob\_1425y	Prob\_1430y	Prob\_1435y	Prob\_1440y	Prob\_1445y	Prob\_1450y	Prob\_1455y	Prob\_1460y	Prob\_1465y	Prob\_1470y	Prob\_1475y	Prob\_1480y	Prob\_1485y	Prob\_1490y	Prob\_1495y	Prob\_1500y	Prob\_1505y	Prob\_1510y	Prob\_1515y	Prob\_1520y	Prob\_1525y	Prob\_1530y	Prob\_1535y	Prob\_1540y	Prob\_1545y	Prob\_1550y	Prob\_1555y	Prob\_1560y	Prob\_1565y	Prob\_1570y	Prob\_1575y	Prob\_1580y	Prob\_1585y	Prob\_1590y	Prob\_1595y	Prob\_1600y	Prob\_1605y	Prob\_1610y	Prob\_1615y	Prob\_1620y	Prob\_1625y	Prob\_1630y	Prob\_1635y	Prob\_1640y	Prob\_1645y	Prob\_1650y	Prob\_1655y	Prob\_1660y	Prob\_1665y	Prob\_1670y	Prob\_1675y	Prob\_1680y	Prob\_1685y	Prob\_1690y	Prob\_1695y	Prob\_1700y	Prob\_1705y	Prob\_1710y	Prob\_1715y	Prob\_1720y	Prob\_1725y	Prob\_1730y	Prob\_1735y	Prob\_1740y	Prob\_1745y	Prob\_1750y	Prob\_1755y	Prob\_1760y	Prob\_1765y	Prob\_1770y	Prob\_1775y	Prob\_1780y	Prob\_1785y	Prob\_1790y	Prob\_1795y	Prob\_1800y	Prob\_1805y	Prob\_1810y	Prob\_1815y	Prob\_1820y	Prob\_1825y	Prob\_1830y	Prob\_1835y	Prob\_1840y	Prob\_1845y	Prob\_1850y	Prob\_1855y	Prob\_1860y	Prob\_1865y	Prob\_1870y	Prob\_1875y	Prob\_1880y	Prob\_1885y	Prob\_1890y	Prob\_1895y	Prob\_1900y	Prob\_1905y	Prob\_1910y	Prob\_1915y	Prob\_1920y	Prob\_1925y	Prob\_1930y	Prob\_1935y	Prob\_1940y	Prob\_1945y	Prob\_1950y	Prob\_1955y	Prob\_1960y	Prob\_1965y	Prob\_1970y	Prob\_1975y	Prob\_1980y	Prob\_1985y	Prob\_1990y	Prob\_1995y	Prob\_2000y	Prob\_2005y	Prob\_2010y	Prob\_2015y	Prob\_2020y	Prob\_2025y	Prob\_2030y	Prob\_2035y	Prob\_2040y	Prob\_2045y	Prob\_2050y	Prob\_2055y	Prob\_2060y	Prob\_2065y	Prob\_2070y	Prob\_2075y	Prob\_2080y	Prob\_2085y	Prob\_2090y	Prob\_2095y	Prob\_2100y	Prob\_2105y	Prob\_2110y	Prob\_2115y	Prob\_2120y	Prob\_2125y	Prob\_2130y	Prob\_2135y	Prob\_2140y	Prob\_2145y	Prob\_2150y	Prob\_2155y	Prob\_2160y	Prob\_2165y	Prob\_2170y	Prob\_2175y	Prob\_2180y	Prob\_2185y	Prob\_2190y	Prob\_2195y	Prob\_2200y	Prob\_2205y	Prob\_2210y	Prob\_2215y	Prob\_2220y	Prob\_2225y	Prob\_2230y	Prob\_2235y	Prob\_2240y	Prob\_2245y	Prob\_2250y	Prob\_2255y	Prob\_2260y	Prob\_2265y	Prob\_2270y	Prob\_2275y	Prob\_2280y	Prob\_2285y	Prob\_2290y	Prob\_2295y	Prob\_2300y	Prob\_2305y	Prob\_2310y	Prob\_2315y	Prob\_2320y	Prob\_2325y	Prob\_2330y	Prob\_2335y	Prob\_2340y	Prob\_2345y	Prob\_2350y	Prob\_2355y	Prob\_2360y	Prob\_2365y	Prob\_2370y	Prob\_2375y	Prob\_2380y	Prob\_2385y	Prob\_2390y	Prob\_2395y	Prob\_2400y	Prob\_2405y	Prob\_2410y	Prob\_2415y	Prob\_2420y	Prob\_2425y	Prob\_2430y	Prob\_2435y	Prob\_2440y	Prob\_2445y	Prob\_2450y	Prob\_2455y	Prob\_2460y	Prob\_2465y	Prob\_2470y	Prob\_2475y	Prob\_2480y	Prob\_2485y	Prob\_2490y	Prob\_2495y	Prob\_2500y	Prob\_2505y	Prob\_2510y	Prob\_2515y	Prob\_2520y	Prob\_2525y	Prob\_2530y	Prob\_2535y	Prob\_2540y	Prob\_2545y	Prob\_2550y	Prob\_2555y	Prob\_2560y	Prob\_2565y	Prob\_2570y	Prob\_2575y	Prob\_2580y	Prob\_2585y	Prob\_2590y	Prob\_2595y	Prob\_2600y	Prob\_2605y	Prob\_2610y	Prob\_2615y	Prob\_2620y	Prob\_2625y	Prob\_2630y	Prob\_2635y	Prob\_2640y	Prob\_2645y	Prob\_2650y	Prob\_2655y	Prob\_2660y	Prob\_2665y	Prob\_2670y	Prob\_2675y	Prob\_2680y	Prob\_2685y	Prob\_2690y	Prob\_2695y	Prob\_2700y	Prob\_2705y	Prob\_2710y	Prob\_2715y	Prob\_2720y	Prob\_2725y	Prob\_2730y	Prob\_2735y	Prob\_2740y	Prob\_2745y	Prob\_2750y	Prob\_2755y	Prob\_2760y	Prob\_2765y	Prob\_2770y	Prob\_2775y	Prob\_2780y	Prob\_2785y	Prob\_2790y	Prob\_2795y	Prob\_2800y	Prob\_2805y	Prob\_2810y	Prob\_2815y	Prob\_2820y	Prob\_2825y	Prob\_2830y	Prob\_2835y	Prob\_2840y	Prob\_2845y	Prob\_2850y	Prob\_2855y	Prob\_2860y	Prob\_2865y	Prob\_2870y	Prob\_2875y	Prob\_2880y	Prob\_2885y	Prob\_2890y	Prob\_2895y	Prob\_2900y	Prob\_2905y	Prob\_2910y	Prob\_2915y	Prob\_2920y	Prob\_2925y	Prob\_2930y	Prob\_2935y	Prob\_2940y	Prob\_2945y	Prob\_2950y	Prob\_2955y	Prob\_2960y	Prob\_2965y	Prob\_2970y	Prob\_2975y	Prob\_2980y	Prob\_2985y	Prob\_2990y	Prob\_2995y	Prob\_3000y	Prob\_3005

