Open science and data sharing *in practice*

Justin Guinney, PhD
Director, Computational Oncology
Sage Bionetworks

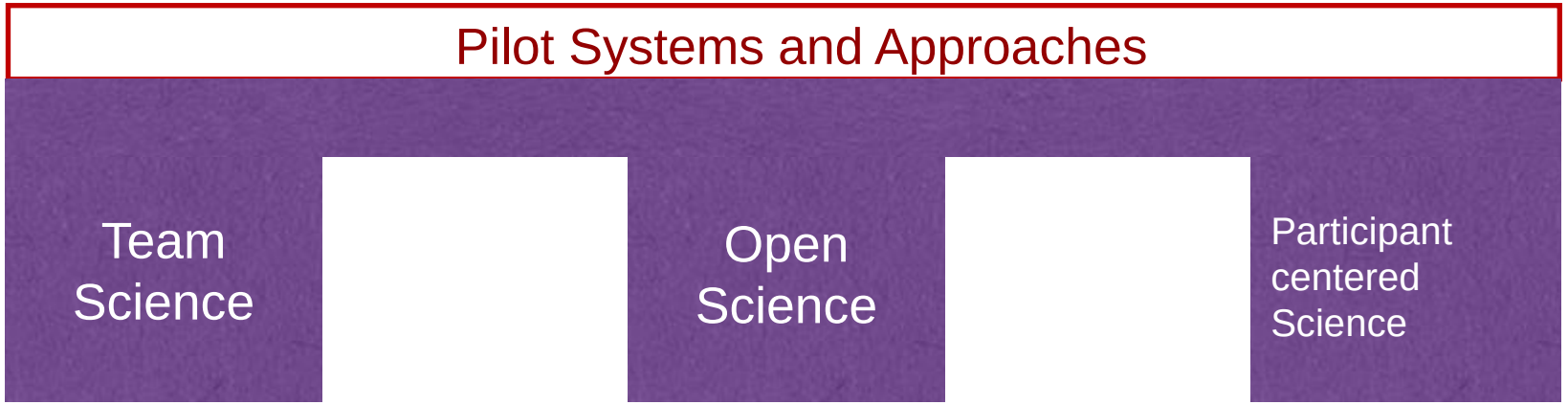
Co-Director
DREAM Challenges

Sage BIONETWORKS

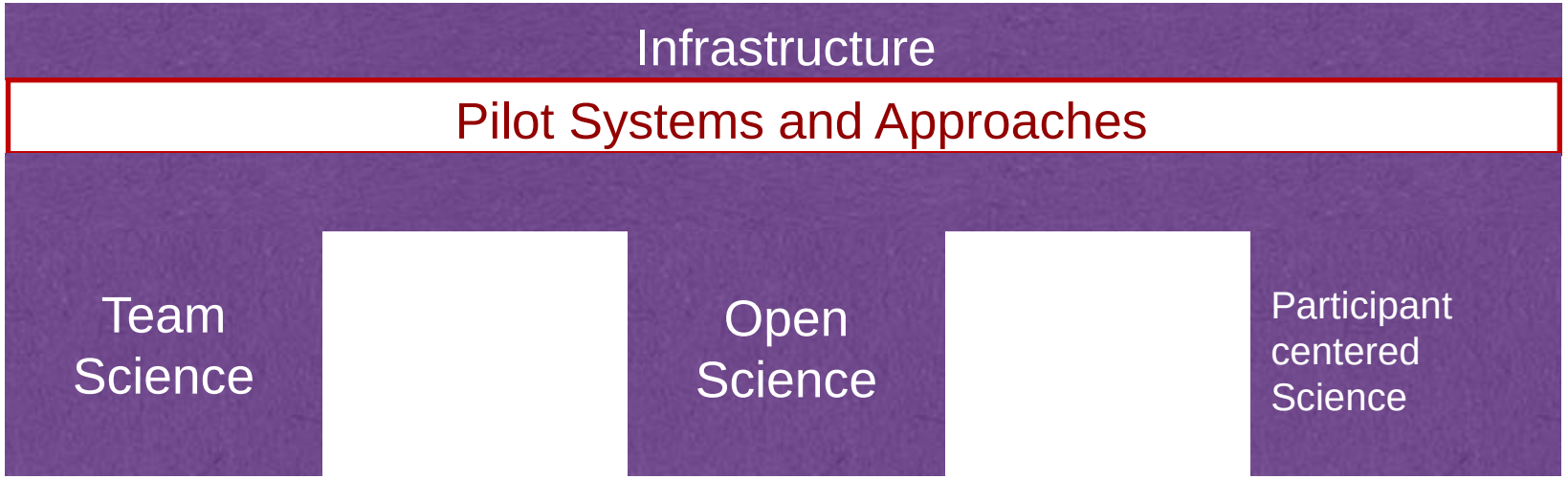
promote open systems, incentives, and norms
to redefine how complex biological data is
gathered, shared, and used

