Pediatric Cancer Research Data Commons

Samuel Volchenboum, MD, PhD

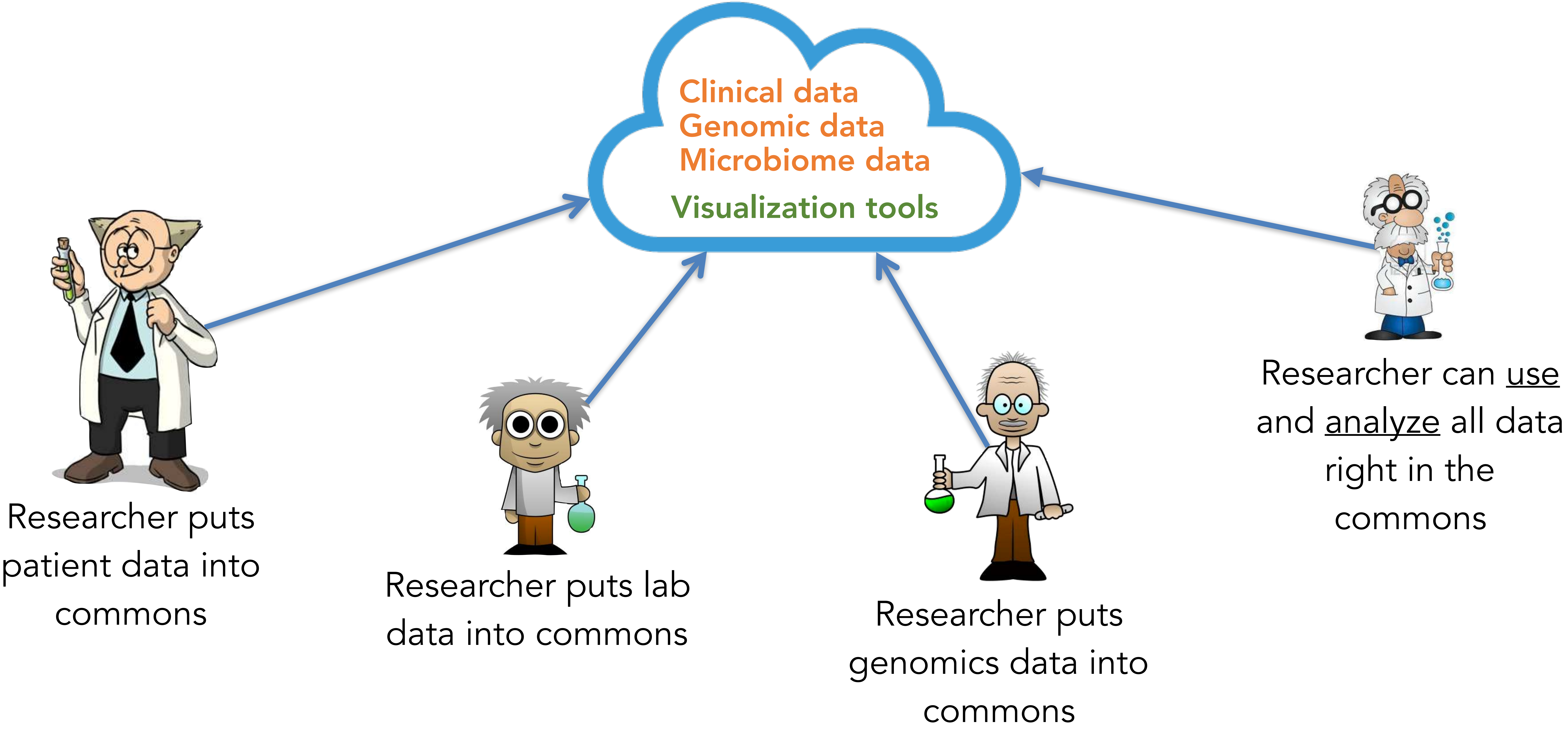
September, 2017

THE UNIVERSITY OF
CHICAGO MEDICINE &
BIOLOGICAL SCIENCES

CENTER FOR
RESEARCH
INFORMATICS

Most research is performed in silos

A commons is a place to store and use data



Clinical data
Genomic data
Microbiome data

Visualization tools

Researcher puts patient data into commons

Researcher puts lab data into commons

Researcher puts genomics data into commons

Researcher can use and analyze all data right in the commons

Why do we need data commons?

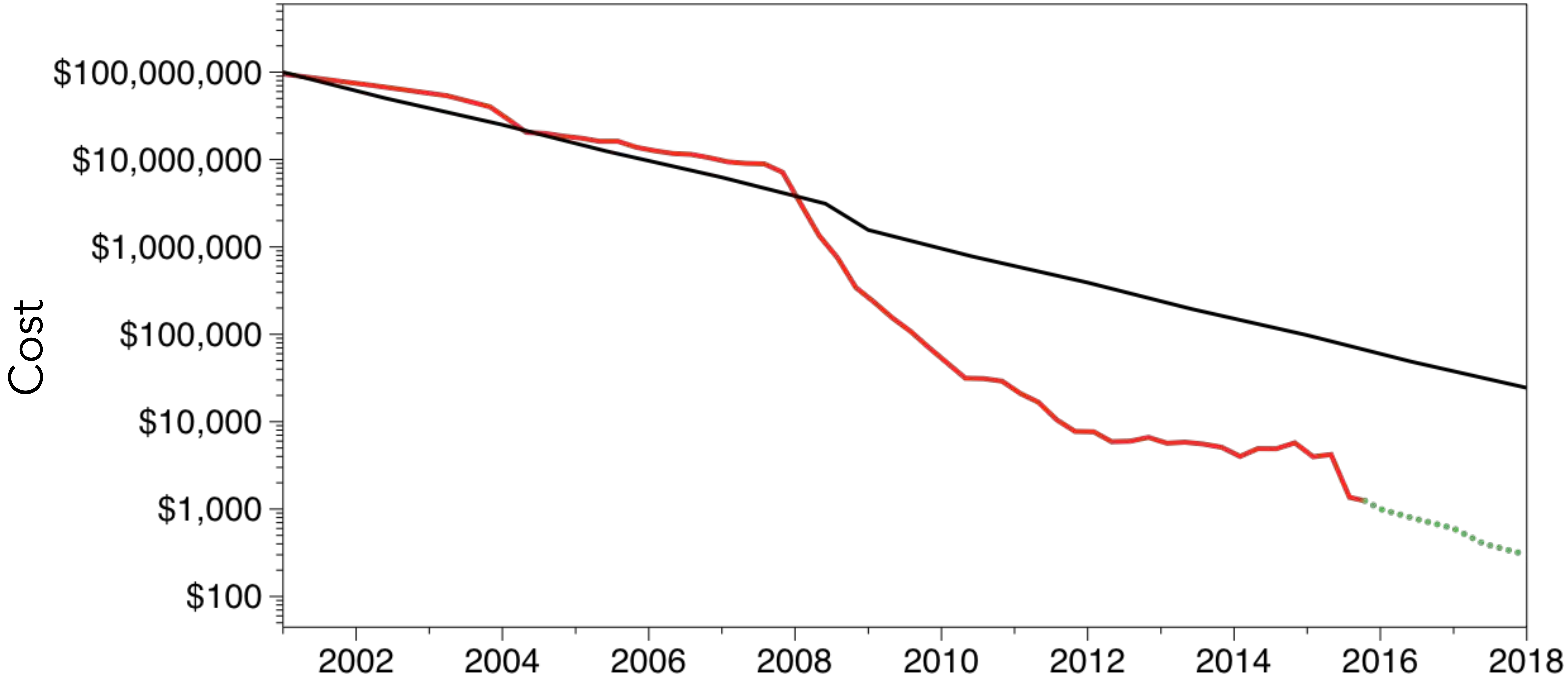
Too much data to store

Takes too long to transfer

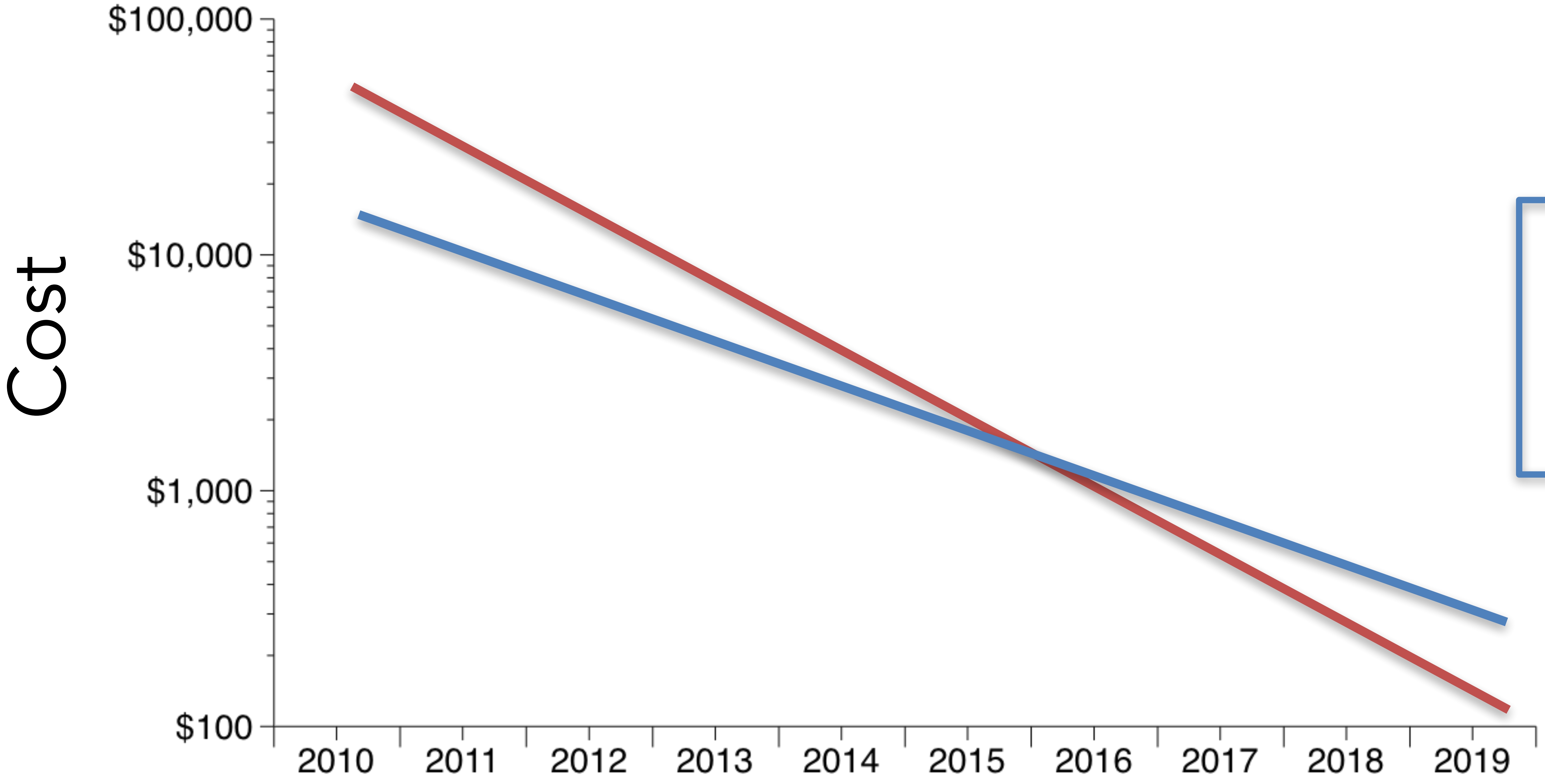
Too expensive to analyze

Lack of data standardization

Falling costs of DNA sequencing



Informatics is now the bottleneck



Sequencing cost dropping ~5x/year

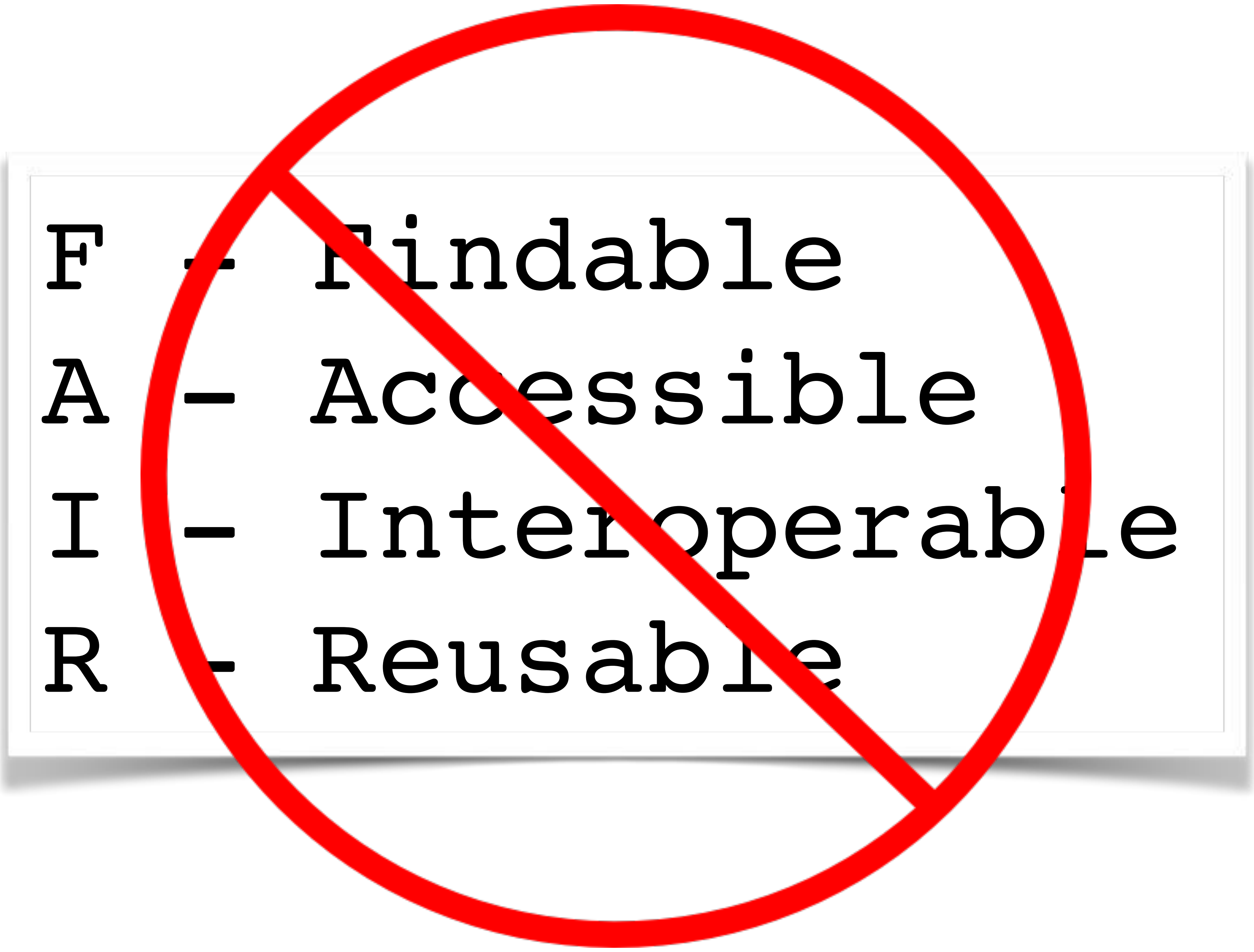
Informatics cost dropping ~2x/year

businessinsider.com

The importance of standards

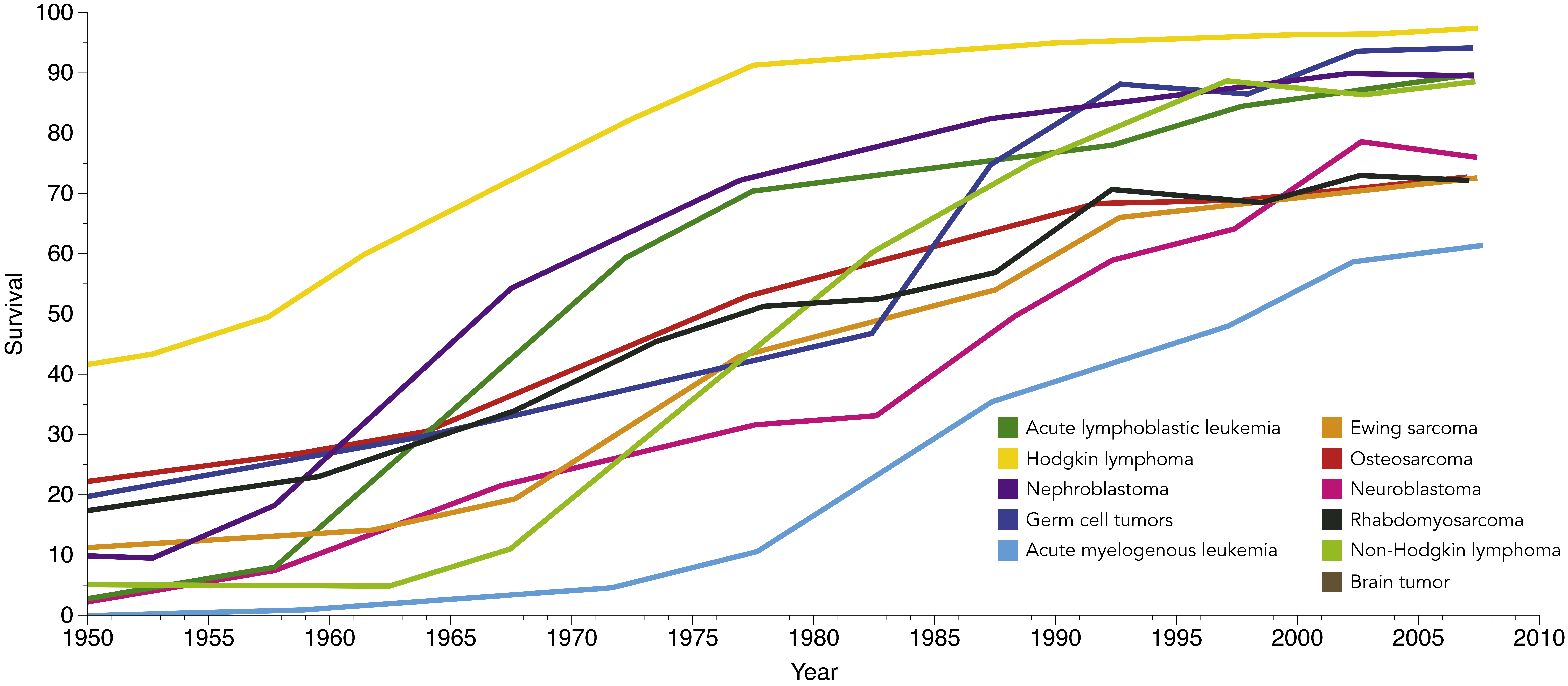
Pediatric rhabdomyosarcoma

	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



F – Findable
A – Accessible
I – Interoperable
R – Reusable

Survival - Pediatric cancer



Why do we need pediatric cancer data commons?