# Neuroblastoma Commons Cohort Discovery

Example: Favorable tumor biology, tissue available



# Neuroblastoma Commons Cohort Discovery

Cohort Search

Search: New

Add a filter: - Please select -

- Please select -
- Revised INPC Prognostic Group
- Favorable
- Unfavorable
- Unknown
- Primary Tumor-Auricular
- 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

Save Search

Clear Filters

Results

Patients in INRG database: 19009  
Patients matching filters: 7091

# Patients

COG 5684

Germany 567

Japan 288

SIOPEN 552

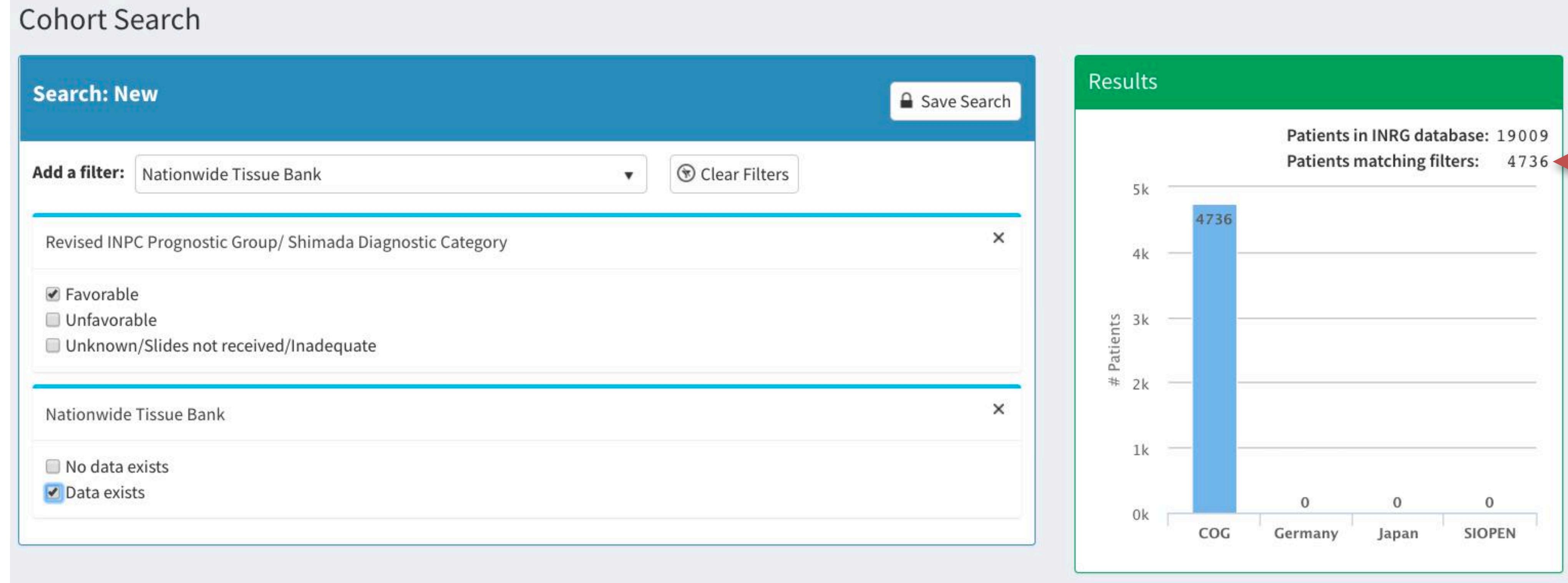
Immediately see cohort counts

Links to external data sets

The screenshot shows a search interface for a neuroblastoma cohort. On the left, a sidebar lists various clinical and genomic filters. A dropdown menu is open under 'Add a filter' with several options like 'Revised INPC Prognostic Group' and 'Favorable' checked. On the right, a bar chart titled 'Results' displays patient counts for different cohorts: COG (5684), Germany (567), Japan (288), and SIOPEN (552). Arrows point from the 'External Data' links in the sidebar to the chart, indicating they lead to more information about these cohorts.



# Neuroblastoma Commons Cohort Discovery



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

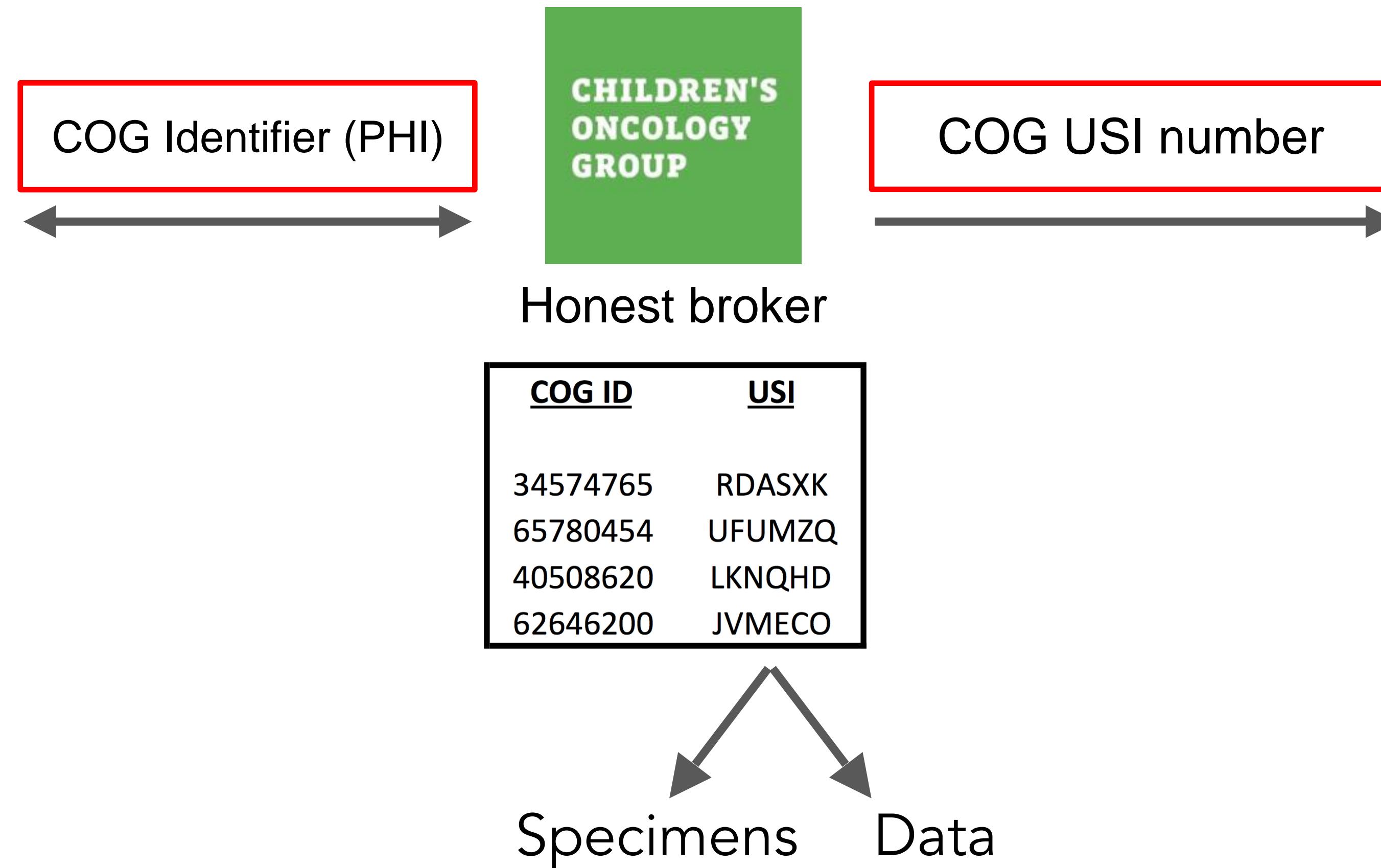
Example: Favorable biology, tissue available