our research is built on three pillars

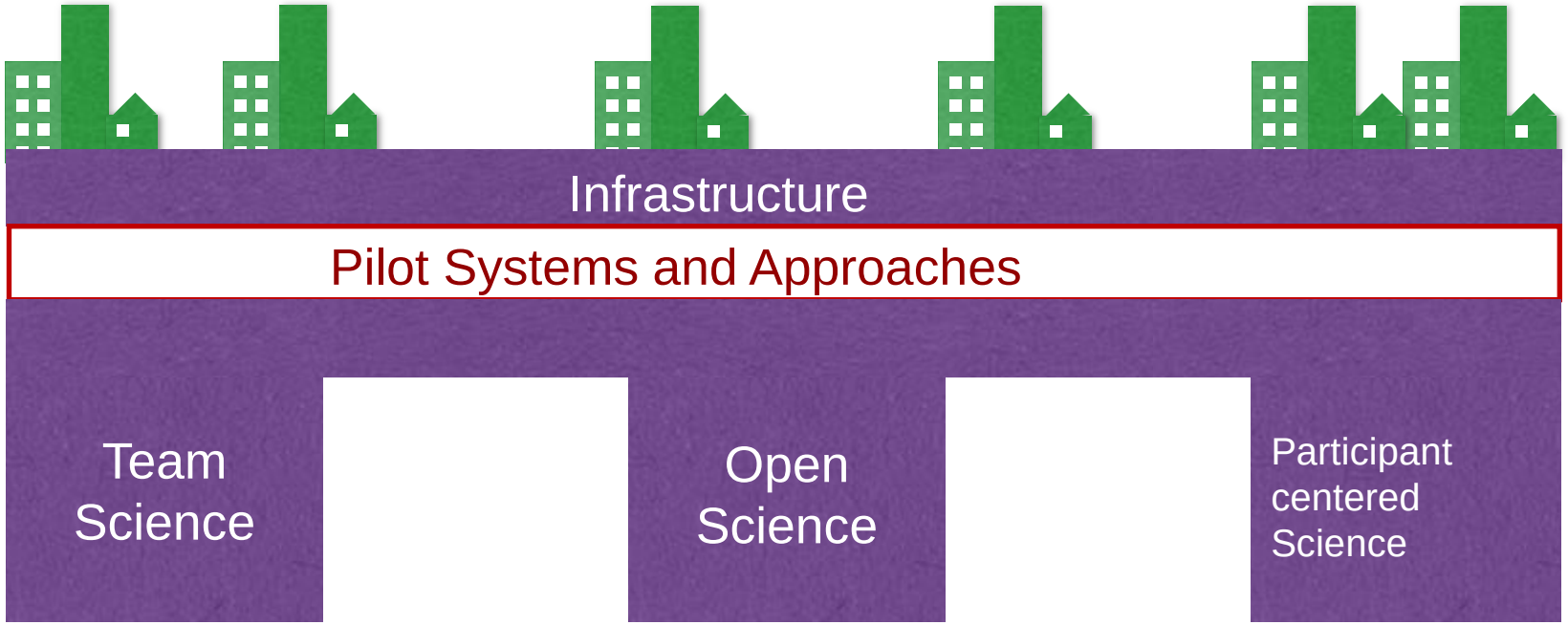
we pilot approaches
to create open systems, incentives, and norms