Adult cancers

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

All	**15,190**
Bone	787
Brain	2,653
Hodgkin lymphoma	973
Kidney	608
Acute lymphoblastic leukemia	2,627
Acute myelogenous leukemia	642
Non-Hodgkin lymphoma	1,116
Soft-tissue	869
Other	4,585

Source: CDC - 2014

Studying a rare disease like pediatric cancer is difficult.

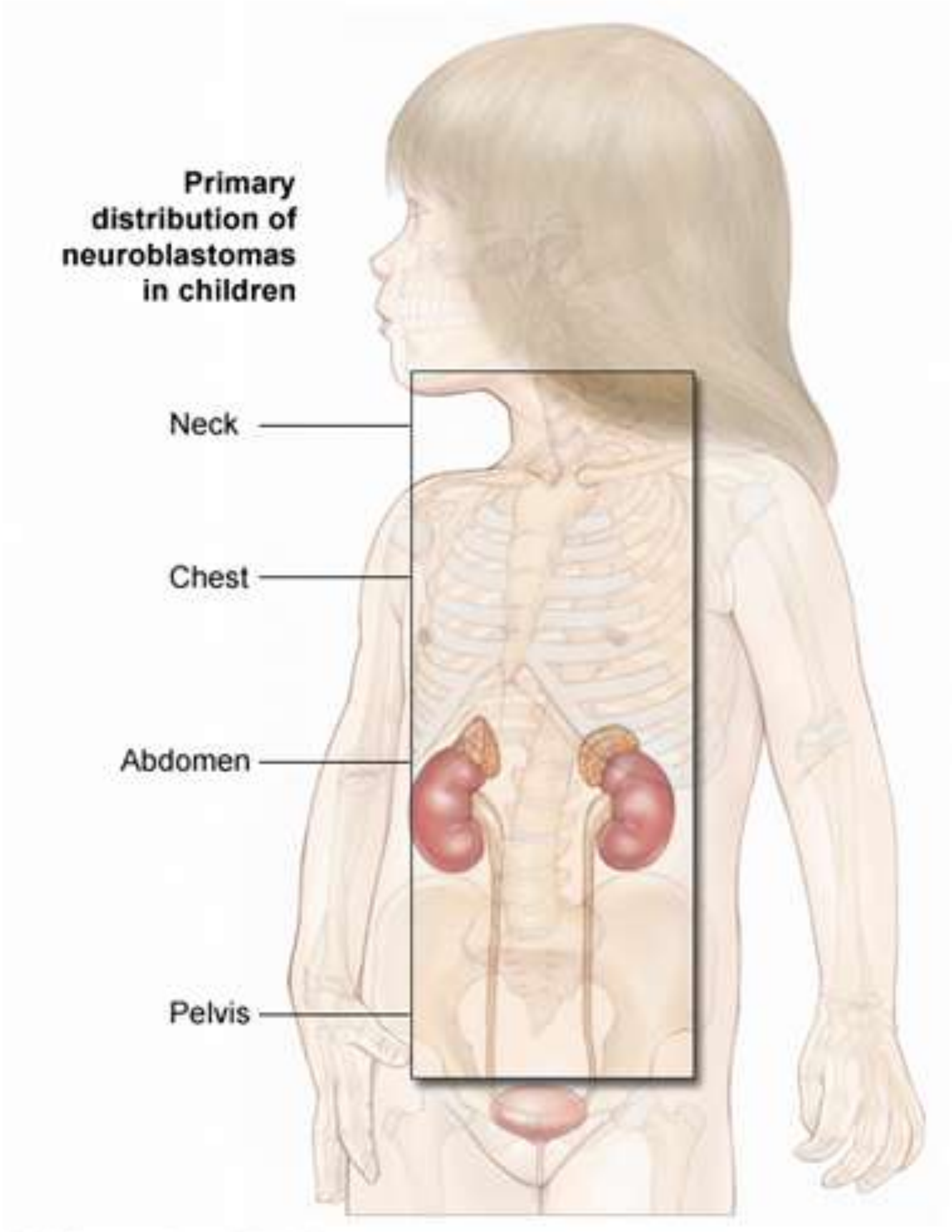
Children’s Oncology Group

1955 Cooperative group system for clinical research

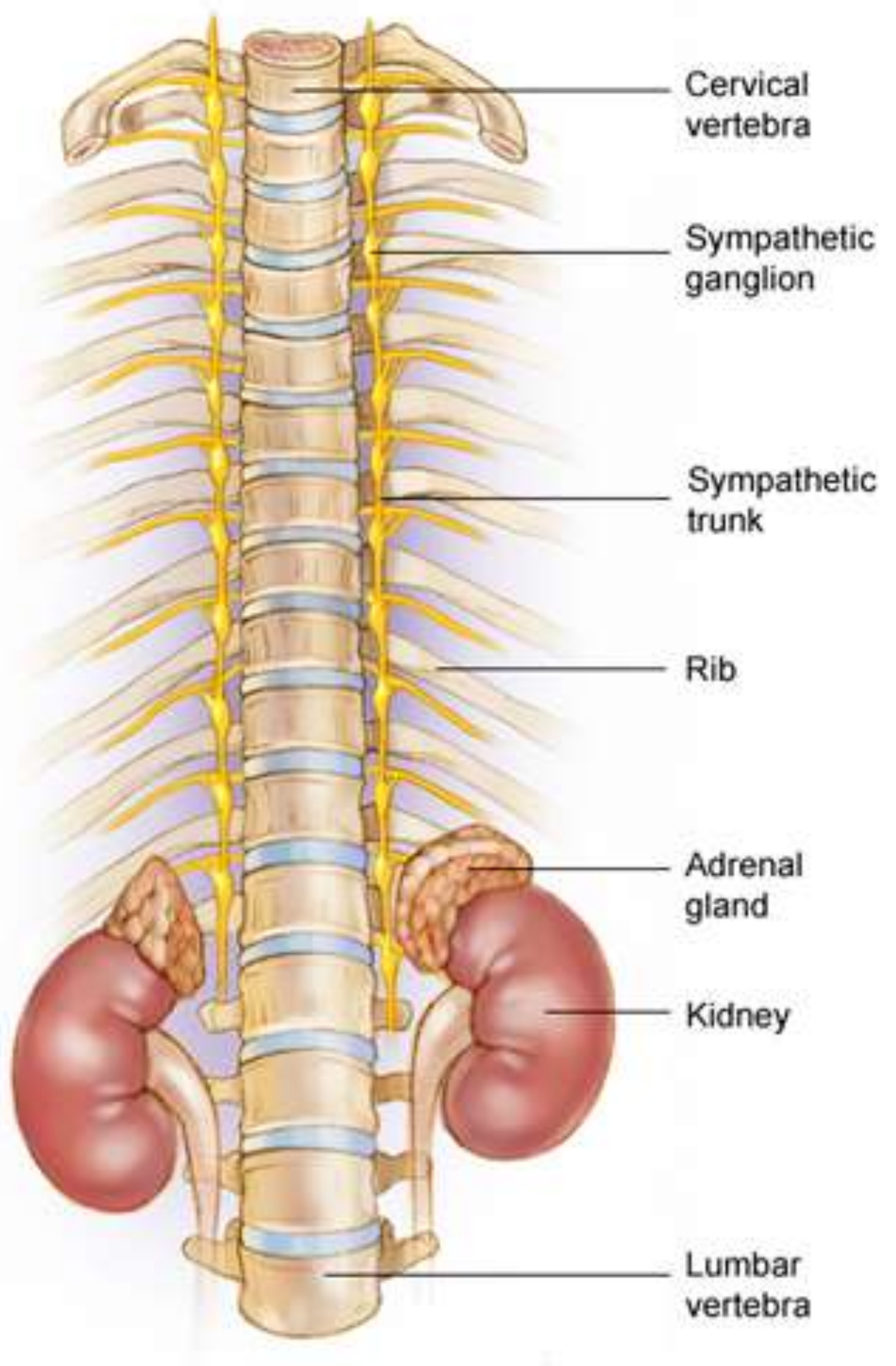
Pediatric Oncology Group (POG)
Children's Cancer Group (CCG)
National Wilms’ Tumor Study Group (NWTS)
Intergroup Rhabdomyosarcoma Study Group (IRSG)

2000 Children’s Oncology Group (COG)

Neuroblastoma



© 2005 American Society of Clinical Oncology



Robert Morreale/Visual Explanations, LLC

The most common solid tumor in children

Neuroblastoma is rare

Site	New Cases 2012
Breast	229,000
Colon	103,000
Prostate	241,000
Lung	226,000
Neuroblastoma	800

Studying such a rare tumor is especially difficult and requires collaborative, consortium-driven research.

International Neuroblastoma Research Group

Established in 2004

Children's Oncology Group (COG)
4235

Germany
1938

Japan
470

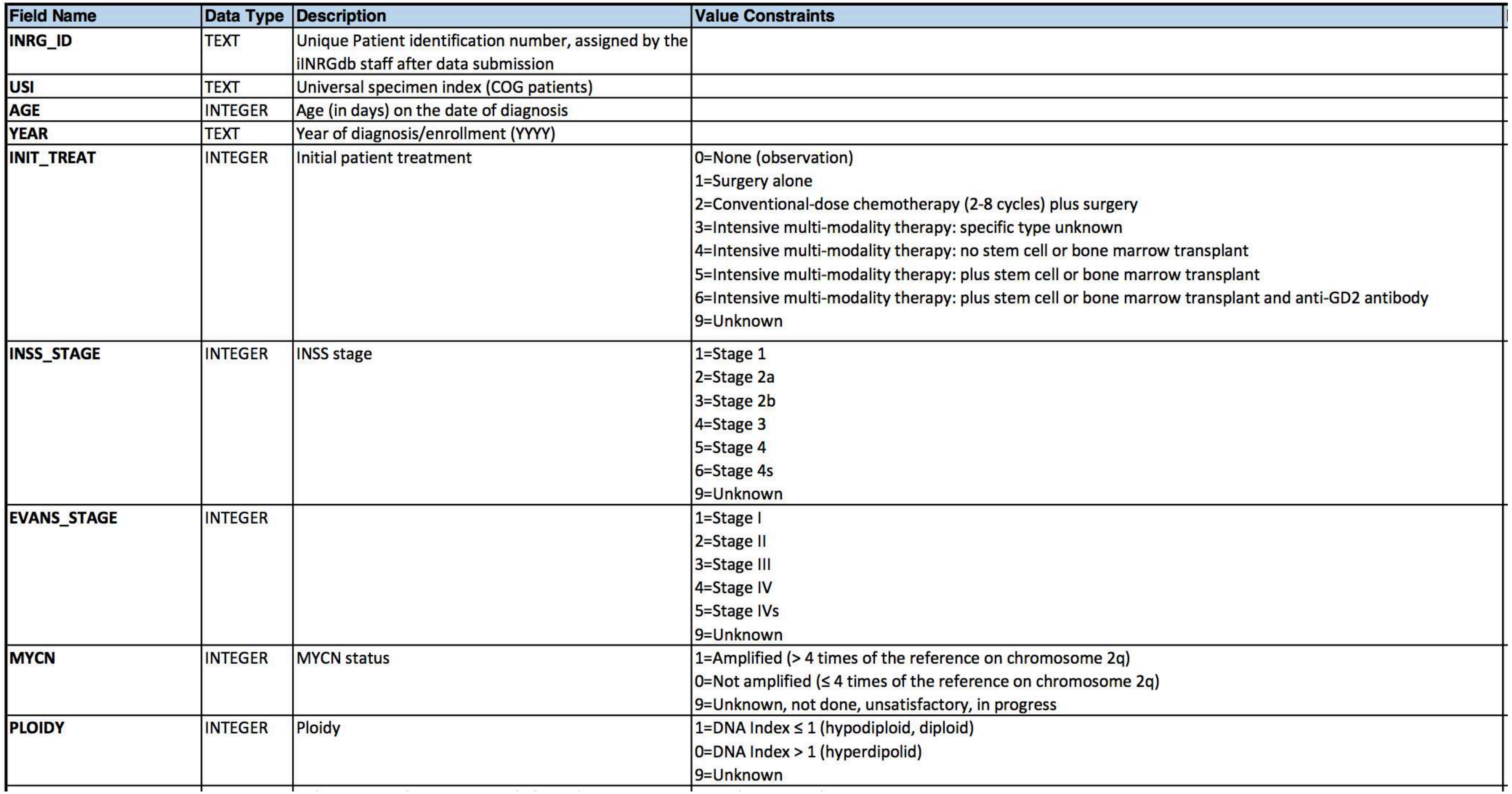
SIOPEN
936

Charged with collecting patients and standardizing terminologies

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

The largest collection of neuroblastoma patients and has resulted in more than a dozen high-impact papers

Neuroblastoma consensus data dictionary



Field Name	Data Type	Description	Value Constraints
INRG_ID	TEXT	Unique Patient identification number, assigned by the iINRGdb staff after data submission	
USI	TEXT	Universal specimen index (COG patients)	
AGE	INTEGER	Age (in days) on the date of diagnosis	
YEAR	TEXT	Year of diagnosis/enrollment (YYYY)	
INIT_TREAT	INTEGER	Initial patient treatment	0=None (observation) 1=Surgery alone 2=Conventional-dose chemotherapy (2-8 cycles) plus surgery 3=Intensive multi-modality therapy: specific type unknown 4=Intensive multi-modality therapy: no stem cell or bone marrow transplant 5=Intensive multi-modality therapy: plus stem cell or bone marrow transplant 6=Intensive multi-modality therapy: plus stem cell or bone marrow transplant and anti-GD2 antibody 9=Unknown
INSS_STAGE	INTEGER	INSS stage	1=Stage 1 2=Stage 2a 3=Stage 2b 4=Stage 3 5=Stage 4 6=Stage 4s 9=Unknown
EVANS_STAGE	INTEGER		1=Stage I 2=Stage II 3=Stage III 4=Stage IV 5=Stage IVs 9=Unknown
MYCN	INTEGER	MYCN status	1=Amplified (> 4 times of the reference on chromosome 2q) 0=Not amplified (≤ 4 times of the reference on chromosome 2q) 9=Unknown, not done, unsatisfactory, in progress
PLOIDY	INTEGER	Ploidy	1=DNA Index ≤ 1 (hypodiploid, diploid) 0=DNA Index > 1 (hyperdipolid) 9=Unknown