# COG Unique Specimen Identifier (USI)



COG institution



NATIONAL CANCER INSTITUTE  
GENOMIC DATA COMMONS



CAVATICA

THE UNIVERSITY OF  
CHICAGO



Commons



THE UNIVERSITY OF  
CHICAGO MEDICINE &  
BIOLOGICAL SCIENCES



CENTER FOR  
**RESEARCH**  
INFORMATICS

# Neuroblastoma data commons growth

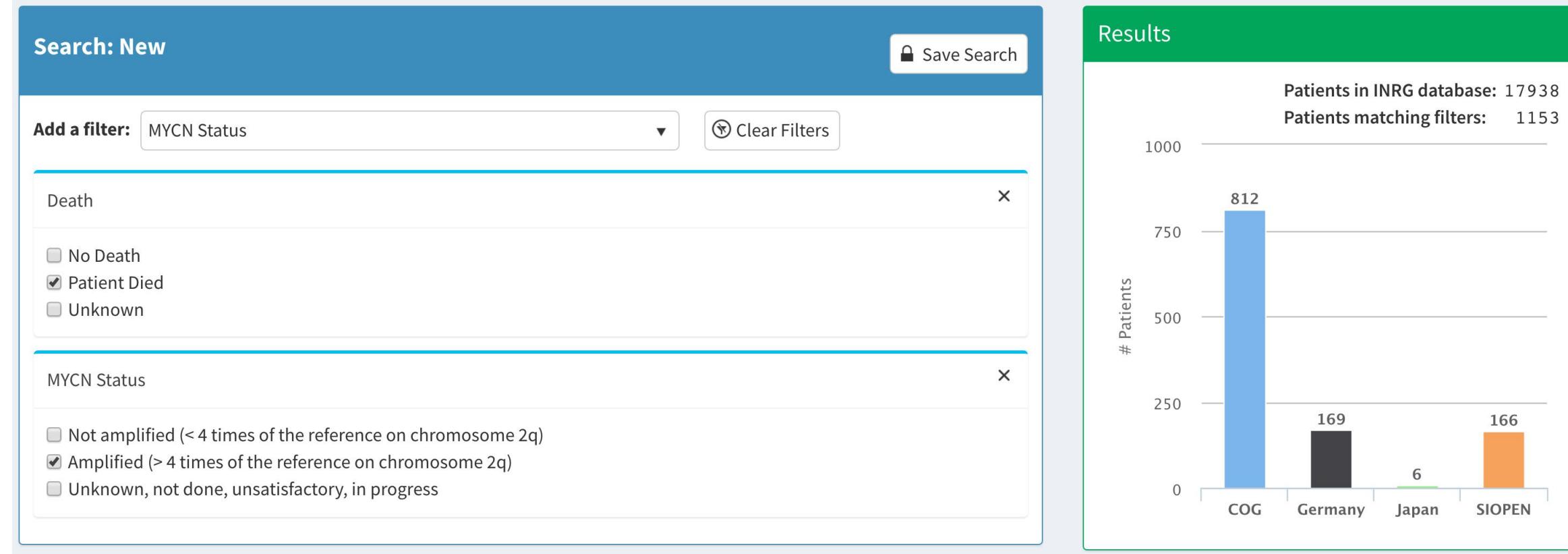
Year	COG	SIOPEN	GPOH	Japan	Total
2004	4235	2157	1938	470	<b>8800</b>
2012	6127	2504	1938	470	<b>11039</b>
2013	11642	2504	1938	470	<b>16554</b>
2015	13060	2504	1938	470	<b>17972</b>
2016	13937	2664	1938	470	<b>19009</b>
2018	14425	3397	2154	470	<b>20446</b>