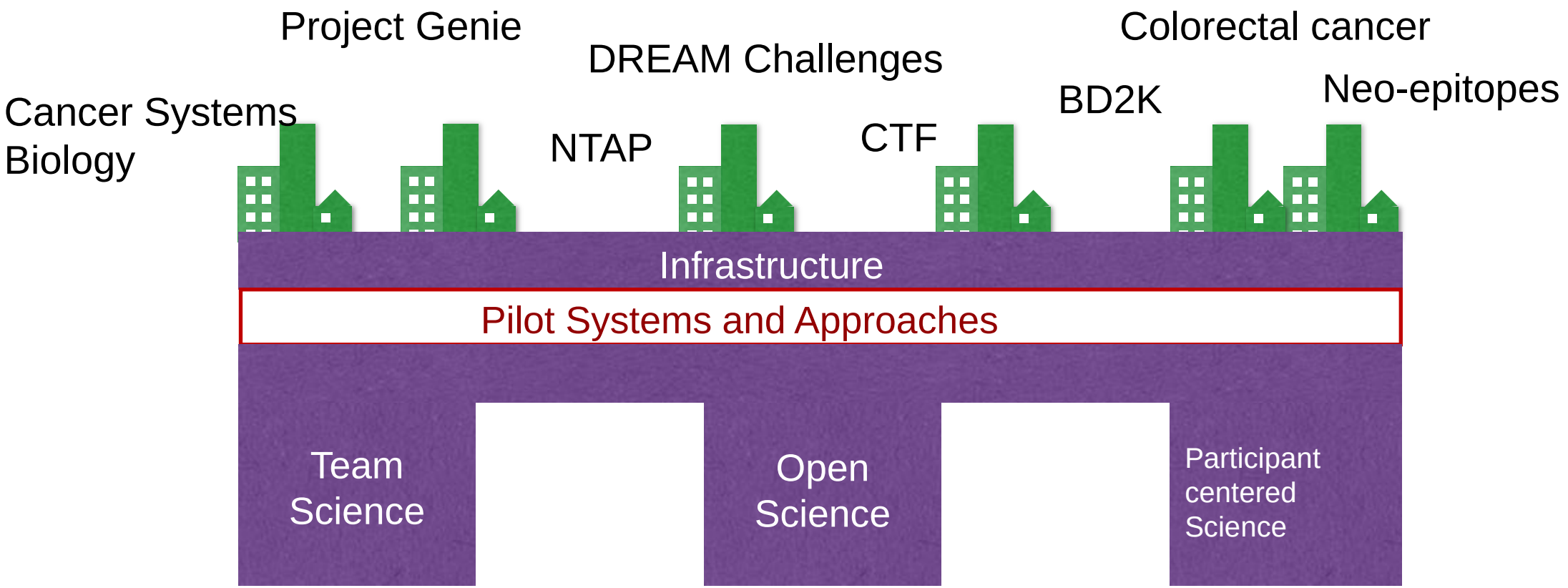
we build infrastructure
to provide robust, reusable solutions



we support research communities
that operate under these principles



cancer communities



Synapse

DREAM CHALLENGES

e-consent

BRIDGE

Data sharing: with whom?

- Sharing with the research community.
- Sharing with collaborators.
- Sharing with oneself.

Barriers to sharing

- Culture / reluctance to share / weak sharing policies
- Disorganization and lack of mechanisms to facilitate sharing

Synapse: data management system
http://synapse.org

Synodos NF2 ★ | Search | Justin Guinney (Justin.Guinney) | Help

Synapse ID: syn2343195 **DOI:** doi:10.7303/syn2343195 **Storage Location:** Synapse Storage

Share | Annotations | Tools

Wiki | Files | Tables | Discussion | Docker beta

- Synodos NF2
  - Synodos Workflow
  - Data
  - Synodos NF2 Data Explorer (SyDE)
  - About us
  - Contact Us
  - FAQs

Edit Order <<

Applying systems biology to create tools and treatment paradigms for NF2-associated meningioma and schwannoma

Synodos for NF2 is a first-of-its-kind Neurofibromatosis research collaboration dedicated to defeating the rare genetic disorder neurofibromatosis type 2 (NF2). This unique consortium brings together a multidisciplinary team of scientists from twelve world-class labs at academic and medical centers of excellence, who have pledged to work closely together – sharing information, datasets, results and more – at every step in research development, with the goal of speeding up the drug discovery process. The end goal of this effort is to find new approaches to the diagnosis and treatment of two primary NF2-related tumors: schwannoma and meningioma. The expectation is to take the work from discovery to pre-clinical development, resulting in the development of an Investigational New Drug (IND) for NF2.