International Neuroblastoma Research Group

Established in 2004

Children's Oncology Group (COG)
4500

Germany
1900

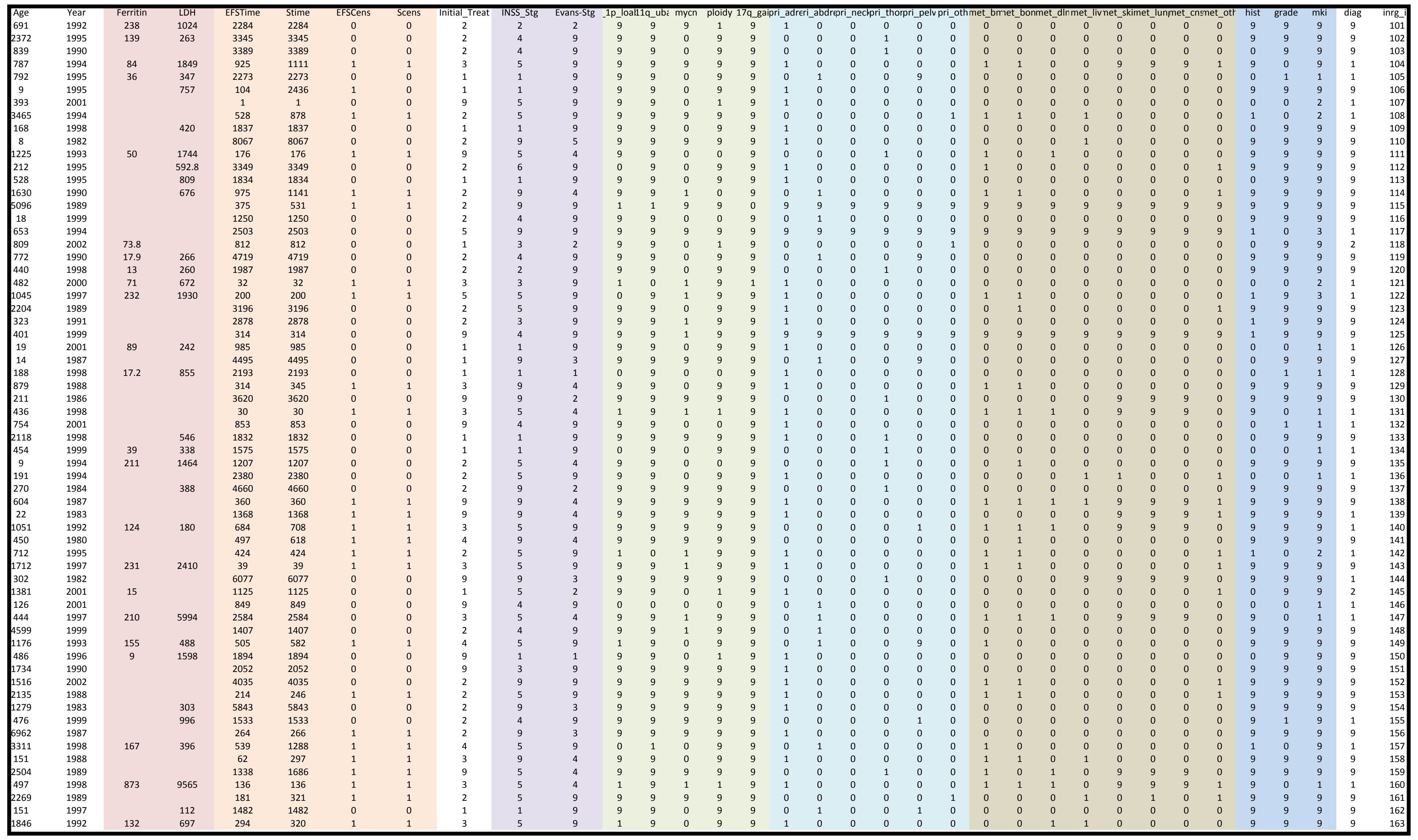
Japan
500

SIOPEN
900

The good news

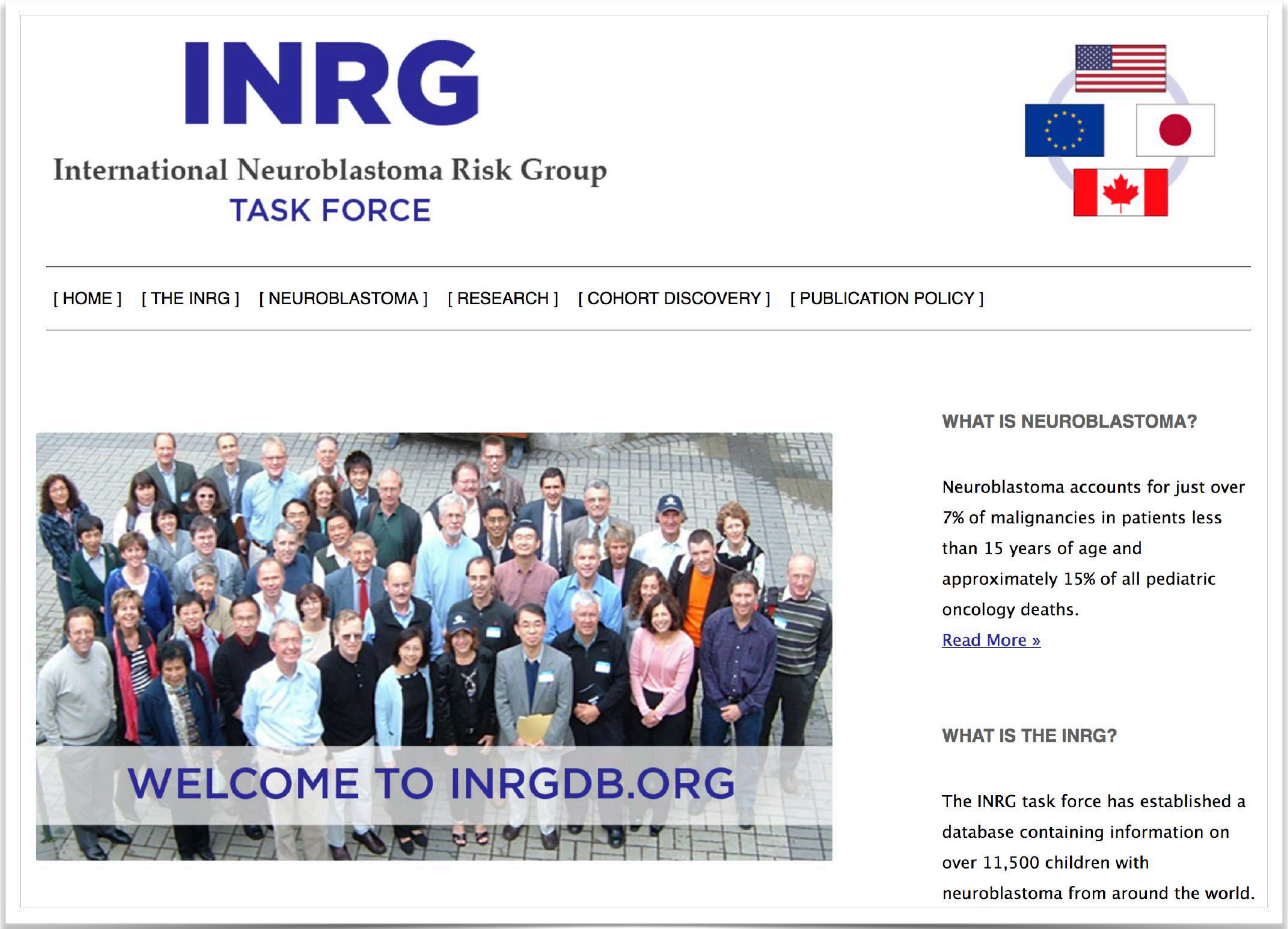
8800 patients!

The bad news?



Age	Year	Ferritin	LDH	EFSTime	Stime	EFSCens	Scens	Initial_Treat	INSS_Stg	Evans-Stg	_1p_loal	11q_uba	mycn	ploidy	17q_gai	pri_adre	pri_abdr	pri_neck	pri_thor	pri_pelv	pri_oth	met_bm	met_bon	met_dln	met_liv	met_ski	met_lun	met_cns	met_oth	hist	grade	mki	diag	inrg_i
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	0	9	9	9	9	101
2372	1995	139	263	3345	3345	0	0	2	4	9	9	9	0	9	9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	9	9	9	9	102
839	1990			3389	3389	0	0	2	4	9	9	9	0	9	9	0	0	0	1	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	9	1	0	0	0	0	0	1	1	0	0	9	9	9	1	9	0	9	1	104
792	1995	36	347	2273	2273	0	0	1	1	9	9	9	0	9	9	0	1	0	0	9	0	0	0	0	0	0	0	0	0	0	1	1	1	105
9	1995		757	104	2436	1	0	1	1	9	9	9	0	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	9	106
393	2001			1	1	0	0	9	5	9	9	9	0	1	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	107
3465	1994			528	878	1	1	2	5	9	9	9	9	9	9	0	0	0	0	0	1	1	1	0	1	0	0	0	0	1	0	2	1	108
168	1998		420	1837	1837	0	0	1	1	9	9	9	0	9	9	1	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	9	1	0	0	0	0	0	0	0	0	1	0	0	0	0	9	9	9	9	110
1225	1993	50	1744	176	176	1	1	9	5	4	9	9	0	0	9	0	0	0	1	0	0	1	0	1	0	0	0	0	0	9	9	9	9	111
212	1995		592.8	3349	3349	0	0	2	6	9	0	9	0	9	9	1	0	0	0	0	0	1	0	0	0	0	0	0	1	9	9	9	9	112
528	1995		809	1834	1834	0	0	1	1	9	9	9	0	9	9	1	0	0	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	1	0	9	0	1	0	0	0	0	1	1	0	0	0	0	0	1	9	9	9	9	114
5096	1989			375	531	1	1	2	9	9	1	1	9	9	0	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	115
18	1999			1250	1250	0	0	2	4	9	9	9	0	0	9	0	1	0	0	0	0	0	0	0	0	0	0	0	0	9	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	9	9	1	0	3	1	117
809	2002	73.8		812	812	0	0	1	3	2	9	9	0	1	9	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	9	9	2	118
772	1990	17.9	266	4719	4719	0	0	2	4	9	9	9	0	9	9	0	1	0	0	9	0	0	0	0	0	0	0	0	0	9	9	9	9	119
440	1998	13	260	1987	1987	0	0	2	2	9	9	9	0	9	9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	9	9	9	9	120
482	2000	71	672	32	32	1	1	3	3	9	1	0	1	9	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	121
1045	1997	232	1930	200	200	1	1	5	5	9	0	9	1	9	9	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	9	3	1	122
2204	1989			3196	3196	0	0	2	5	9	9	9	0	9	9	1	0	0	0	0	0	0	1	0	0	0	0	0	1	9	9	9	9	123
323	1991			2878	2878	0	0	2	3	9	9	9	1	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	9	9	9	124
401	1999			314	314	0	0	9	4	9	9	9	1	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	0	9	9	1	0	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	1	0	0	9	0	0	0	0	0	0	0	0	0	0	9	9	9	127
188	1998	17.2	855	2193	2193	0	0	1	1	1	0	9	0	0	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	128
879	1988			314	345	1	1	3	9	4	9	9	0	9	9	1	0	0	0	0	0	1	1	0	0	0	0	0	0	9	9	9	9	129
211	1986			3620	3620	0	0	9	9	2	9	9	9	9	9	0	0	0	1	0	0	0	0	0	0	9	9	9	0	9	9	9	9	130
436	1998			30	30	1	1	3	5	4	1	9	1	1	9	1	0	0	0	0	0	1	1	1	0	9	9	9	0	9	0	1	1	131
754	2001			853	853	0	0	9	4	9	9	9	0	0	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	132
2118	1998		546	1832	1832	0	0	1	1	9	9	9	9	9	9	1	0	0	1	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	0	9	9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	134
9	1994	211	1464	1207	1207	0	0	2	5	4	9	9	0	0	9	0	0	0	1	0	0	0	1	0	0	0	0	0	0	9	9	9	9	135
191	1994			2380	2380	0	0	2	5	9	9	9	0	9	9	1	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	1	1	136
270	1984		388	4660	4660	0	0	2	9	2	9	9	9	9	9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	9	9	9	9	137
604	1987			360	360	1	1	9	9	4	9	9	9	9	9	1	0	0	0	0	0	1	1	1	1	9	9	9	1	9	9	9	9	138
22	1983			1368	1368	1	1	9	9	4	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	9	9	9	1	9	9	9	1	139
1051	1992	124	180	684	708	1	1	3	5	9	9	9	9	9	9	0	0	0	0	1	0	1	1	1	0	9	9	9	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	1	0	0	0	0	0	0	9	9	9	9	141
712	1995			424	424	1	1	2	5	9	1	0	1	9	9	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	2	1	142
1712	1997	231	2410	39	39	1	1	3	5	9	9	9	1	9	9	1	0	0	0	0	0	1	1	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	1	0	0	0	0	0	9	9	9	9	0	9	9	9	1	144
1381	2001	15		1125	1125	0	0	1	5	2	9	9	0	1	9	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	9	9	2	145
126	2001			849	849	0	0	9	4	9	0	0	0	0	9	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	146
444	1997	210	5994	2584	2584	0	0	3	5	4	9	9	1	9	9	0	1	0	0	0	0	1	1	1	0	9	9	9	0	9	0	1	1	147
4599	1999			1407	1407	0	0	2	4	9	9	9	1	9	9	0	1	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	9	148
1176	1993	155	488	505	582	1	1	4	5	9	1	9	0	9	9	0	1	0	0	9	0	1	0	0	0	0	0	0	0	9	9	9	9	149
486	1996	9	1598	1894	1894	0	0	9	1	1	9	9	0	1	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	150
1734	1990			2052	2052	0	0	9	3	9	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	9	151
1516	2002			4035	4035	0	0	2	9	9	9	9	9	9	9	1	0	0	0	0	0	1	1	0	0	0	0	0	1	9	9	9	9	152
2135	1988			214	246	1	1	2	5	9	9	9	9	9	9	1	0	0	0	0	0	1	1	0	0	0	0	0	1	9	9	9	9	153
1279	1983		303	5843	5843	0	0	2	9	3	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	9	154
476	1999		996	1533	1533	0	0	2	4	9	9	9	9	9	9	0	0	0	0	1	0	0	0	0	0	0	0	0	0	9	1	9	1	155
6962	1987			264	266	1	1	2	9	3	9	9	9	9	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	9	156
3311	1998	167	396	539	1288	1	1	4	5	9	0	1	0	9	9	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	9	1	157
151	1988			62	297	1	1	3	9	4	9	9	0	9	9	1	0	0	0	0	0	1	1	0	1	0	0	0	0	9	9	9	9	158
2504	1989			1338	1686	1	1	9	5	4	9	9	9	9	9	0	0	0	1	0	0	1	0	1	0	9	9	9	0	9	9	9	9	159
497	1998	873	9565	136	136	1	1	3	5	4	1	9	1	1	9	1	0	0	0	0	0	1	1	1	0	9	9	9	1	9	0	1	1	160
2269	1989			181	321	1	1	2	5	9	9	9	9	9	9	0	0	0	0	0	1	0	0	0	1	0	1	0	1	9	9	9	9	161
151	1997		112	1482	1482	0	0	1	1	9	9	9	0	9	9	0	1	0	0	1	0	0	0	0	0	0	0	0	0	9	9	9	9	162
1846	1992	132	697	294	320	1	1	3	5	9	1	9	0	9	9	1	0	0	0	0	0	0	0	1	1	0	0	0	0	9	9	9	9	163

Neuroblastoma Commons Cohort Discovery



INRG

International Neuroblastoma Risk Group

TASK FORCE

[HOME] [THE INRG] [NEUROBLASTOMA] [RESEARCH] [COHORT DISCOVERY] [PUBLICATION POLICY]

WELCOME TO INRGDB.ORG

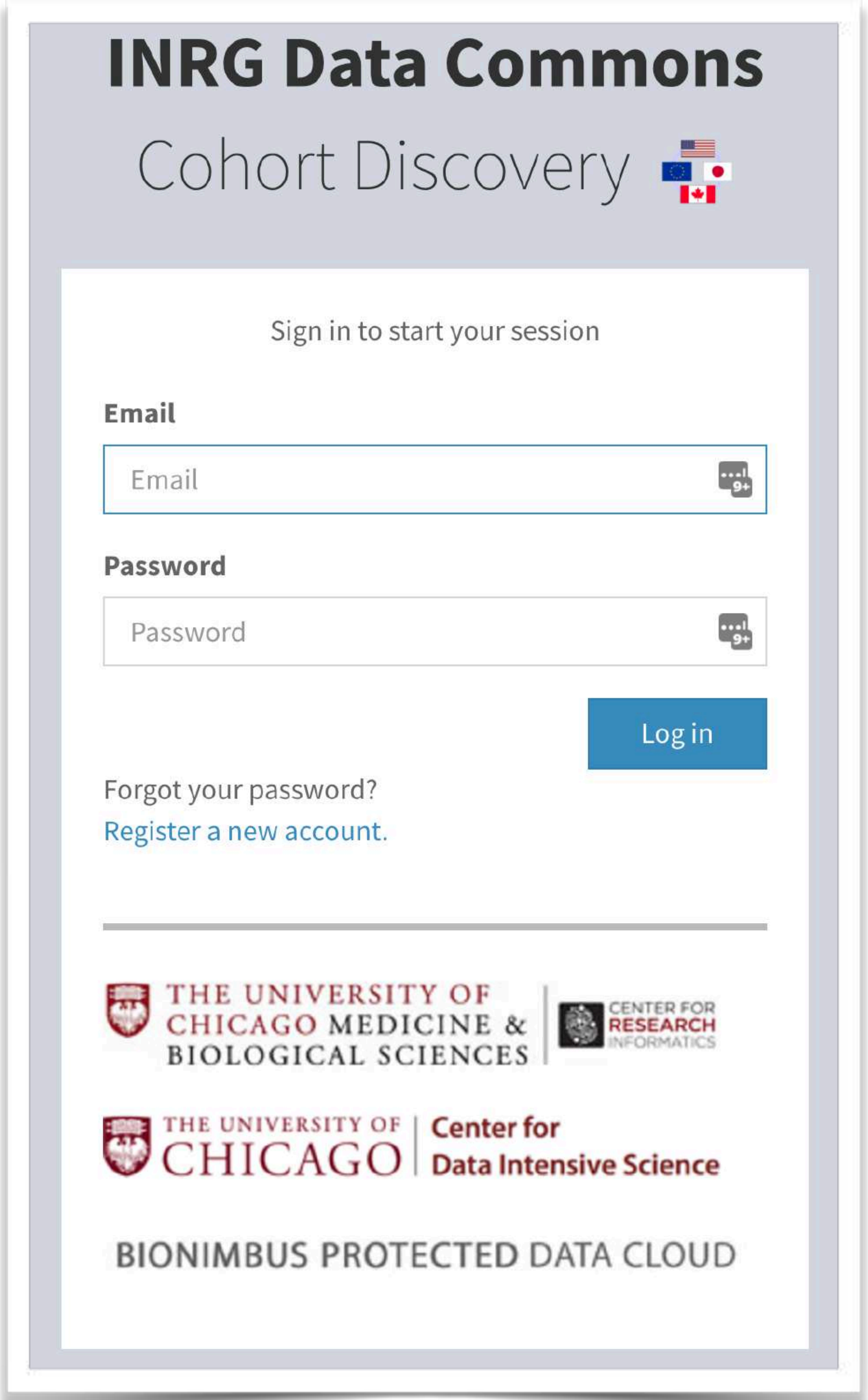
WHAT IS NEUROBLASTOMA?

Neuroblastoma accounts for just over 7% of malignancies in patients less than 15 years of age and approximately 15% of all pediatric oncology deaths.

Read More »

WHAT IS THE INRG?

The INRG task force has established a database containing information on over 11,500 children with neuroblastoma from around the world.



INRG Data Commons

Cohort Discovery

Sign in to start your session

Email

Email

Password

Password

Log in

Forgot your password?

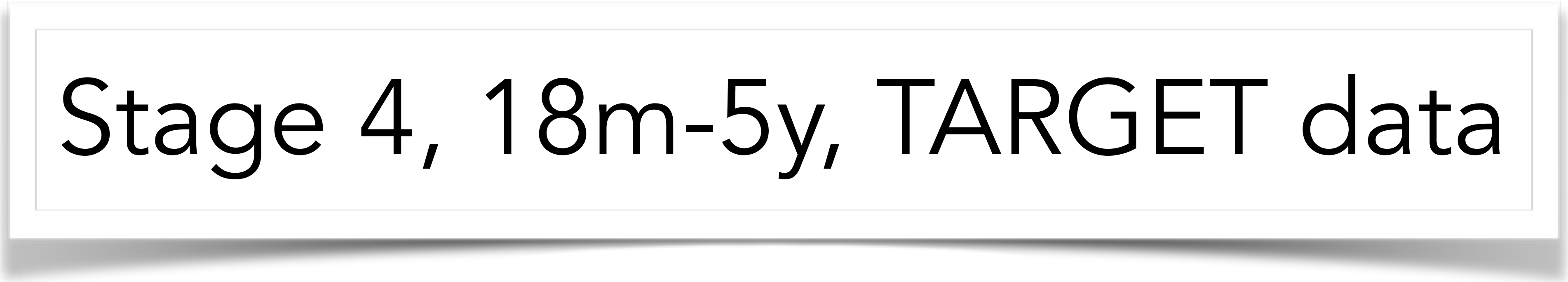
Register a new account.