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



# Neuroblastoma Data Commons

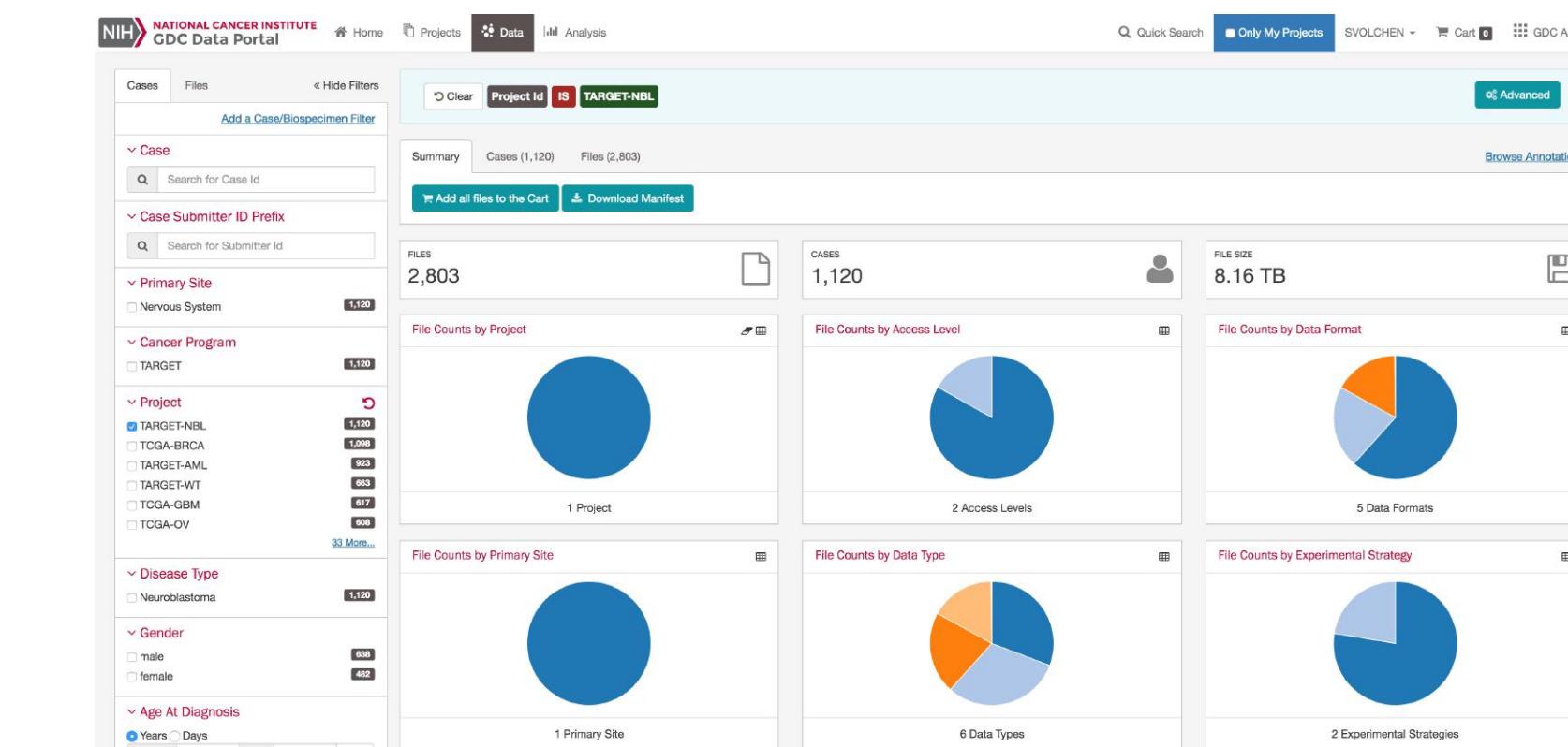
## Cohort Search



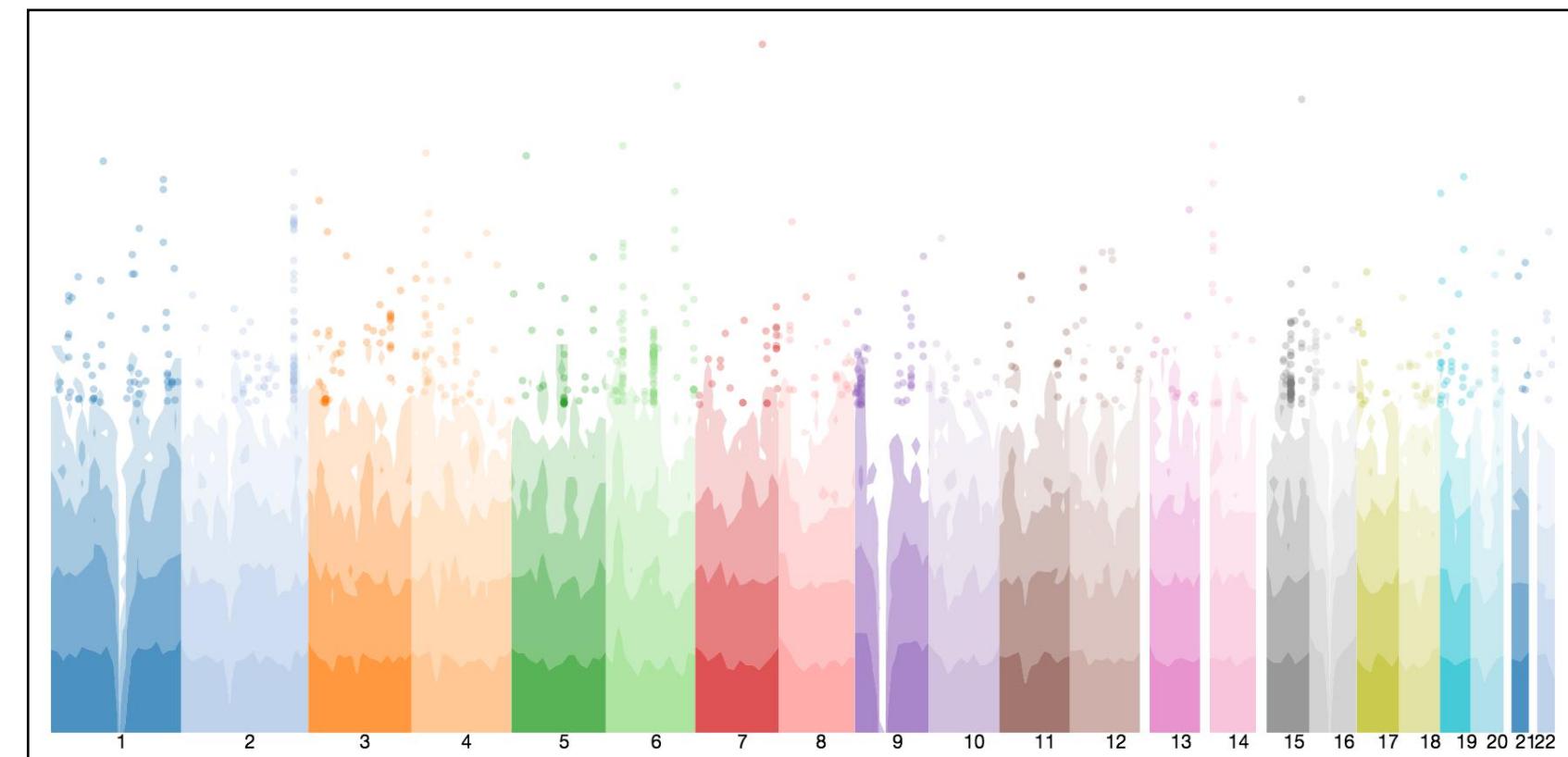
INRG cohort discovery

```
> registerJavaRodaum()
> idir <- "/home/ubuntu/demo_inrg"
> pheno.file <- file.path(idir, "meta", "clinical.deseq_fmt.filtered.txt")
> count.dir <- file.path(idir, "files", "fmt")
> clin.df <- read.delim(pheno.file, header=TRUE, sep="\t", stringsAsFactors = FALSE)
> count.files <- unlist(lapply(clin.df$pt_id, function(x) paste0(x, ".htseq.counts")))
> count.files
[1] "PAIPGU.htseq.counts" "PAISNS.htseq.counts" "PAITCI.htseq.counts"
[4] "PAIVHE.htseq.counts" "PAIXIF.htseq.counts" "PAKYZS.htseq.counts"
[7] "PALBFW.htseq.counts" "PALCBW.htseq.counts" "PALETP.htseq.counts"
[10] "PALEVG.htseq.counts" "PALIIN.htseq.counts" "PALUKC.htseq.counts"
[13] "PALNVP.htseq.counts" "PALTEG.htseq.counts" "PALUYS.htseq.counts"
[16] "PALVKK.htseq.counts" "PALWVJ.htseq.counts" "PALXTB.htseq.counts"
[19] "PALZZV.htseq.counts" "PAMXF.htseq.counts" "PAMNLH.htseq.counts"
[22] "PAMVRA.htseq.counts" "PAMYCE.htseq.counts" "PAMZGT.htseq.counts"
[25] "PAMZMG.htseq.counts" "PANBJH.htseq.counts" "PANBMJ.htseq.counts"
[28] "PANBSP.htseq.counts" "PANGKK.htseq.counts" "PANLH.htseq.counts"
[31] "PANKFE.htseq.counts" "PANNMS.htseq.counts" "PANRW.htseq.counts"
[34] "PANUVK.htseq.counts" "PANYGR.htseq.counts" "PANZPV.htseq.counts"
[37] "PANZVU.htseq.counts" "PAPBGH.htseq.counts" "PAPBZI.htseq.counts"
[40] "PAPEFE.htseq.counts" "PAPHPE.htseq.counts" "PAPKWN.htseq.counts"
[43] "PAPUAR.htseq.counts" "PAPUEB.htseq.counts" "PAPZYP.htseq.counts"
[46] "PARACM.htseq.counts"
> sample.table <- data.frame(sampleName=clin.df$pt_id,
+                               fileName=count.files,
+                               mycn=as.factor(clin.df$mycn))
> [
```