The consortium is funded by Children's Tumor Foundation(CTF)

Synodos NF2

Search | Justin Guinney (Justin.Guinney) | Help

Synapse ID: syn2343195 **DOI:** doi:10.7303/syn2343195 **Storage Location:** Synapse Storage

Share | Annotations | Tools

Wiki | Files | Tables | Discussion | Docker beta

Upload or Link to File | Add Folder

Name	Modified On	ID
Data	06/03/2016 11:27:01AM	syn6133098
DrugScreening	06/03/2016 11:27:11AM	syn6133099
Proteomics_Kinome	06/03/2016 11:28:50AM	syn6133104
baseline	06/27/2016 11:17:25AM	syn6178252
processed	06/27/2016 11:17:49AM	syn6178255
Synodos_kinome_baseline_LFQ_data.tsv	08/19/2016 07:20:35AM	syn6182623
Synodos_kinome_baseline_iTRAQ_data.tsv	08/19/2016 07:24:00AM	syn6182638
rawData	06/27/2016 11:17:43AM	syn6178254
treatment	06/27/2016 11:17:32AM	syn6178253
SyDE	06/09/2016 07:07:07PM	syn6140303

Created by Brian Bot (BrianMBot) on Tuesday, February 4, 2014 7:46 AM

Modified by Abhishek Pratap (apratap) on Wednesday, June 15, 2016 12:15 PM

Synodos Compounds ☆

Synapse ID: syn6138291 **Conditions for use:** None (change) report issue

Show simple search

SELECT * FROM syn6138291 Query

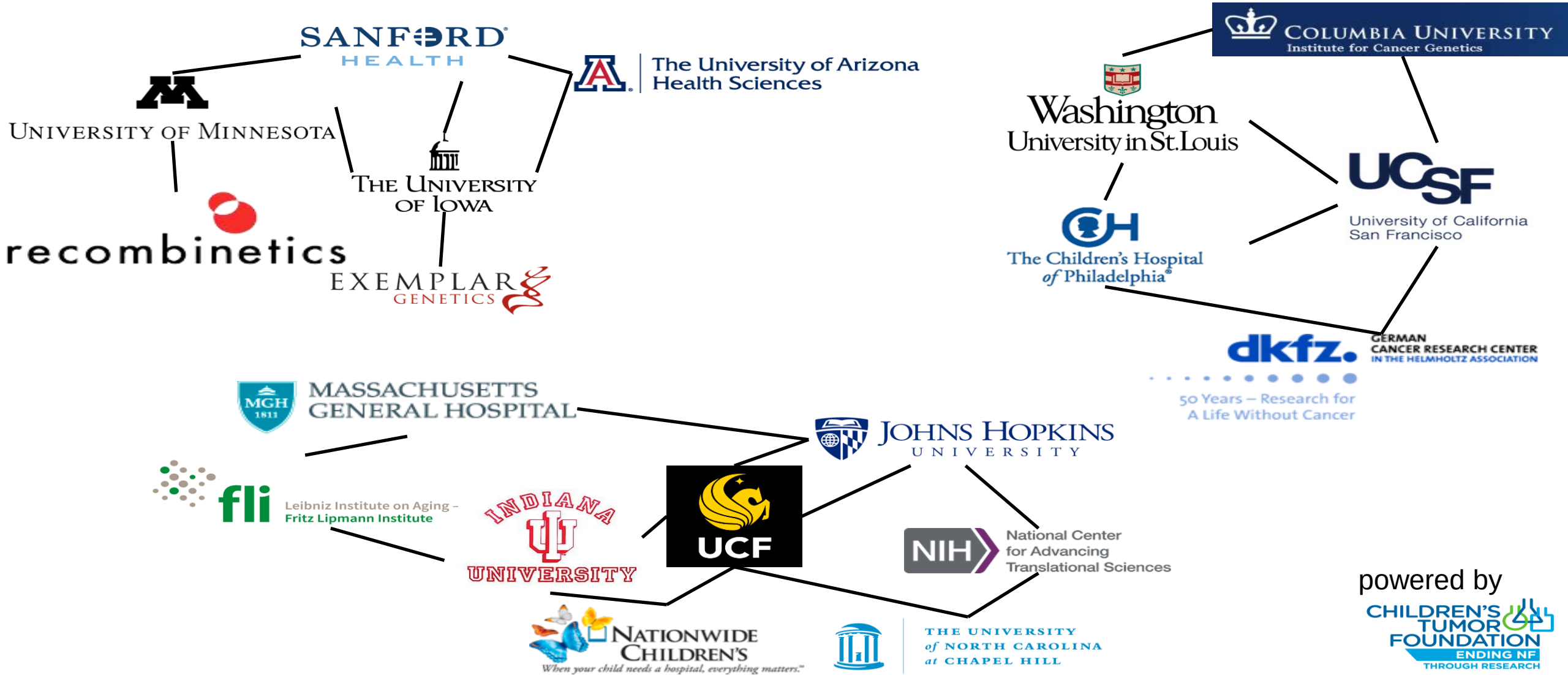
Product Name	Targets	Information	Pathway
Axitinib	c-Kit,VEGFR,PDGFR	multi-target inhibitor of VEGFR1, VEGFR2, VEGFR3, PDGFR_ and c-Kit with IC50 of 0.1 nM, 0.2 nM, 0.1-0.3 nM, 1.6 nM and 1.7 nM, respectively.	Protein Tyrosine Kinase