THE UNIVERSITY OF CHICAGO MEDICINE & BIOLOGICAL SCIENCES | CENTER FOR RESEARCH INFORMATICS

THE UNIVERSITY OF CHICAGO | Center for Data Intensive Science

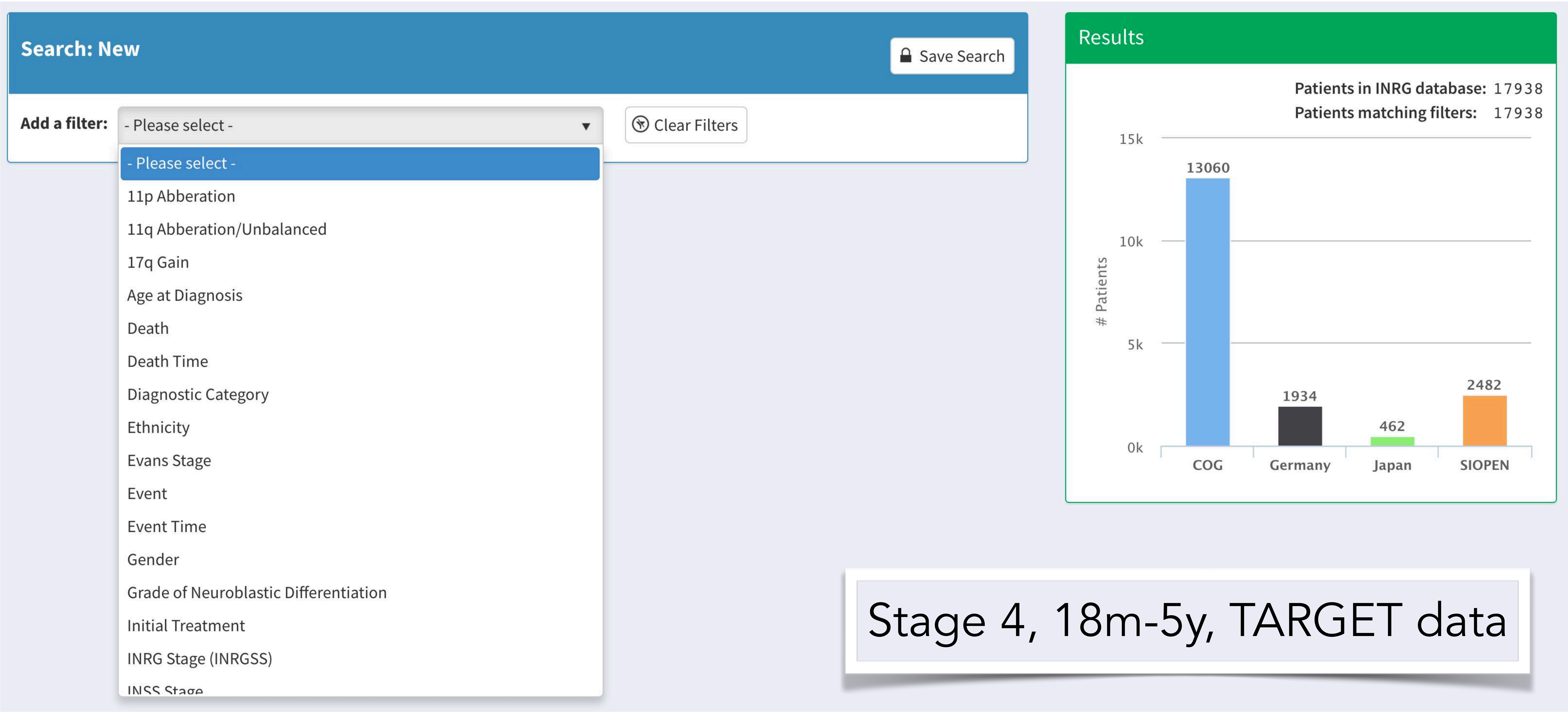
BIONIMBUS PROTECTED DATA CLOUD

Demo query 1



Stage 4, 18m-5y, TARGET data

Neuroblastoma Commons Cohort Discovery



Search: New

Add a filter: - Please select -

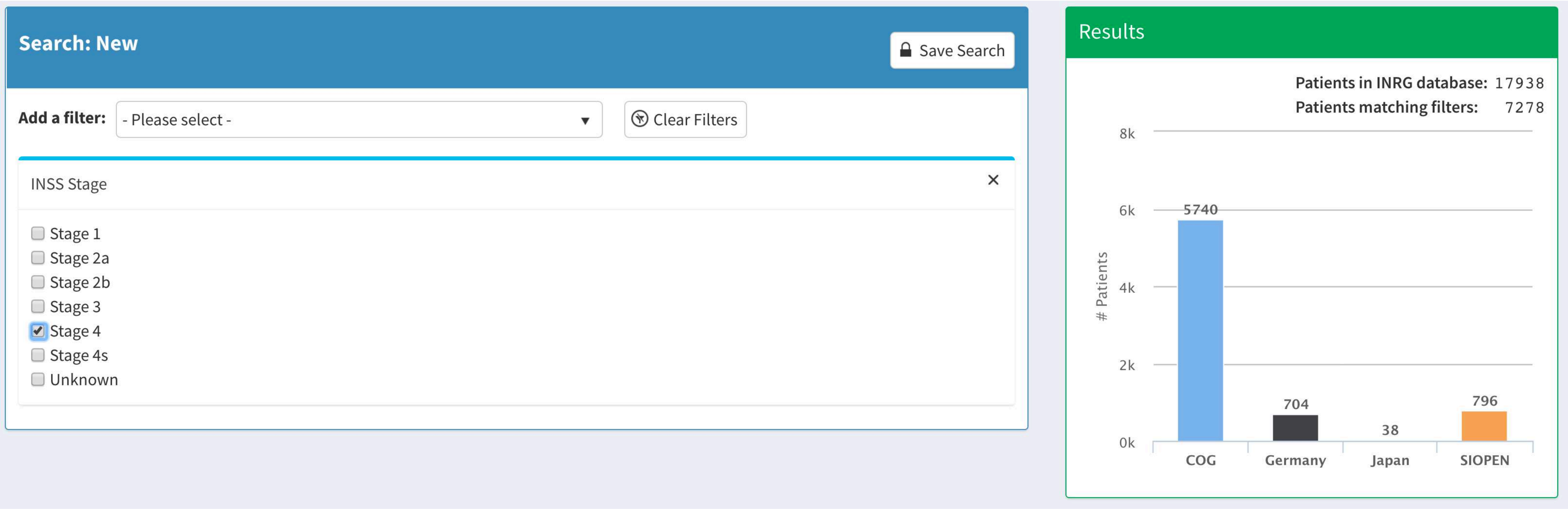
Results

Patients in INRG database: 17938

Patients matching filters: 17938

Stage 4, 18m-5y, TARGET data

Neuroblastoma Commons Cohort Discovery



Search: New

🔒 Save Search

Add a filter: - Please select - ▼

Clear Filters

INSS Stage ✕

- ☐ Stage 1
- ☐ Stage 2a
- ☐ Stage 2b
- ☐ Stage 3
- ☑ Stage 4
- ☐ Stage 4s
- ☐ Unknown

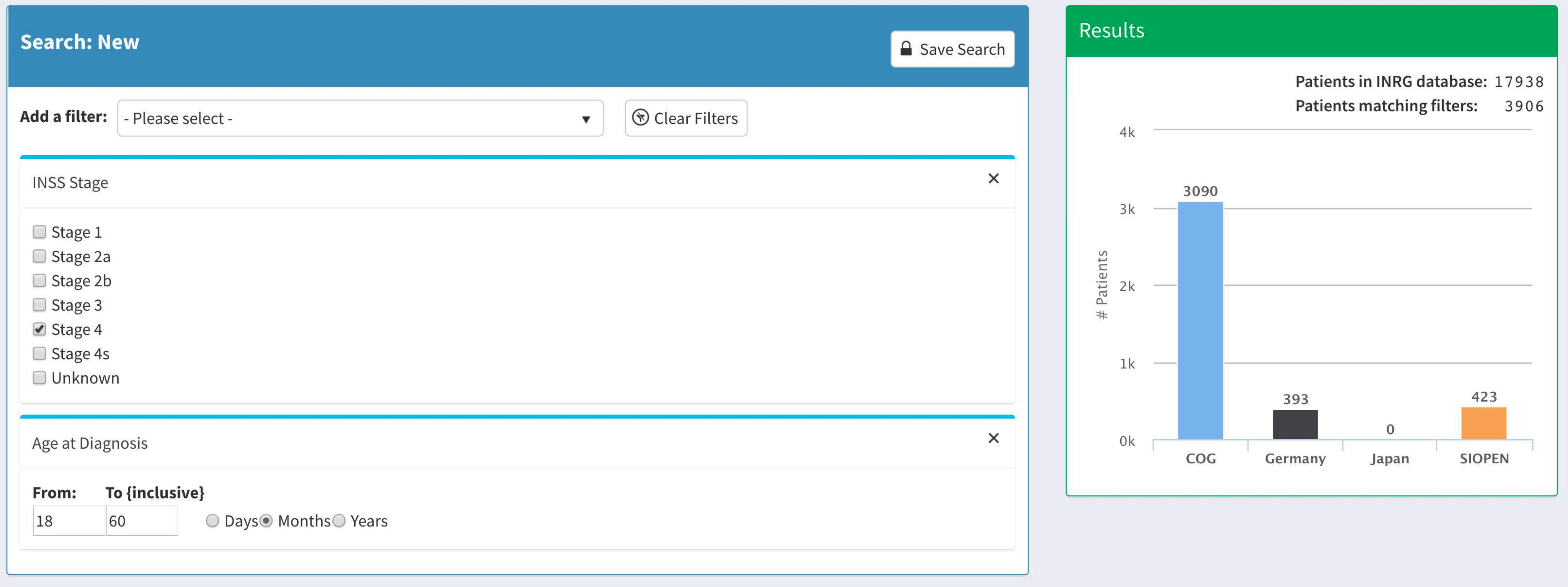
Results

Patients in INRG database: 17938

Patients matching filters: 7278

Stage 4, 18m-5y, TARGET data

Neuroblastoma Commons Cohort Discovery



Search: New

Save Search

Add a filter: - Please select - ▾

Clear Filters

INSS Stage ×

- □ Stage 1
- □ Stage 2a
- □ Stage 2b
- □ Stage 3
- ☑ Stage 4
- □ Stage 4s
- □ Unknown

Age at Diagnosis ×

From: 18 **To {inclusive}** 60

○ Days ● Months ○ Years

Results

Patients in INRG database: 17938

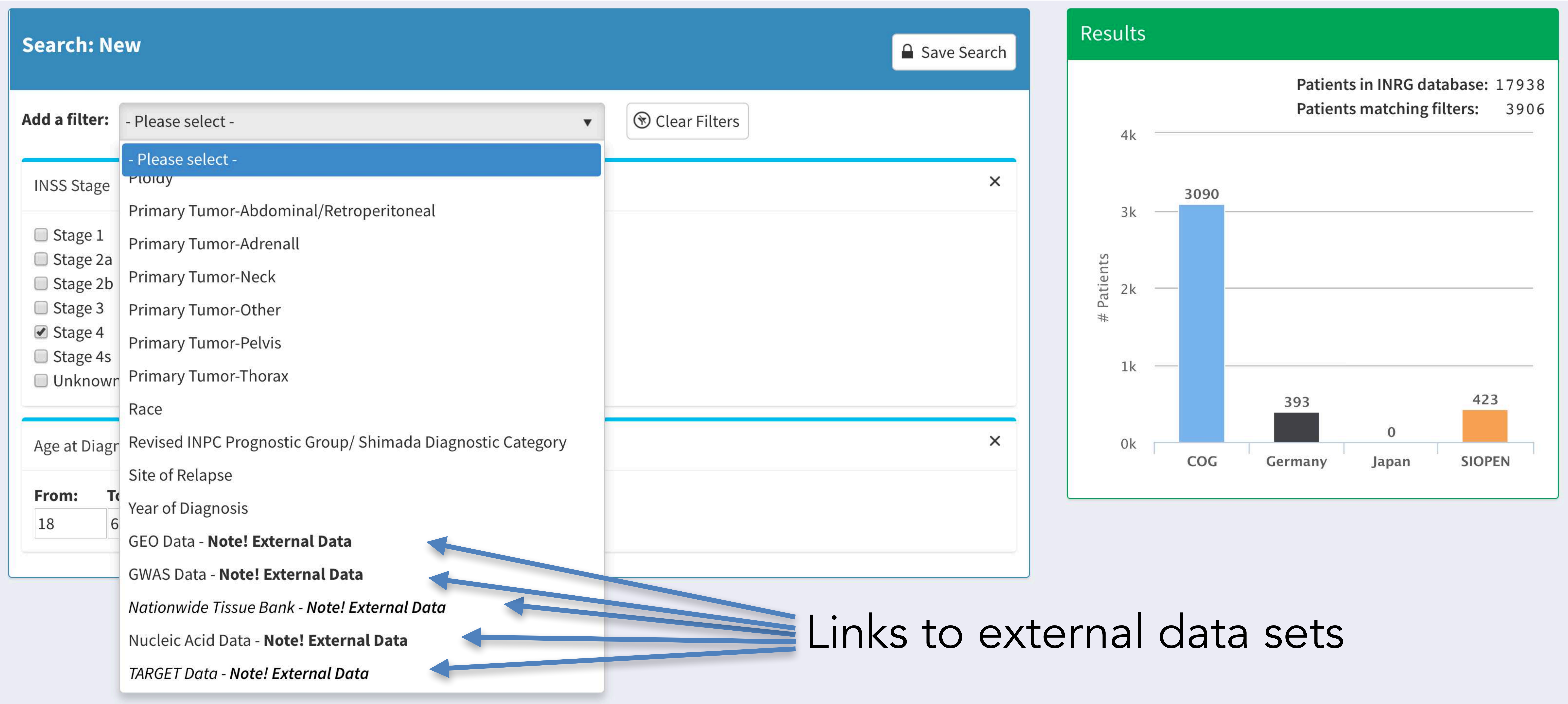
Patients matching filters: 3906

Stage 4, 18m-5y, TARGET data

THE UNIVERSITY OF CHICAGO MEDICINE & BIOLOGICAL SCIENCES

CENTER FOR RESEARCH INFORMATICS

Neuroblastoma Commons Cohort Discovery



Search: New

🔒 Save Search

Add a filter: - Please select - ▾

Clear Filters

- - Please select -
- Ploidy
- Primary Tumor-Abdominal/Retroperitoneal
- Primary Tumor-Adrenall
- Primary Tumor-Neck
- Primary Tumor-Other
- Primary Tumor-Pelvis
- Primary Tumor-Thorax
- Race
- Revised INPC Prognostic Group/ Shimada Diagnostic Category
- Site of Relapse
- 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***

INSS Stage ✕

- □ Stage 1
- □ Stage 2a
- □ Stage 2b
- □ Stage 3
- ☑ Stage 4
- □ Stage 4s
- □ Unknown

Age at Diagn ✕

From: 18 **To** 6

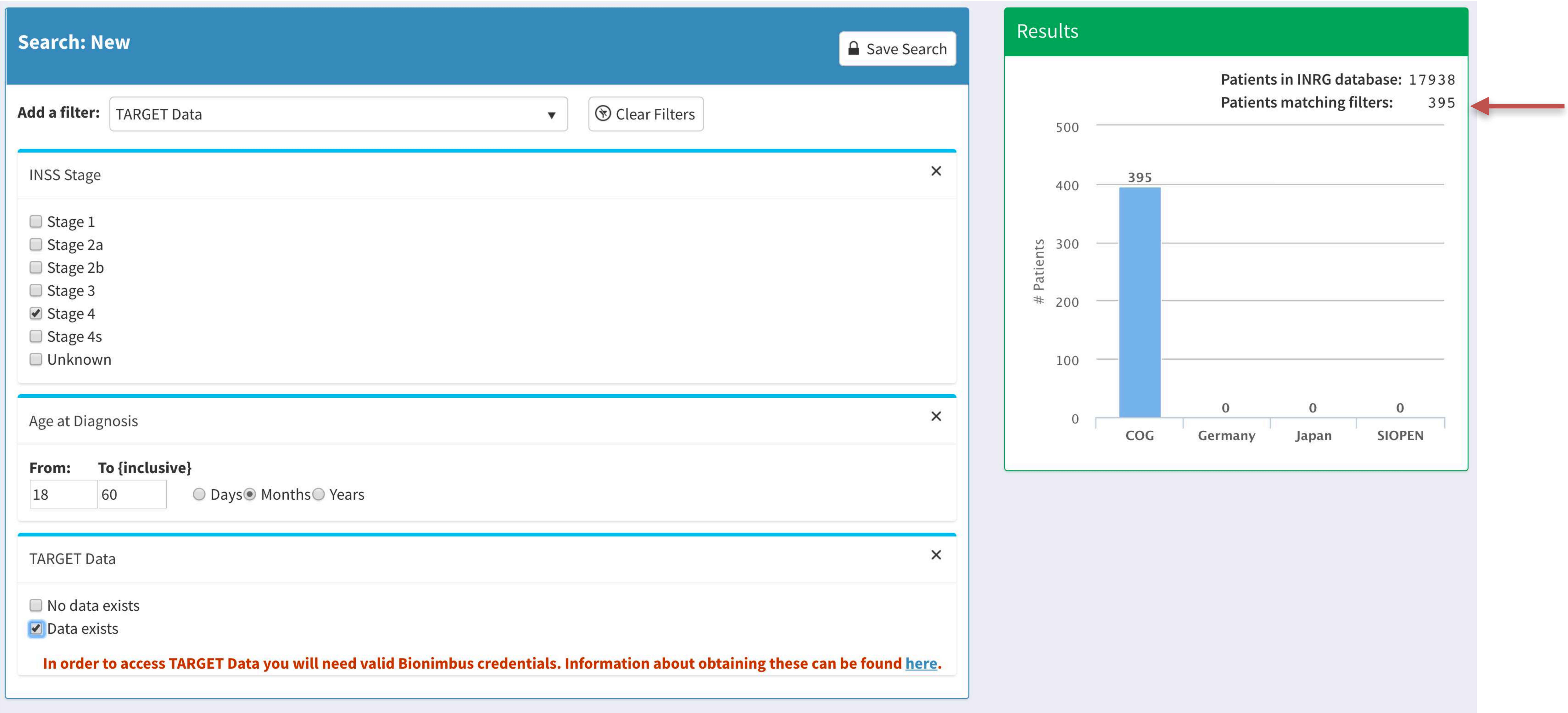
Results

Patients in INRG database: 17938

Patients matching filters: 3906

Links to external data sets

Neuroblastoma Commons Cohort Discovery



Search: New

Add a filter: TARGET Data ▾

INSS Stage ✕

- ☐ Stage 1
- ☐ Stage 2a
- ☐ Stage 2b
- ☐ Stage 3
- ☑ Stage 4
- ☐ Stage 4s
- ☐ Unknown

Age at Diagnosis ✕

From: 18 **To {inclusive}** 60 ○ Days ● Months ○ Years

TARGET Data ✕

- ☐ No data exists
- ☑ Data exists

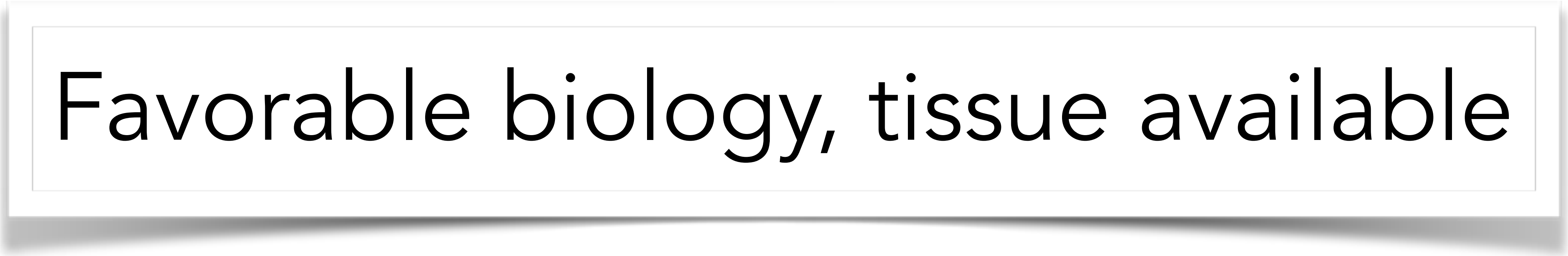
In order to access TARGET Data you will need valid Bionimbus credentials. Information about obtaining these can be found here.