Command-line analysis



Genomic data commons search



Visualization



# 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

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



THE UNIVERSITY OF  
CHICAGO MEDICINE &  
BIOLOGICAL SCIENCES



CENTER FOR  
**RESEARCH**  
INFORMATICS

# Remapping NBL to a standard ontology

Current	Tumor type	Code
	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	<a href="#">C0017075</a>
	Ganglioneuroblastoma - intermixed	<a href="#">C1517444</a>
	Ganglioneuroma	<a href="#">C0017075</a>
	Ganglioneuroblastoma - nodular	<a href="#">C1517445</a>
	Unknown	<a href="#">C0087135</a>



# Ganglioneuroblastoma

NIH NATIONAL CANCER INSTITUTE [www.cancer.gov](http://www.cancer.gov)

EVS Enterprise Vocabulary Services

**NCI**metathesaurus

NCIm Version: 201706 Version 2.8 (using LexEVS 6.5)

unknown

Contains  Exact Match  Begins With  
 Name  Code  Property  Relationship

Source ALL Advanced Search

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 ( <a href="#">see NCI Thesaurus info</a> )
--------------------	---

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>



# 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	<a href="#">C0001625</a>
Retroperitoneal	<a href="#">CL318004</a>
Neck	<a href="#">C0027530</a>
Thorax	<a href="#">C0817096</a>
Pelvis	<a href="#">C0030797</a>
Not otherwise specified	<a href="#">C1518425</a>
Unknown	<a href="#">C0439673</a>



# 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



THE UNIVERSITY OF  
CHICAGO MEDICINE &  
BIOLOGICAL SCIENCES



CENTER FOR  
**RESEARCH**  
INFORMATICS

# Rhabdomyosarcoma data standardization

	Current	
	Code	Description
<b>SEX</b>	1	Male
	2	Female

	1	Alveolar rhabdomyosarcoma	
	2	Embryonal rhabdomyosarcoma	
	3	Botryoid rhabdomyosarcoma	
	4	Not otherwise specified	
	5	Undifferentiated sarcoma	
<b>HISTOLOGY</b>	6	Sarcoma, not classifiable	
	7	Spindle cell sarcoma	
	8	Ectomesenchymoma	
	9	Other	
	10	Mixed rhabdomyosarcoma	
	99	Unknown	

<b>Male</b>	<a href="#">C0086582</a>
<b>Female</b>	<a href="#">C0015780</a>
<b>Unknown</b>	<a href="#">C0439673</a>
<b>Alveolar rhabdomyosarcoma (ARMS)</b>	<a href="#">C0206655</a>
<b>Botryoid rhabdomyosarcoma (BRMS)</b>	<a href="#">C1306573</a>
<b>Embryonal rhabdomyosarcoma (ERMS)</b>	<a href="#">C0206656</a>
<b>Pleomorphic rhabdomyosarcoma (PRMS)</b>	<a href="#">C0334480</a>
<b>Rhabdomyosarcoma (RMS), not classifiable</b>	<a href="#">C0035412</a>
<b>Rhabdomyosarcoma (RMS), inadequate tissue for classification</b>	
<b>Rhabdomyosarcoma (RMS), w Mixed Embryonal and Alveolar Features</b>	<a href="#">C1709053</a>
<b>Spindle cell</b>	<a href="#">C0205945</a>



# 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.)
  - $\leq 5 \text{ cm}$
  - $>5 \text{ cm}$
  - Unknown



# Consensus harmonization of primary site

CWS COG EpSSG

Major Primary Site	CWS	COG	EpSSG/MMT Name
ORBIT	Eyelid Orbit	1=Eye 2=Orbit	Eyelid Orbit
HEAD & NECK (non PM)	Scalp	10=Scalp	Soft tissue of scalp External auricular canal Ear soft tissue, external ear Temporal muscle Parotid, soft tissue Gum Base of tongue Lip Lower lip Upper lip Tongue Larynx Oropharynx
	Parotid Oral Cavity	9=Paratoid 7=Oral cavity	Lingual tonsil Mandible soft tissue Bone of face (Maxillar) Masseter Oral cavity Cheek Hypopharynx
	Larynx Oropharynx	5=Larynx 8=Orophaynx	Cheek Hypopharynx Thyroid Neck Neck Supra-clavicular soft tissues Neck, nodes Nos Chin Soft tissue face (non specified region) Face specified region Nasolabial fold (skin) Nostril
	Cheek Hypopharynx Thyroid & Parathyroid Neck	3=Cheek 4=Hypopharynx 11=Thyroid & Parathyroid 6=Neck 12=Other Head & Neck	Thyroid Neck Neck, nodes Nos Chin Soft tissue face (non specified region) Face specified region Nasolabial fold (skin) Nostril