Selumetinib	MEK	potent, highly selective MEK1 inhibitor with IC50 of 14 nM, also inhibits ERK1/2 phosphorylation with IC50 of 10 nM, no inhibition to p38_, MKK6, EGFR, ErbB2, ERK2, B-Raf, etc. Phase 1/2.	MAPK
Bortezomib	Proteasome	potent 20S proteasome inhibitor with Ki of 0.6 nM.	Proteases
Lapatinib	HER2,EGFR	potent EGFR and ErbB2 inhibitor with IC50 of 10.8 and 9.2 nM, respectively.	Protein Tyrosine Kinase
Panobinostat	HDAC	novel broad-spectrum HDAC inhibitor with IC50 of 5 nM. Phase 3.	Epigenetics
Perifosine	AKT	novel Akt inhibitor with IC50 of 4.7 _M, targets pleckstrin homology domain of Akt. Phase 2.	PI3K/Akt/mTOR
Vorinostat	Autophagy,HDAC	HDAC inhibitor with IC50 of ~10 nM.	Epigenetics
GDC-0941	PI3K	potent inhibitor of PI3K_/_ with IC50 of 3 nM, with modest selectivity against p110_ (11-fold) and p110_ (25-fold).	PI3K/Akt/mTOR
Vismodegib	Hedgehog/SMO	potent, novel and specific hedgehog inhibitor with IC50 of 3 nM and also inhibits P-gp with IC50 of 3.0 _M.	Stem Cells & Wnt
OSU-03012 (AR-12)	PDK-1	potent inhibitor of recombinant PDK-1 with IC50 of 5 _M and 2-fold increase in potency over OSU-02067.	PI3K/Akt/mTOR
Everolimus	mTOR	an mTOR inhibitor of FKBP12 with IC50 of 1.6-2.4 nM.	PI3K/Akt/mTOR
Ganetespib	HSP (e.g. HSP90)	HSP90 inhibitor with IC50 of 4 nM in OSA 8 cells, induces apoptosis of OSA cells while normal osteoblasts are not affected;	Cytoskeletal

Synapse

- Dash boarding for meta-data
- Access controls for sharing
- Governance facilities and auditing
- Docker store for methods and pipelines
- Embedding of visualizations and tools