Results

Patients in INRG database: 17938

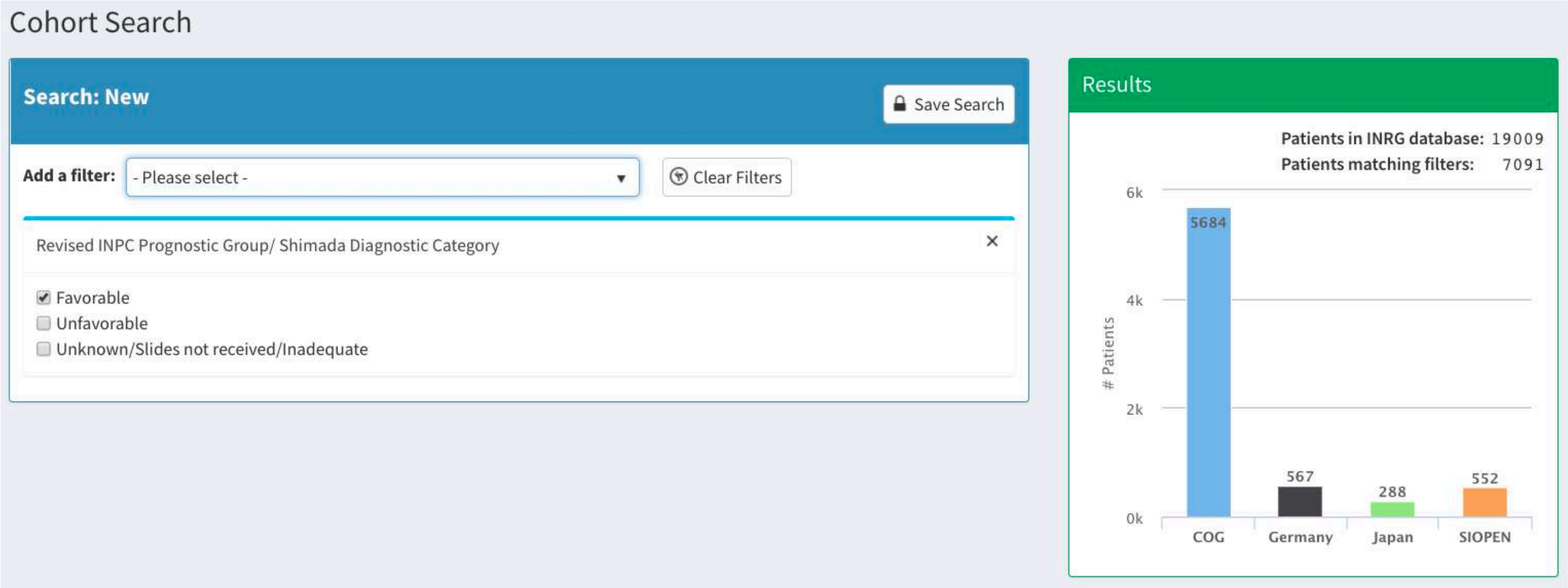
Patients matching filters: 395

Demo query 2



Favorable biology, tissue available

Neuroblastoma Commons Cohort Discovery



Cohort Search

Search: New

Add a filter: - Please select -

Revised INPC Prognostic Group/ Shimada Diagnostic Category

☑ Favorable
□ Unfavorable
□ Unknown/Slides not received/Inadequate

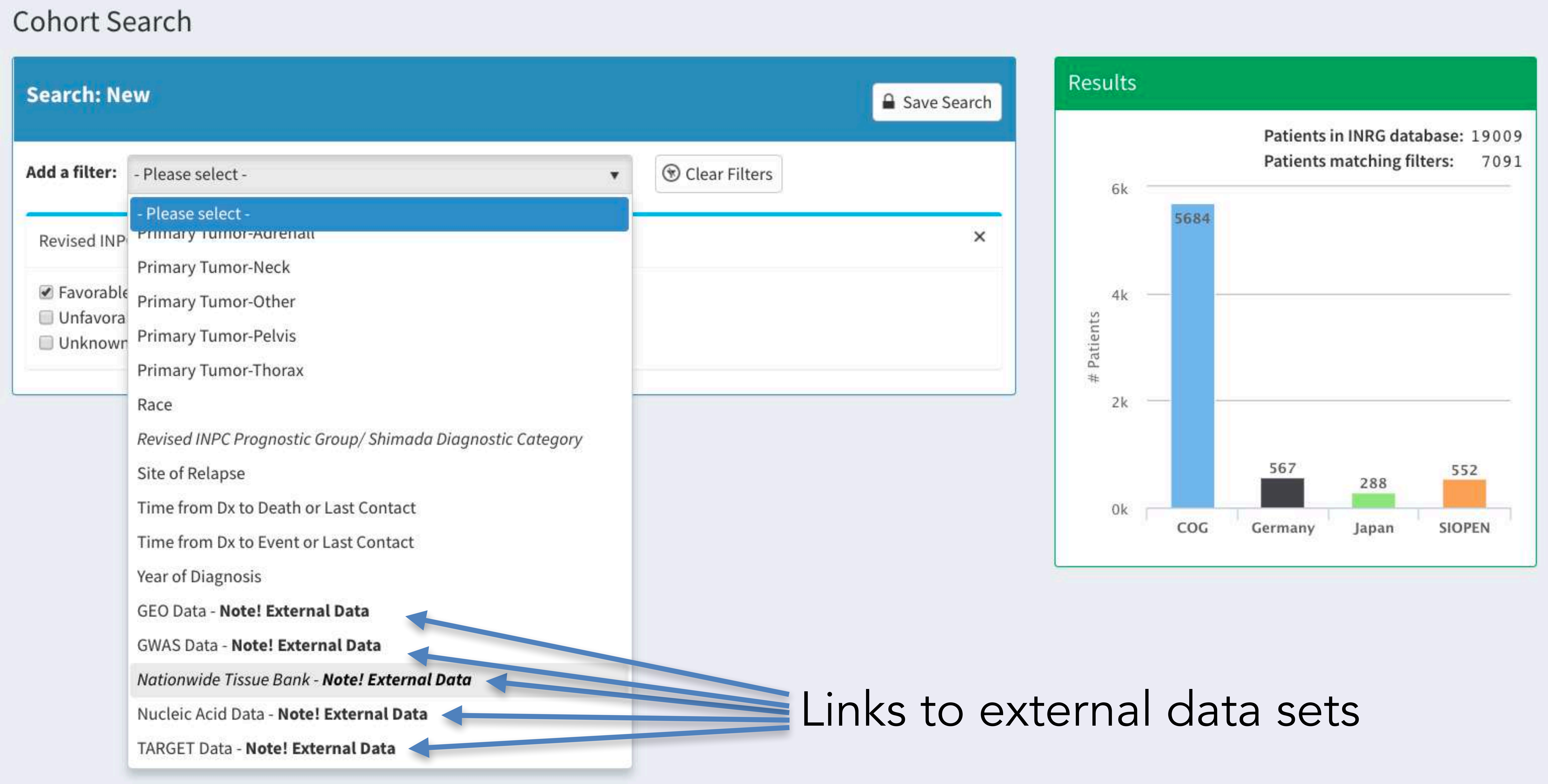
Results

Patients in INRG database: 19009
Patients matching filters: 7091

Favorable biology, tissue available

THE UNIVERSITY OF CHICAGO MEDICINE & BIOLOGICAL SCIENCES | CENTER FOR RESEARCH INFORMATICS

Neuroblastoma Commons Cohort Discovery



Cohort Search

Search: New

🔒 Save Search

Add a filter: - Please select - ▾

Clear Filters

- Please select -

Primary Tumor-Adrenal

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

□ Unfavora

□ Unknown

×

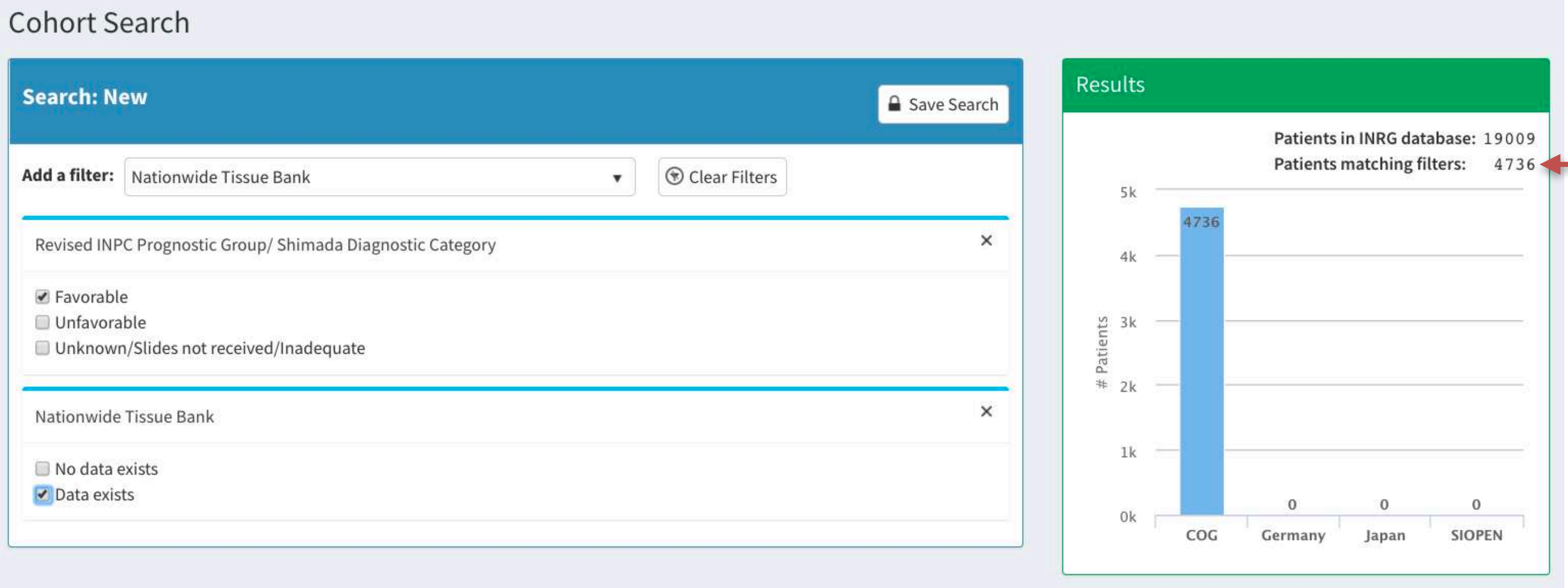
Results

Patients in INRG database: 19009

Patients matching filters: 7091

Links to external data sets

Neuroblastoma Commons Cohort Discovery



Cohort Search

Search: New

Save Search

Add a filter: Nationwide Tissue Bank

Clear Filters

Revised INPC Prognostic Group/ Shimada Diagnostic Category

☑ Favorable
□ Unfavorable
□ Unknown/Slides not received/Inadequate

Nationwide Tissue Bank

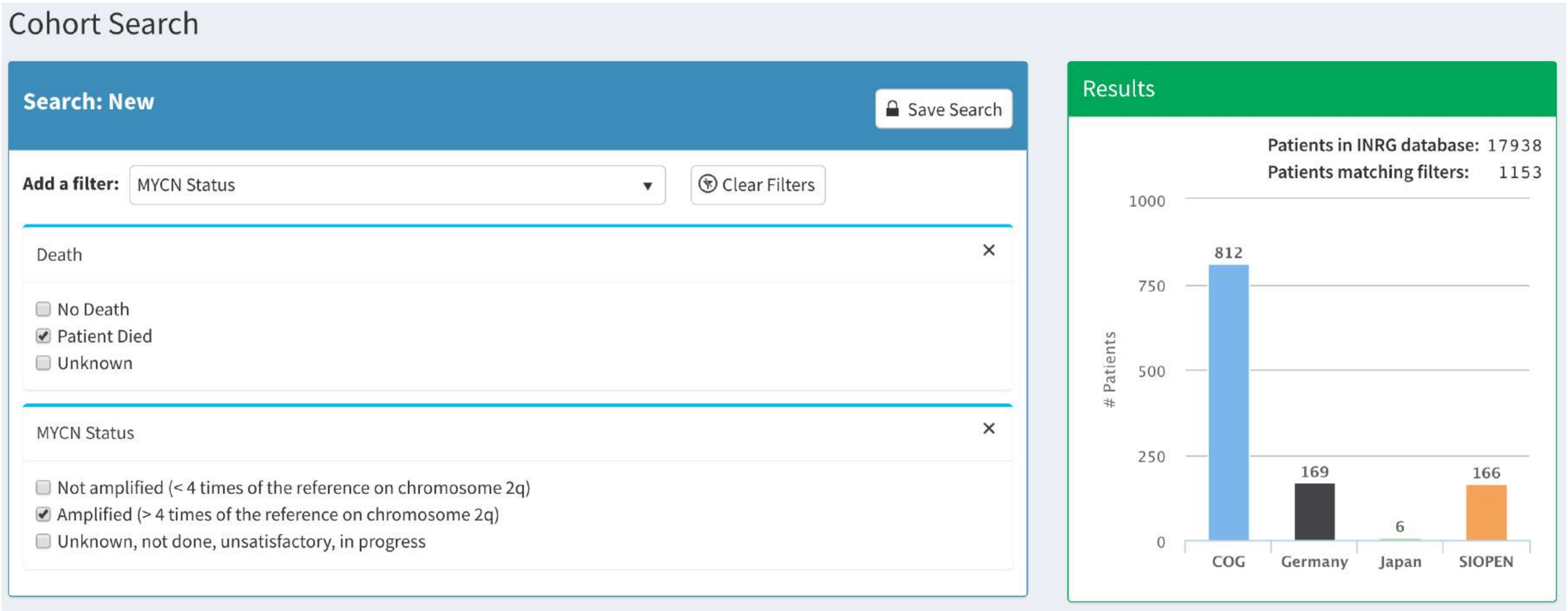
□ No data exists
☑ Data exists

Results

Patients in INRG database: 19009
Patients matching filters: 4736

Favorable biology, tissue available

Neuroblastoma Data Commons



Cohort Search

Search: New

Save Search

Add a filter: MYCN Status

Clear Filters

Death

- ☐ No Death
- ☑ Patient Died
- ☐ Unknown

MYCN Status

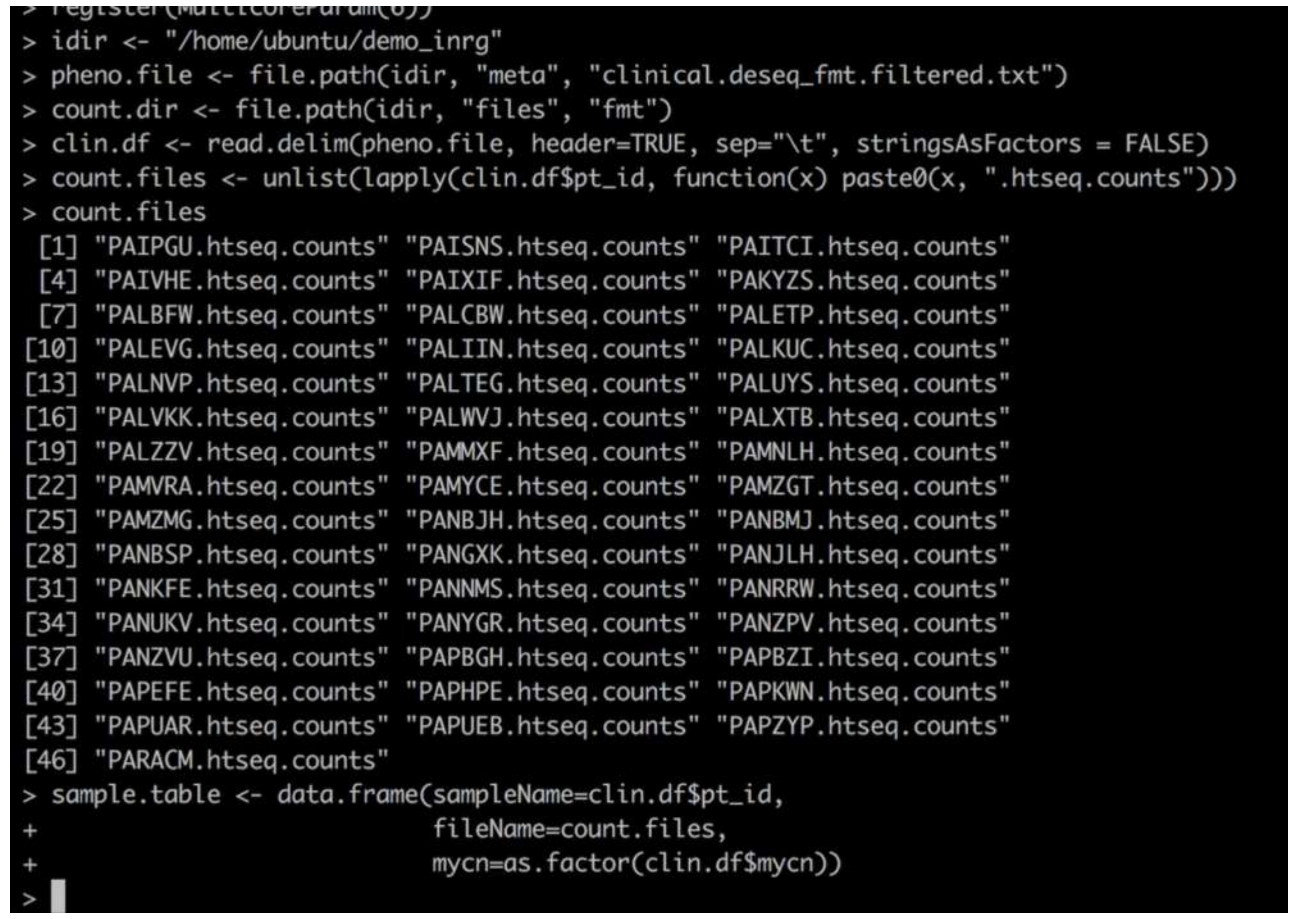
- ☐ Not amplified (< 4 times of the reference on chromosome 2q)
- ☑ Amplified (> 4 times of the reference on chromosome 2q)
- ☐ Unknown, not done, unsatisfactory, in progress