Consensus

Orbit	Eyelid Orbit Other orbit	<a href="#">C0015426</a> <a href="#">C0700042</a> <a href="#">C0700042</a>
Head and neck	Cheek Hypopharynx Larynx Neck Oral cavity Oropharynx Parotid Scalp Thyroid and parathyroid Other face Other head and neck	<a href="#">C0007966</a> <a href="#">C0020629</a> <a href="#">C0023078</a> <a href="#">C0027530</a> <a href="#">C1711367</a> <a href="#">C0521367</a> <a href="#">C3272625</a> <a href="#">C0036270</a> <a href="#">C0574117</a> <a href="#">C0015450</a> <a href="#">C0460004</a>



# 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

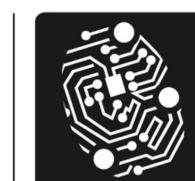


# Developing a paradigm for treatment data

## 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]

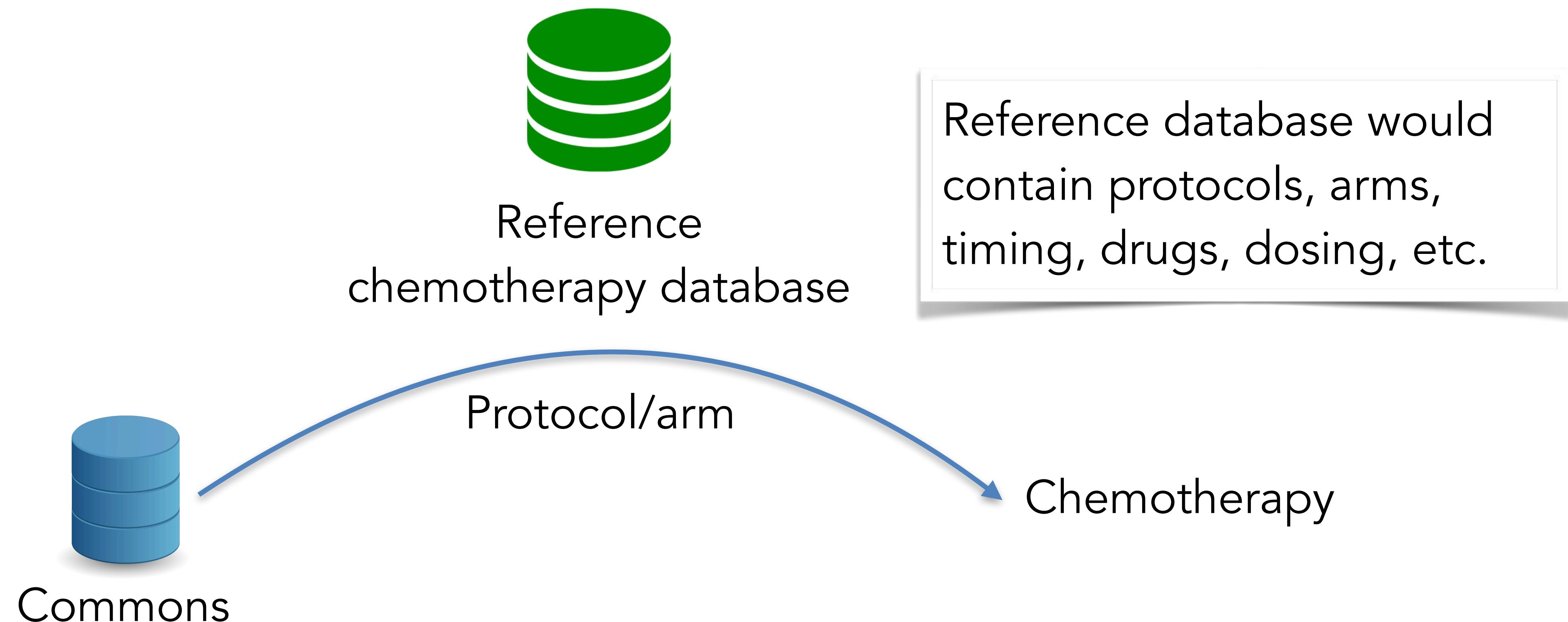
**Cycle # 1**

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 <sup>2</sup> /dose	1		a. Physical Exam (Ht., Wt., BSA, VS), b. CBC, differential, platelets <sup>1</sup> , Serum creatinine <sup>2</sup>  Protocol Section 7.0 for complete list of required observations, including required at study entry ,and Section 15.0 acquisition or shipment guidelines
Etoposide (ETOP)	IV over 1 hour	Pts ≤ 12kg: 4 mg/kg/dose Pts > 12 kg: 120 mg/m <sup>2</sup> /dose	1-3	Give following CARBO on Day 1 only	
Filgrastim (G-CSF)				<b>Carboplatin/Etoposide (CUI C0280617)</b>  <a href="#">Suggest changes to this concept</a> <a href="#">Add to Cart</a>	<b>OTHER STUDIES AS FOR GOOD PATIENT</b>