Results

Patients in INRG database: 17938

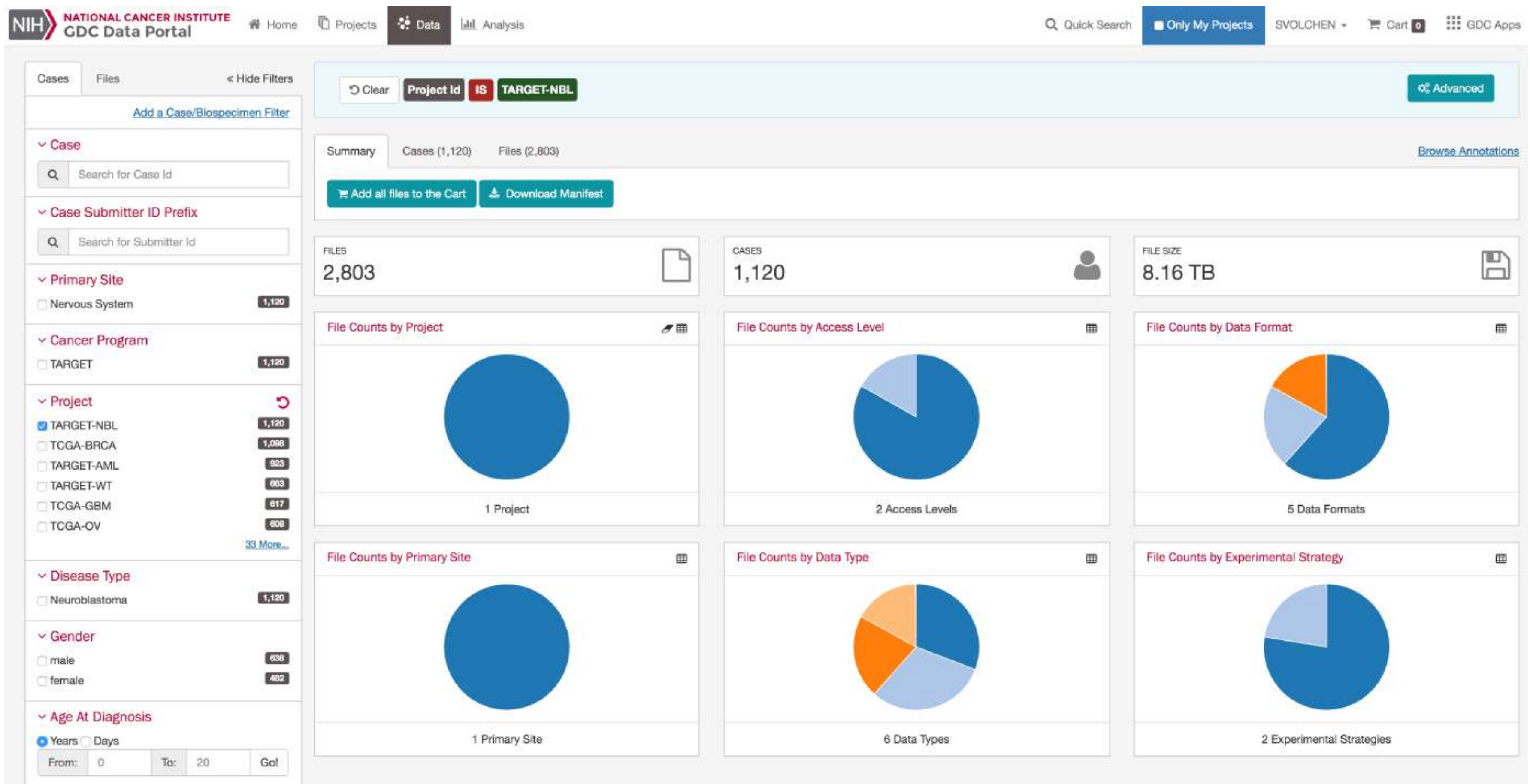
Patients matching filters: 1153

INRG cohort discovery



```
> register(MulticoreParam(6))
> 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$pt_id, function(x) paste0(x, ".htseq.counts")))
> count.files
 [1] "PAIPGU.htseq.counts" "PAISNS.htseq.counts" "PAITCI.htseq.counts"
 [4] "PAIVHE.htseq.counts" "PAIXIF.htseq.counts" "PAKYZS.htseq.counts"
 [7] "PALBFW.htseq.counts" "PALCBW.htseq.counts" "PALETP.htseq.counts"
[10] "PALEVG.htseq.counts" "PALIIN.htseq.counts" "PALKUC.htseq.counts"
[13] "PALNVP.htseq.counts" "PALTEG.htseq.counts" "PALUYS.htseq.counts"
[16] "PALVKK.htseq.counts" "PALWVJ.htseq.counts" "PALXTB.htseq.counts"
[19] "PALZZV.htseq.counts" "PAMMXF.htseq.counts" "PAMNLH.htseq.counts"
[22] "PAMVRA.htseq.counts" "PAMYCE.htseq.counts" "PAMZGT.htseq.counts"
[25] "PAMZMG.htseq.counts" "PANBJH.htseq.counts" "PANBMJ.htseq.counts"
[28] "PANBSP.htseq.counts" "PANGXK.htseq.counts" "PANJLH.htseq.counts"
[31] "PANKFE.htseq.counts" "PANNMS.htseq.counts" "PANRRW.htseq.counts"
[34] "PANUKV.htseq.counts" "PANYGR.htseq.counts" "PANZPV.htseq.counts"
[37] "PANZVU.htseq.counts" "PAPBGH.htseq.counts" "PAPBZI.htseq.counts"
[40] "PAPEFE.htseq.counts" "PAPHPE.htseq.counts" "PAPKWN.htseq.counts"
[43] "PAPUAR.htseq.counts" "PAPUEB.htseq.counts" "PAPZYP.htseq.counts"
[46] "PARACM.htseq.counts"
> sample.table <- data.frame(sampleName=clin.df$pt_id,
+                            fileName=count.files,
+                            mycn=as.factor(clin.df$mycn))
> 
```

Command-line analysis



Project Id IS TARGET-NBL

Case: Search for Case Id

Case Submitter ID Prefix: Search for Submitter Id

Primary Site
- ☐ Nervous System 1,120

Cancer Program
- ☐ TARGET 1,120

Project
- ☑ TARGET-NBL 1,120
- ☐ TCGA-BRCA 1,098
- ☐ TARGET-AML 923
- ☐ TARGET-WT 663
- ☐ TCGA-GBM 617
- ☐ TCGA-OV 608

Disease Type
- ☐ Neuroblastoma 1,120

Gender
- ☐ male 638
- ☐ female 482

Age At Diagnosis
- ● Years ○ Days
- From: 0 To: 20

FILES 2,803

CASES 1,120

FILE SIZE 8.16 TB

File Counts by Project — 1 Project

File Counts by Access Level — 2 Access Levels

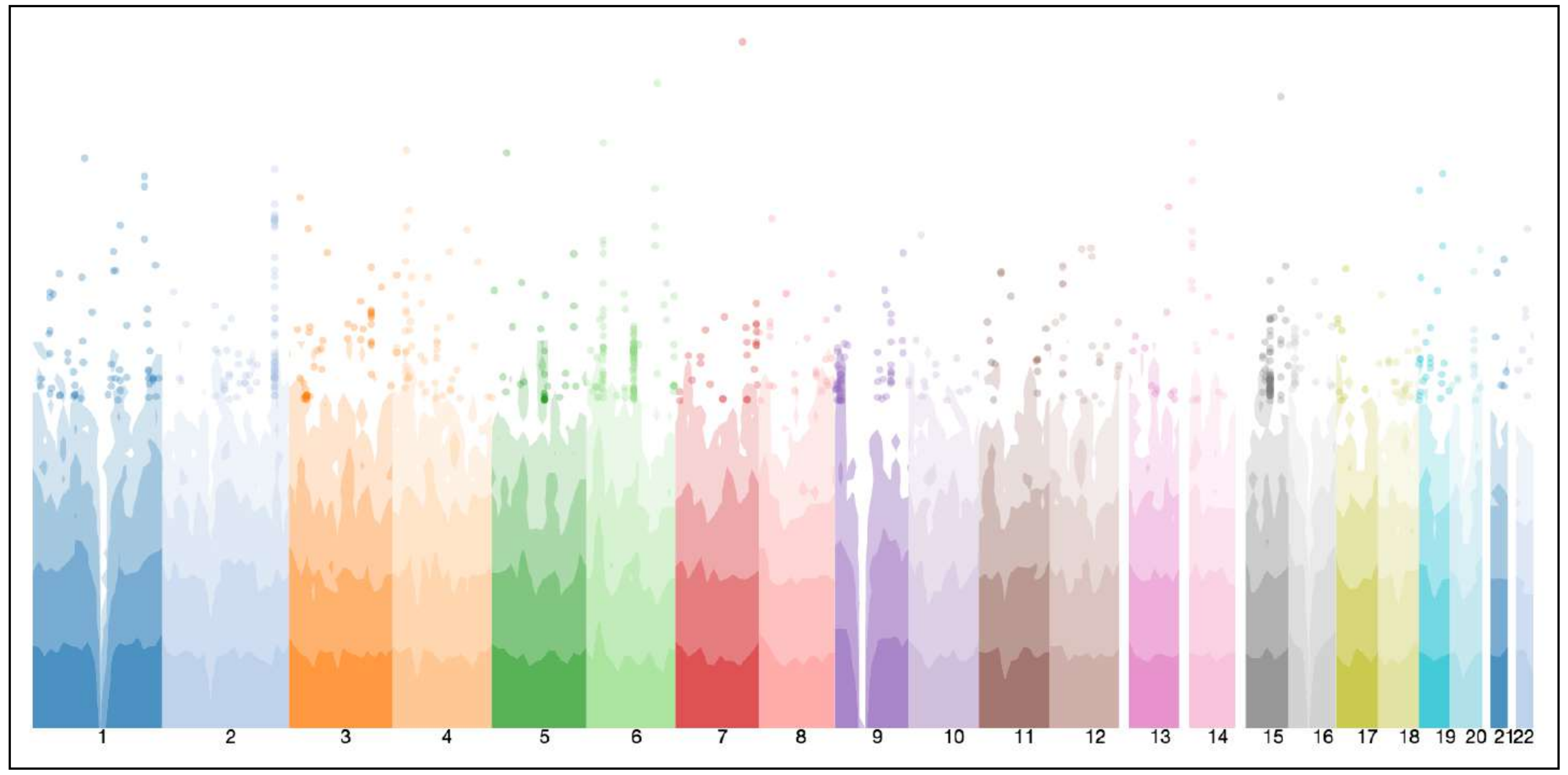
File Counts by Data Format — 5 Data Formats