Does not include dose information or a reference



# 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



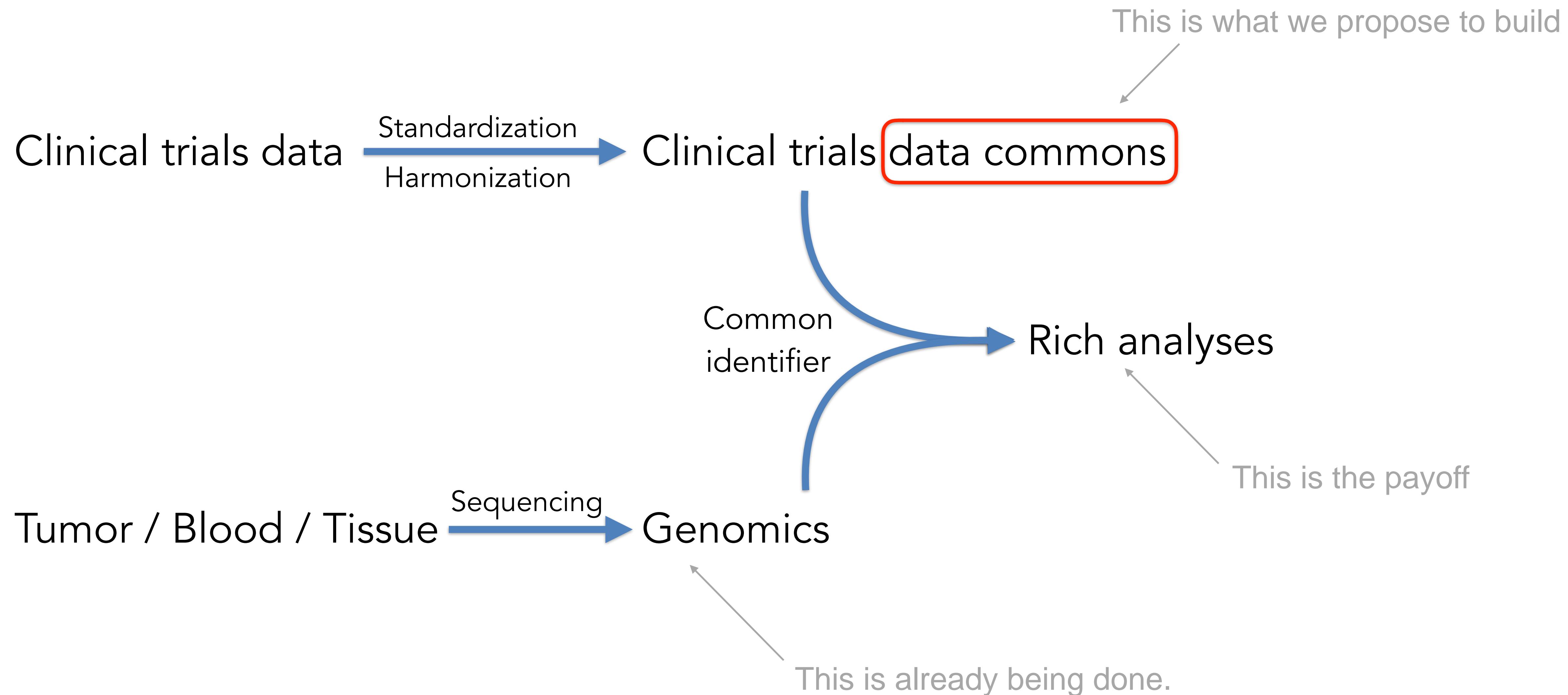
# April 10 COG data standards workshop



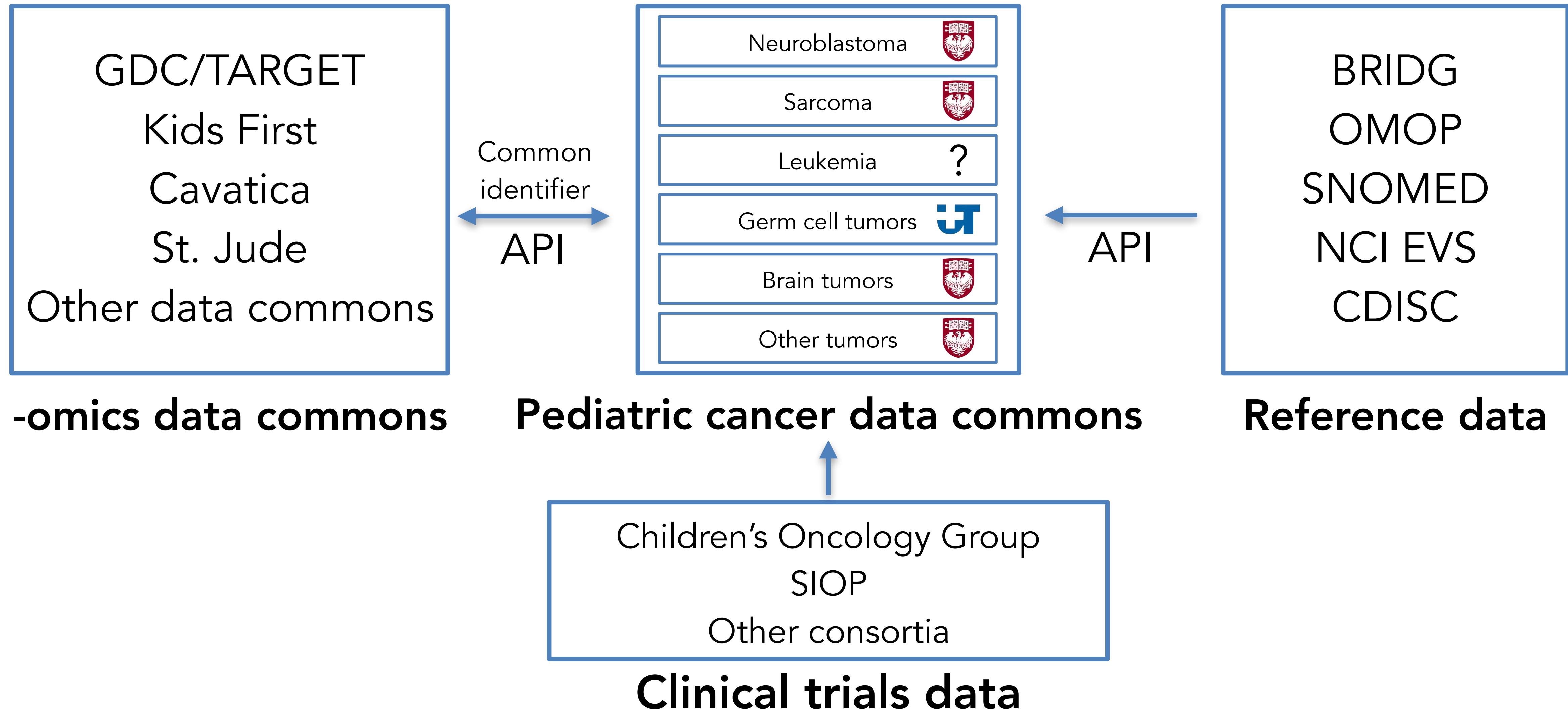
Plan to develop informatics task force for COG  
Make recommendations for informatics strategic plan for COG