File Counts by Primary Site — 1 Primary Site

File Counts by Data Type — 6 Data Types

File Counts by Experimental Strategy — 2 Experimental Strategies

Genomic data commons search

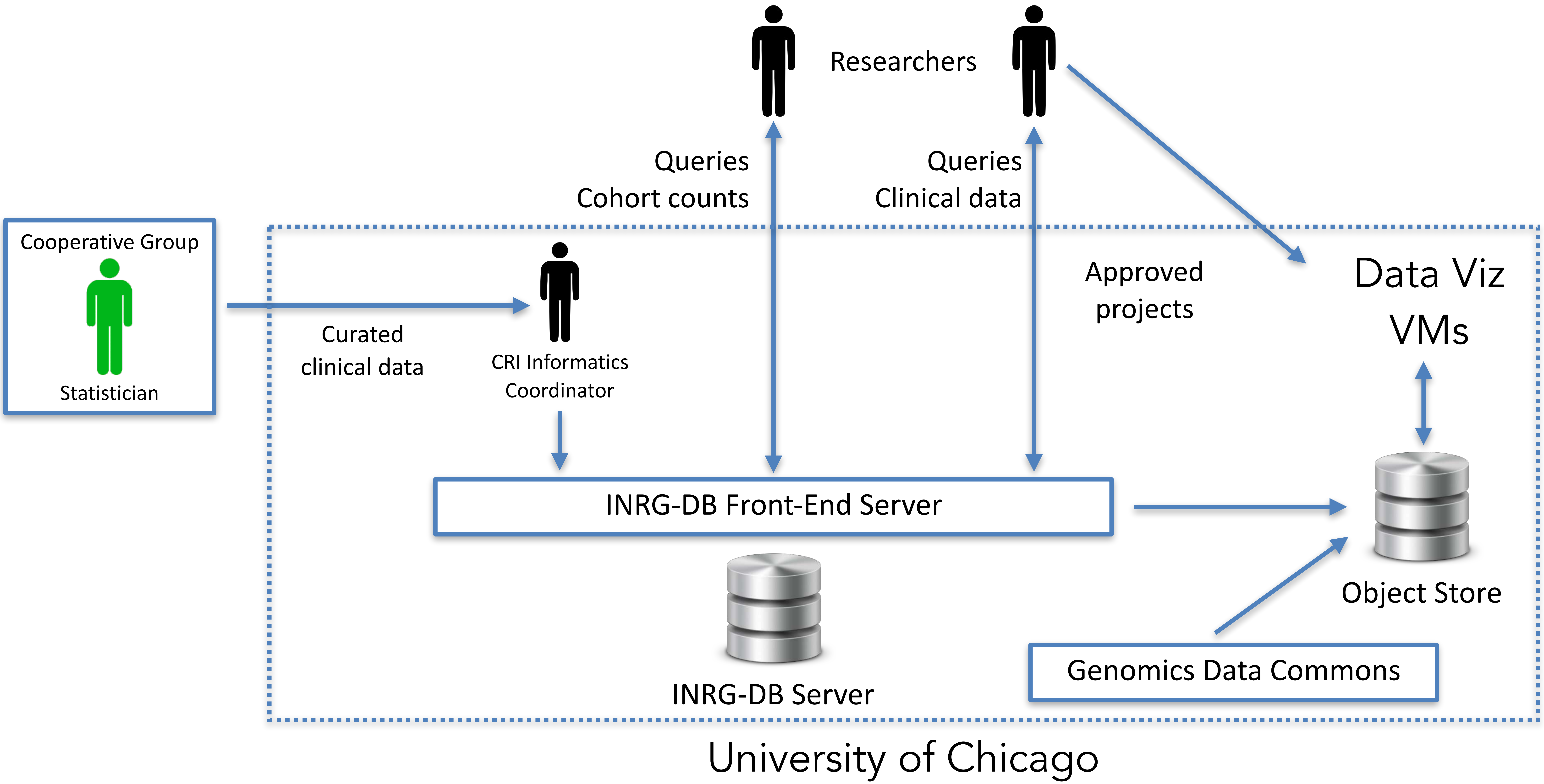


Visualization

Data 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**

Neuroblastoma data commons



Researchers

Queries
Cohort counts

Queries
Clinical data

Cooperative Group

Statistician

Curated clinical data

CRI Informatics Coordinator

Approved projects

Data Viz VMs

INRG-DB Front-End Server

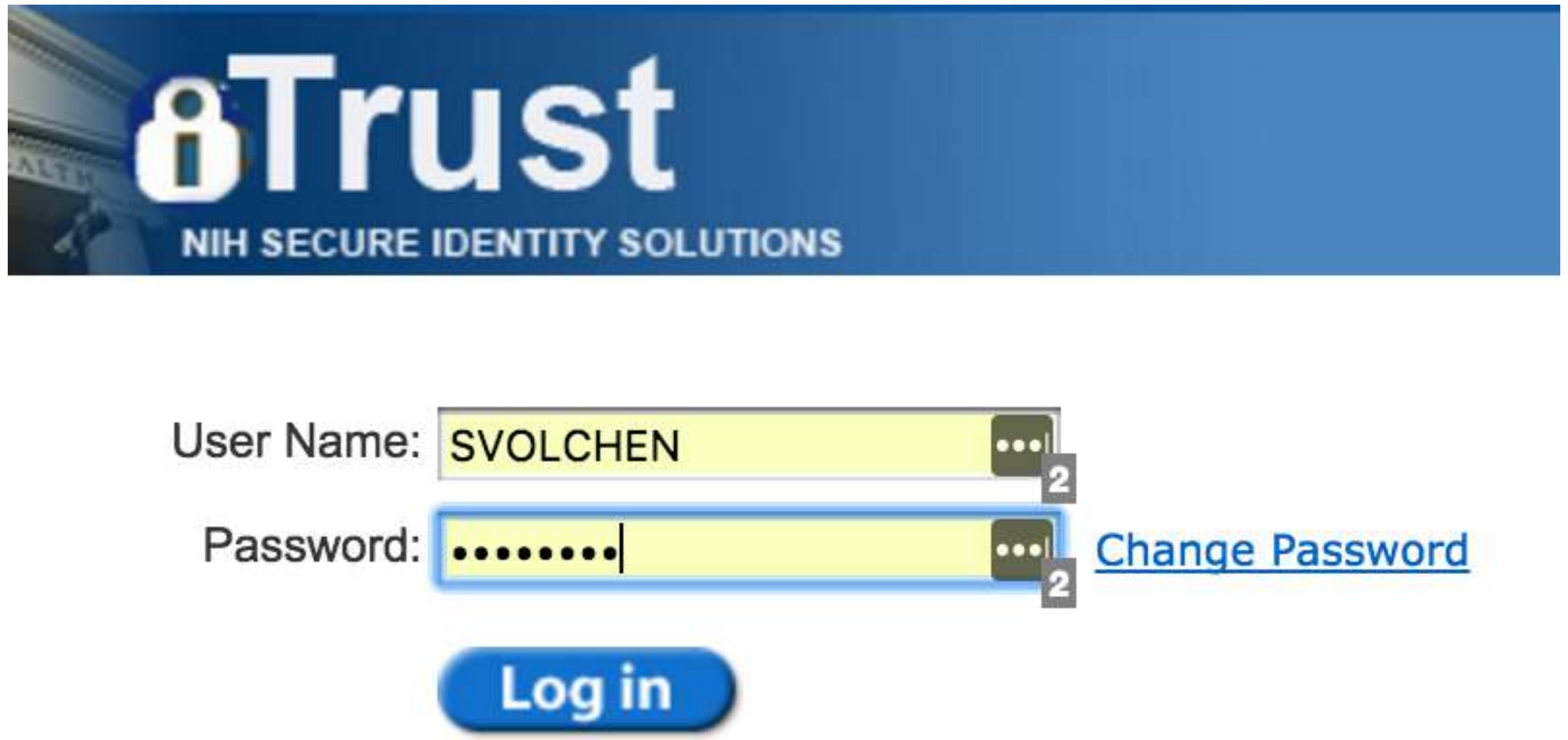
INRG-DB Server

Object Store

Genomics Data Commons

University of Chicago

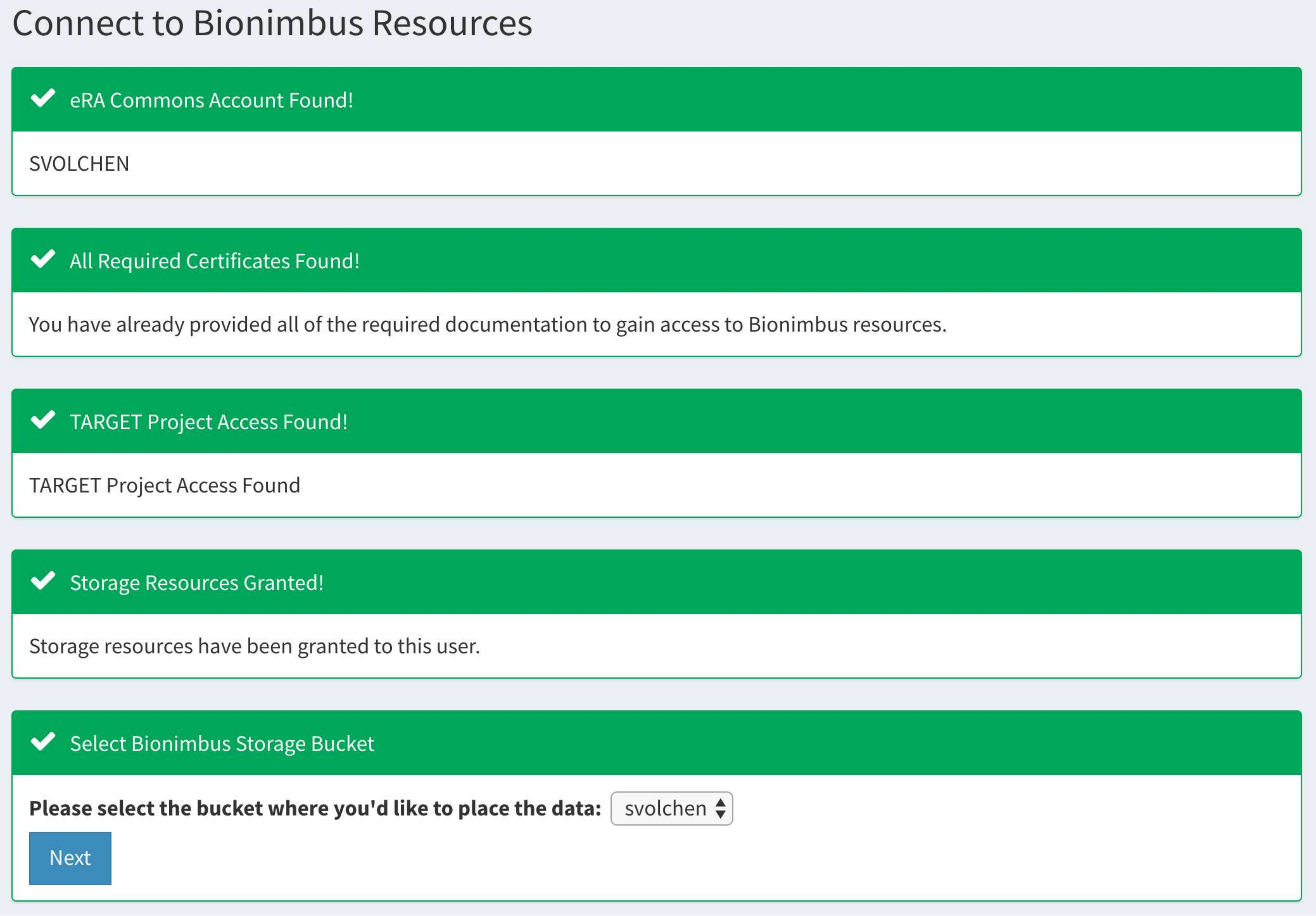
Federated Authentication



User Name: SVOLCHEN

Password: ••••••••

Governance / Regulation / Compliance



Connect to Bionimbus Resources

✓ eRA Commons Account Found!

SVOLCHEN

✓ All Required Certificates Found!

You have already provided all of the required documentation to gain access to Bionimbus resources.

✓ TARGET Project Access Found!

TARGET Project Access Found

✓ Storage Resources Granted!

Storage resources have been granted to this user.

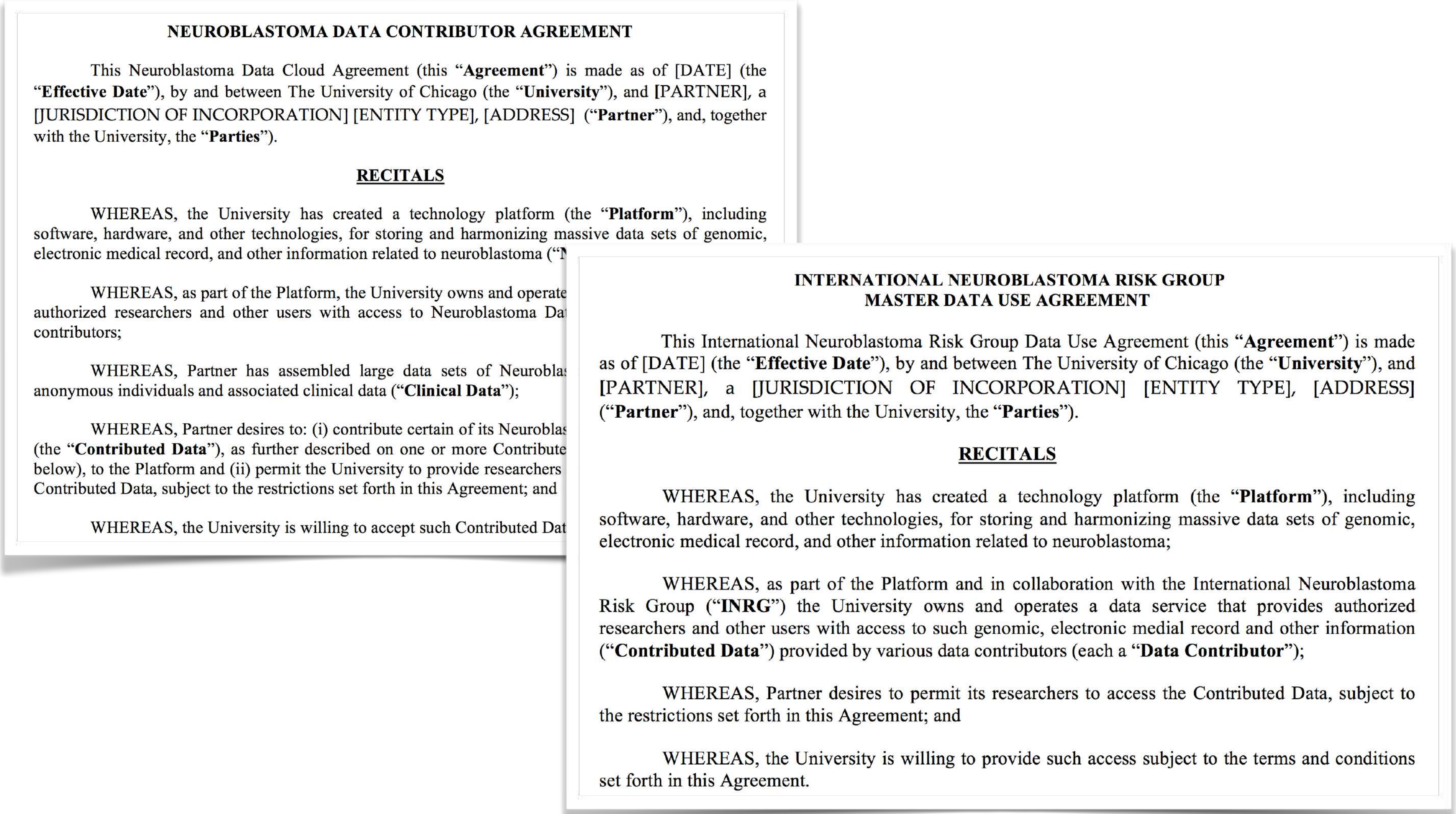
✓ Select Bionimbus Storage Bucket

Please select the bucket where you'd like to place the data: svolchen

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

Governance / Regulation / Compliance



NEUROBLASTOMA DATA CONTRIBUTOR AGREEMENT

This Neuroblastoma Data Cloud Agreement (this “**Agreement**”) is made as of [DATE] (the “**Effective Date**”), by and between The University of Chicago (the “**University**”), and [PARTNER], a [JURISDICTION OF INCORPORATION] [ENTITY TYPE], [ADDRESS] (“**Partner**”), and, together with the University, the “**Parties**”).

RECITALS

WHEREAS, the University has created a technology platform (the “**Platform**”), including software, hardware, and other technologies, for storing and harmonizing massive data sets of genomic, electronic medical record, and other information related to neuroblastoma (“N

WHEREAS, as part of the Platform, the University owns and operate authorized researchers and other users with access to Neuroblastoma Da contributors;

WHEREAS, Partner has assembled large data sets of Neuroblas anonymous individuals and associated clinical data (“**Clinical Data**”);

WHEREAS, Partner desires to: (i) contribute certain of its Neuroblas (the “**Contributed Data**”), as further described on one or more Contribute below), to the Platform and (ii) permit the University to provide researchers Contributed Data, subject to the restrictions set forth in this Agreement; and

WHEREAS, the University is willing to accept such Contributed Dat

INTERNATIONAL NEUROBLASTOMA RISK GROUP
MASTER DATA USE AGREEMENT

This International Neuroblastoma Risk Group Data Use Agreement (this “**Agreement**”) is made as of [DATE] (the “**Effective Date**”), by and between The University of Chicago (the “**University**”), and [PARTNER], a [JURISDICTION OF INCORPORATION] [ENTITY TYPE], [ADDRESS] (“**Partner**”), and, together with the University, the “**Parties**”).

RECITALS

WHEREAS, the University has created a technology platform (the “**Platform**”), including software, hardware, and other technologies, for storing and harmonizing massive data sets of genomic, electronic medical record, and other information related to neuroblastoma;

WHEREAS, as part of the Platform and in collaboration with the International Neuroblastoma Risk Group (“**INRG**”) the University owns and operates a data service that provides authorized researchers and other users with access to such genomic, electronic medial record and other information (“**Contributed Data**”) provided by various data contributors (each a “**Data Contributor**”);

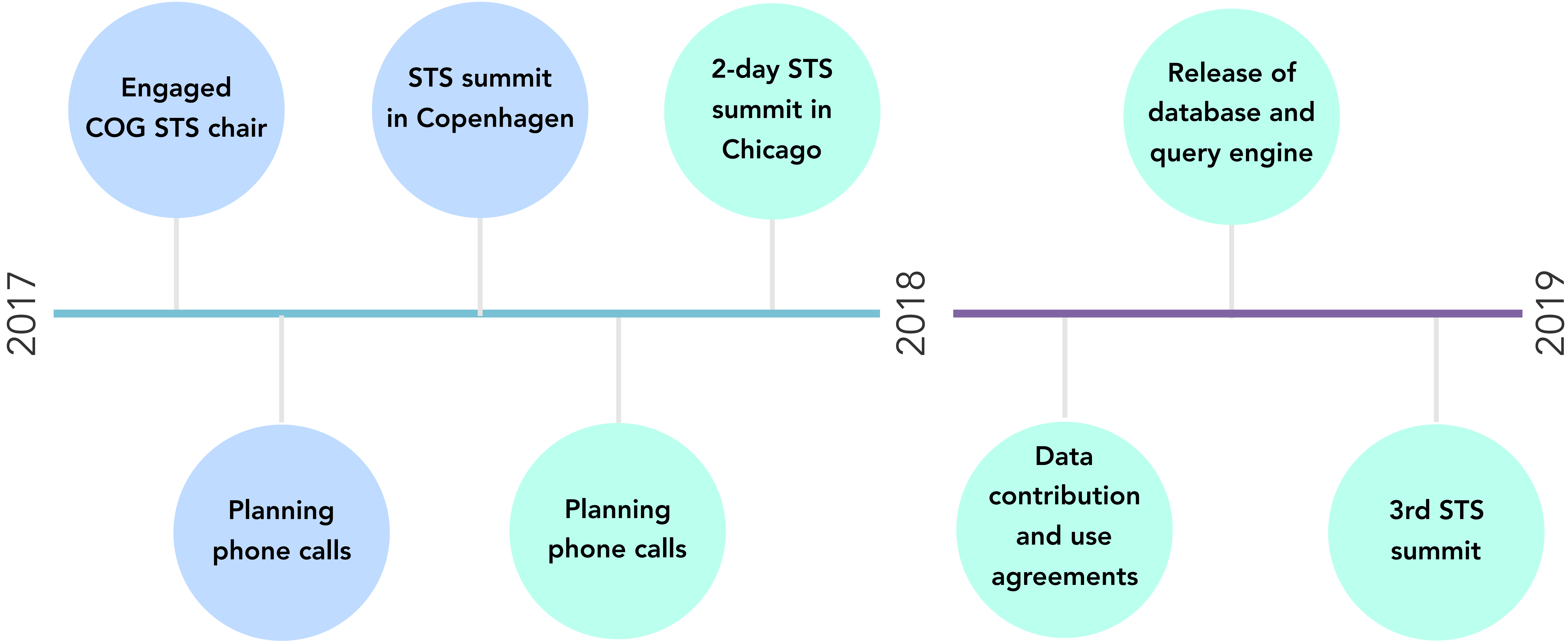
WHEREAS, Partner desires to permit its researchers to access the Contributed Data, subject to the restrictions set forth in this Agreement; and

WHEREAS, the University is willing to provide such access subject to the terms and conditions set forth in this Agreement.

Building pediatric cancer commons

- Acute myeloid leukemia
- Acute lymphoblastic leukemia
- Bone tumors
- Central nervous system tumors
- Germ cell tumors
- Hodgkin Disease
- Neuroblastoma
- Non-Hodgkin lymphoma
- Renal tumors
- Soft tissue sarcoma