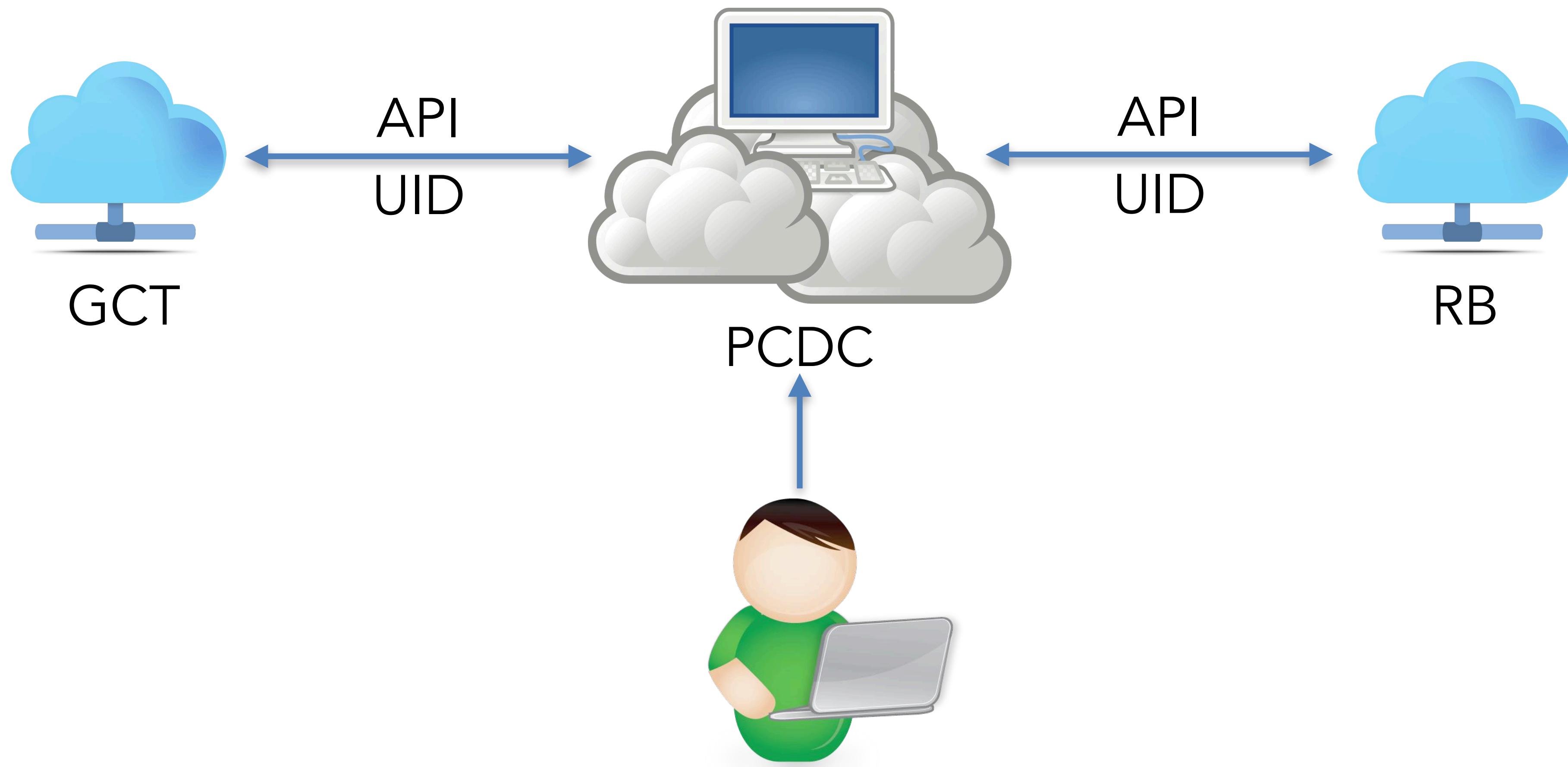
# Goal: Build a commons of clinical trials data to enrich the genomic information



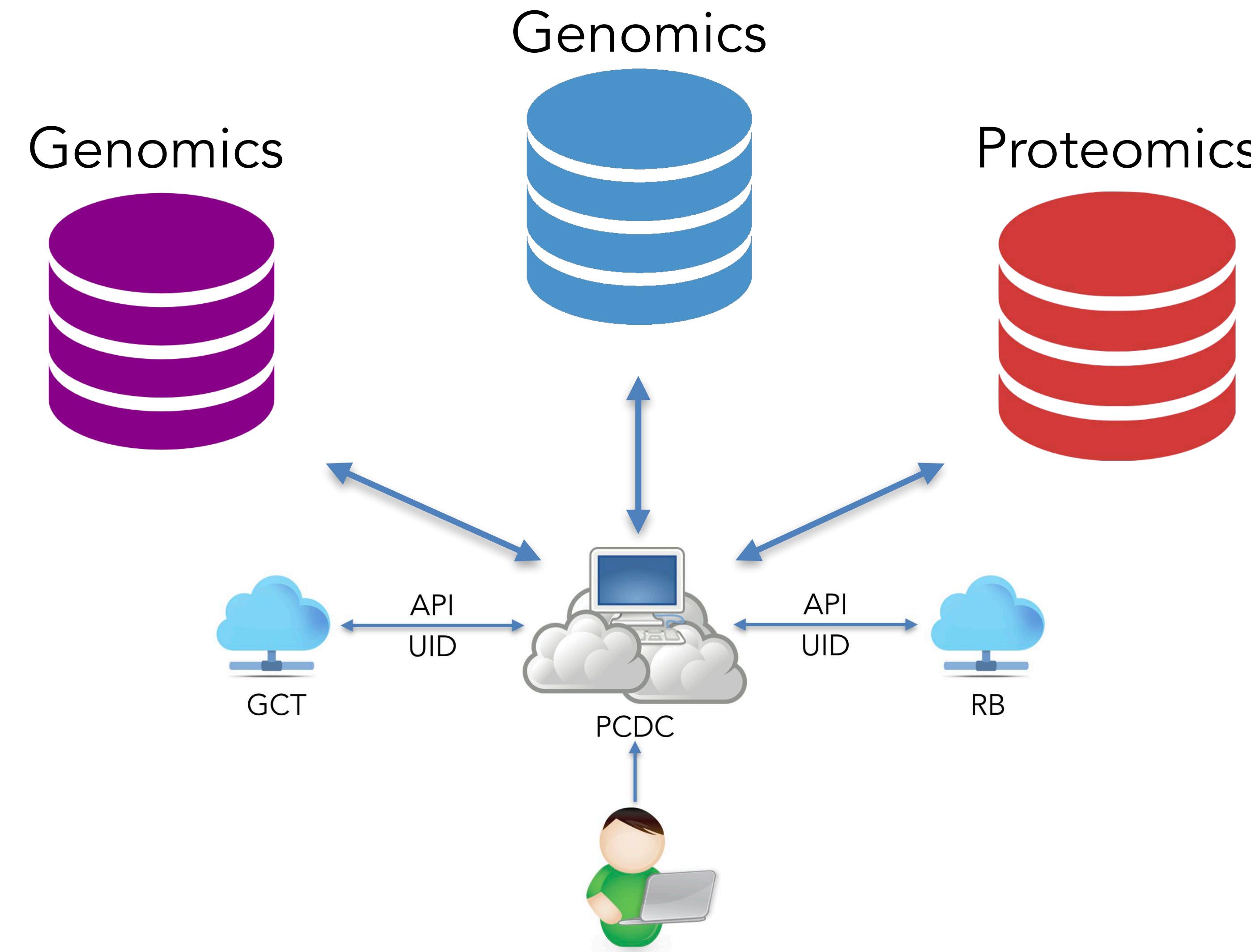
# Paradigm for pediatric cancer clinical trials data commons



# 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





WILLIAM GUY FORBECK  
RESEARCH FOUNDATION  
[wgfrf.org](http://wgfrf.org)

**SB** St. Baldrick's  
FOUNDATION  
*Conquer Childhood Cancers*

Alex's  
Lemonade  
Stand

The Super  
Jake Foundation

LITTLE HEROES

Children's Neuroblastoma  
Cancer Foundation