Building a soft-tissue sarcoma data commons



2017

Engaged COG STS chair

Planning phone calls

STS summit in Copenhagen

Planning phone calls

2-day STS summit in Chicago

2018

Data contribution and use agreements

Release of database and query engine

3rd STS summit

2019

THE UNIVERSITY OF CHICAGO MEDICINE & BIOLOGICAL SCIENCES

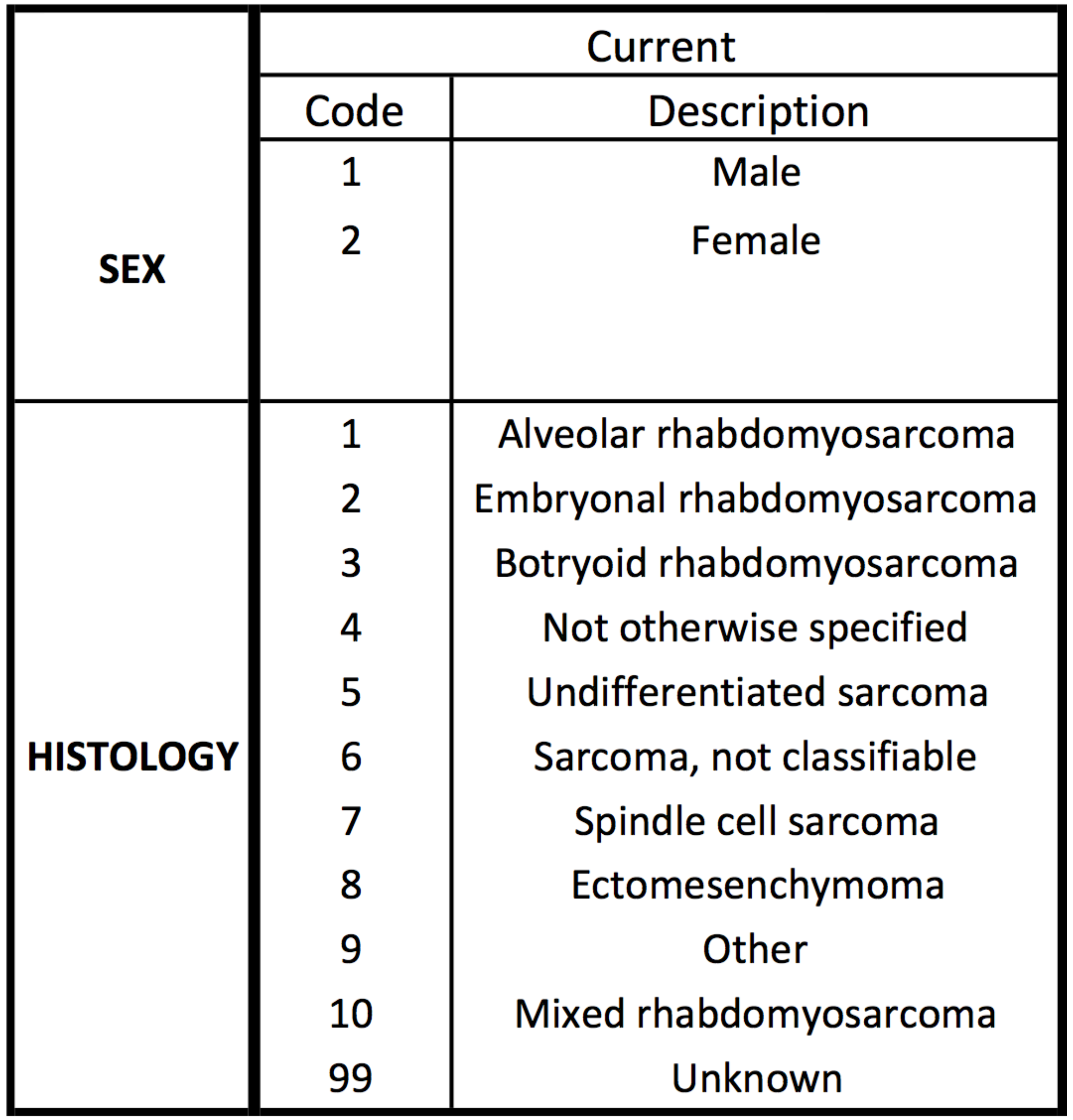
CENTER FOR RESEARCH INFORMATICS

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

The importance of standards

Pediatric rhabdomyosarcoma



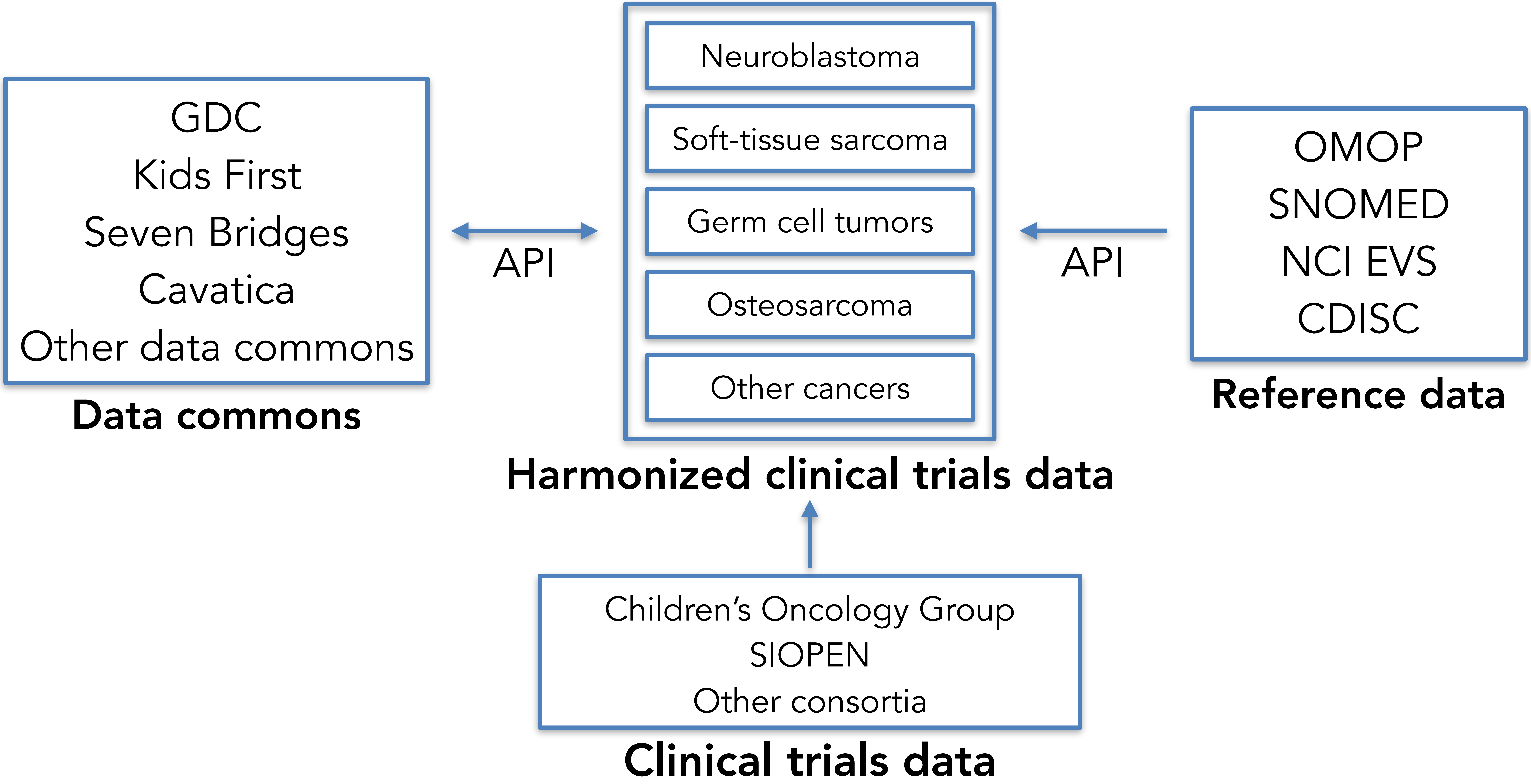
	Current	
	Code	Description
SEX	1	Male
	2	Female
HISTOLOGY	1	Alveolar rhabdomyosarcoma
	2	Embryonal rhabdomyosarcoma
	3	Botryoid rhabdomyosarcoma
	4	Not otherwise specified
	5	Undifferentiated sarcoma
	6	Sarcoma, not classifiable
	7	Spindle cell sarcoma
	8	Ectomesenchymoma
	9	Other
	10	Mixed rhabdomyosarcoma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

The importance of standards

Pediatric rhabdomyosarcoma

	Current		Proposed		
	Code	Description	Code	Description	Source
SEX	1	Male	C20197	Male	NCI
	2	Female	C16576	Female	
			C45908	Intersex	
			C17998	Unknown	
HISTOLOGY	1	Alveolar rhabdomyosarcoma	404053004	Alveolar rhabdomyosarcoma	SNOMED
	2	Embryonal rhabdomyosarcoma	404051002	Embryonal rhabdomyosarcoma	
	3	Botryoid rhabdomyosarcoma	404052009	Botryoid rhabdomyosarcoma	
	4	Not otherwise specified	302847003	Rhabdomyosarcoma	
	5	Undifferentiated sarcoma	128734000	Undifferentiated sarcoma	
	6	Sarcoma, not classifiable	302847003	Rhabdomyosarcoma	
	7	Spindle cell sarcoma	9801004	Spindle cell sarcoma	
	8	Ectomesenchymoma	128750008	Rhabdomyosarcoma with ganglionic differentiation	
	9	Other	--	--	
	10	Mixed rhabdomyosarcoma	62383007	Mixed type rhabdomyosarcoma	
	99	Unknown	C17998	Unknown	NCI

Paradigm for pediatric cancer commons



GDC
Kids First
Seven Bridges
Cavatica
Other data commons

Data commons

API

Neuroblastoma
Soft-tissue sarcoma
Germ cell tumors
Osteosarcoma
Other cancers

Harmonized clinical trials data

API

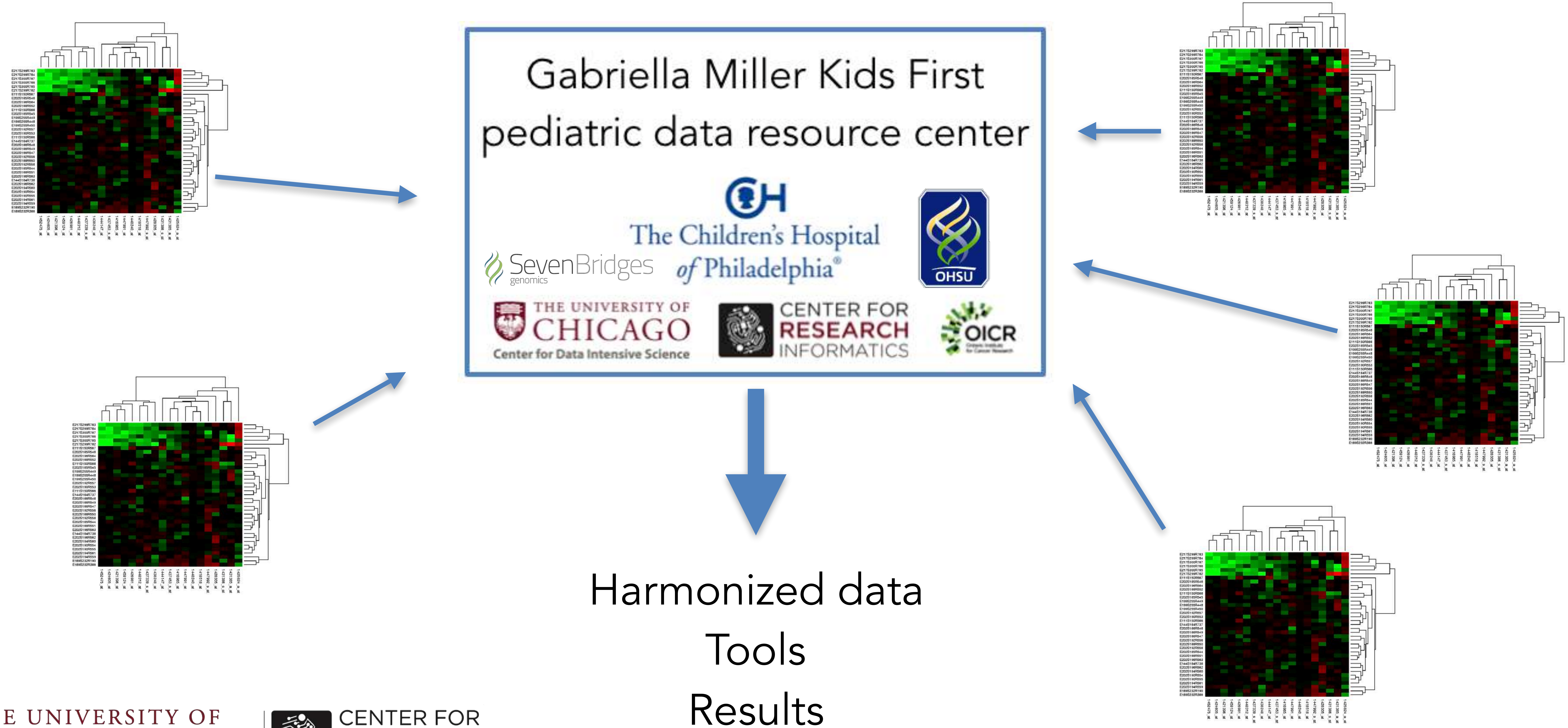
OMOP
SNOMED
NCI EVS
CDISC

Reference data

Children's Oncology Group
SIOPEN
Other consortia

Clinical trials data

Gabriella Miller Kids First Pediatric Data Resource Center



Gabriella Miller Kids First pediatric data resource center

The Children's Hospital of Philadelphia®

SevenBridges genomics

OHSU

THE UNIVERSITY OF CHICAGO Center for Data Intensive Science

CENTER FOR RESEARCH INFORMATICS

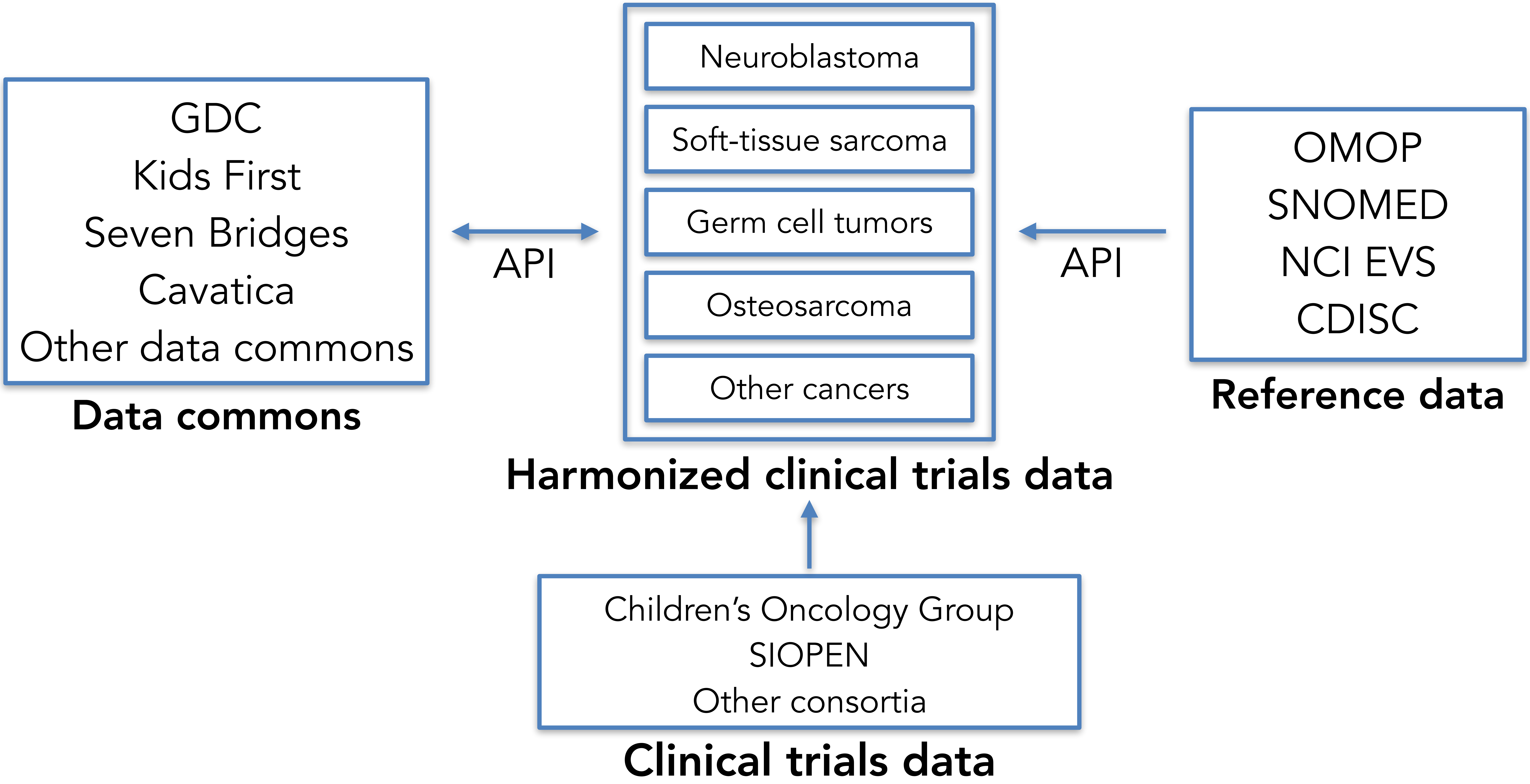
OICR

Harmonized data

Tools

Results

Paradigm for pediatric cancer commons



GDC
Kids First
Seven Bridges
Cavatica
Other data commons

Data commons

API

Neuroblastoma
Soft-tissue sarcoma
Germ cell tumors
Osteosarcoma
Other cancers

Harmonized clinical trials data

API

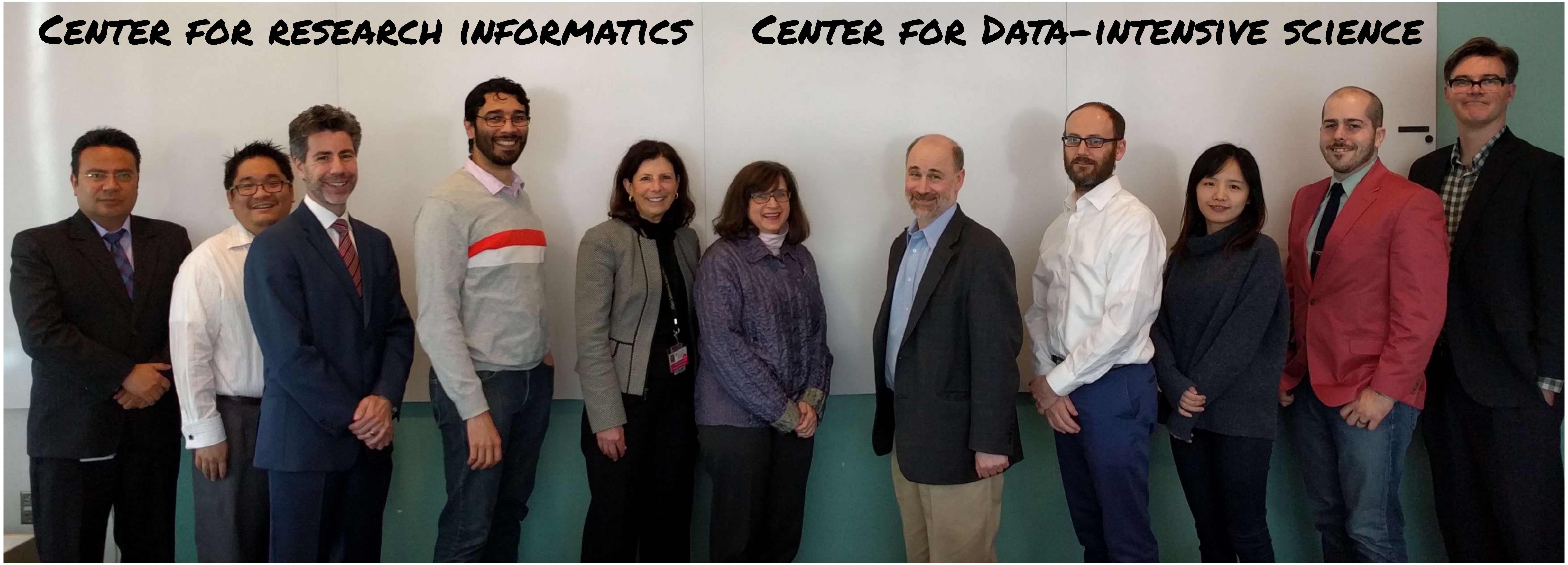
OMOP
SNOMED
NCI EVS
CDISC

Reference data

Children's Oncology Group
SIOPEN
Other consortia

Clinical trials data

Acknowledgements


Center for research informatics
Center for Data-intensive science

WILLIAM GUY FORBECK
RESEARCH FOUNDATION
wgfrf.org

St. Baldrick's
FOUNDATION
Conquer Childhood Cancers

Alex's Lemonade Stand

The Super Jake Foundation

LITTLE HEROES

Children's Neuroblastoma Cancer Foundation

THE UNIVERSITY OF CHICAGO MEDICINE & BIOLOGICAL SCIENCES

CENTER FOR RESEARCH INFORMATICS