CTF & Sage

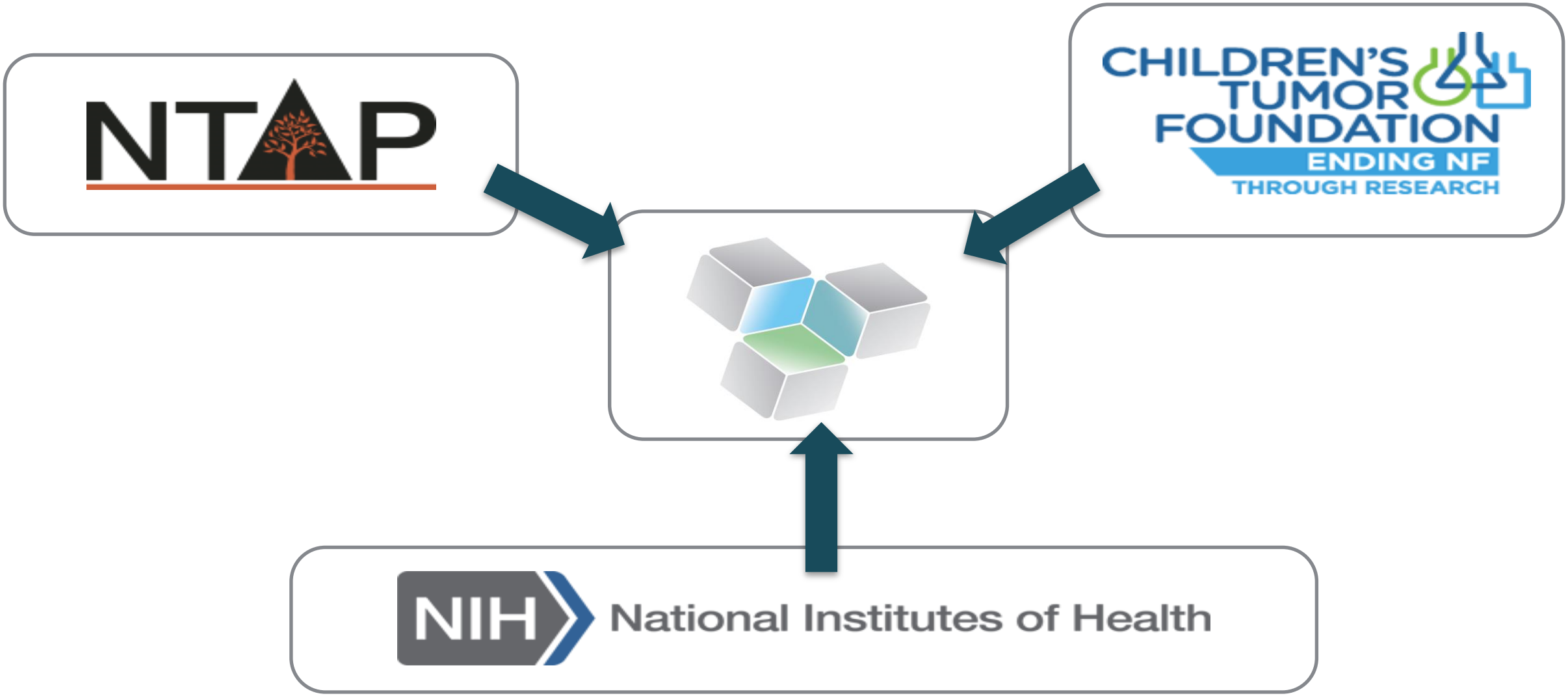
Building **networks** among CTF researchers



powered by

Children's Tumor Foundation
Ending NF through Research

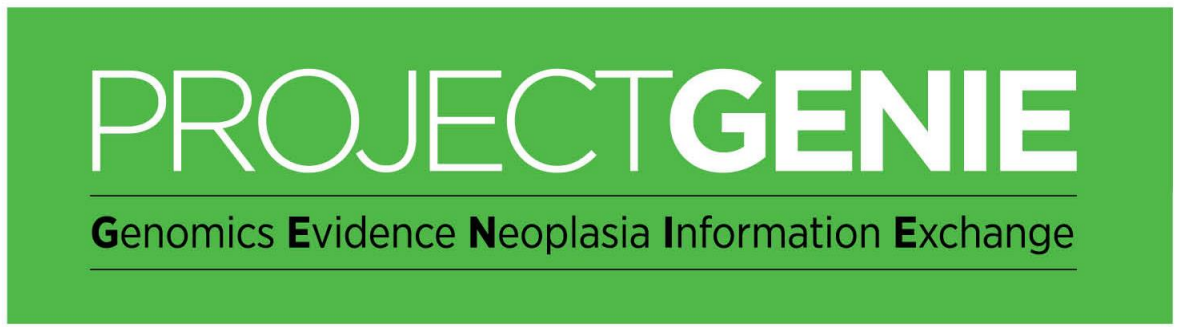
Building a **network** for the **NF community**



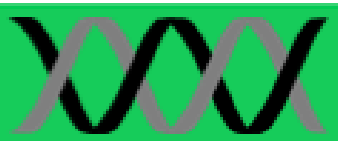
Data sharing vignettes

1. AACR Project GENIE
2. DREAM Challenges


AACR
American Association
for Cancer Research
FINDING CURES TOGETHER®


PROJECTGENIE
Genomics Evidence Neoplasia Information Exchange

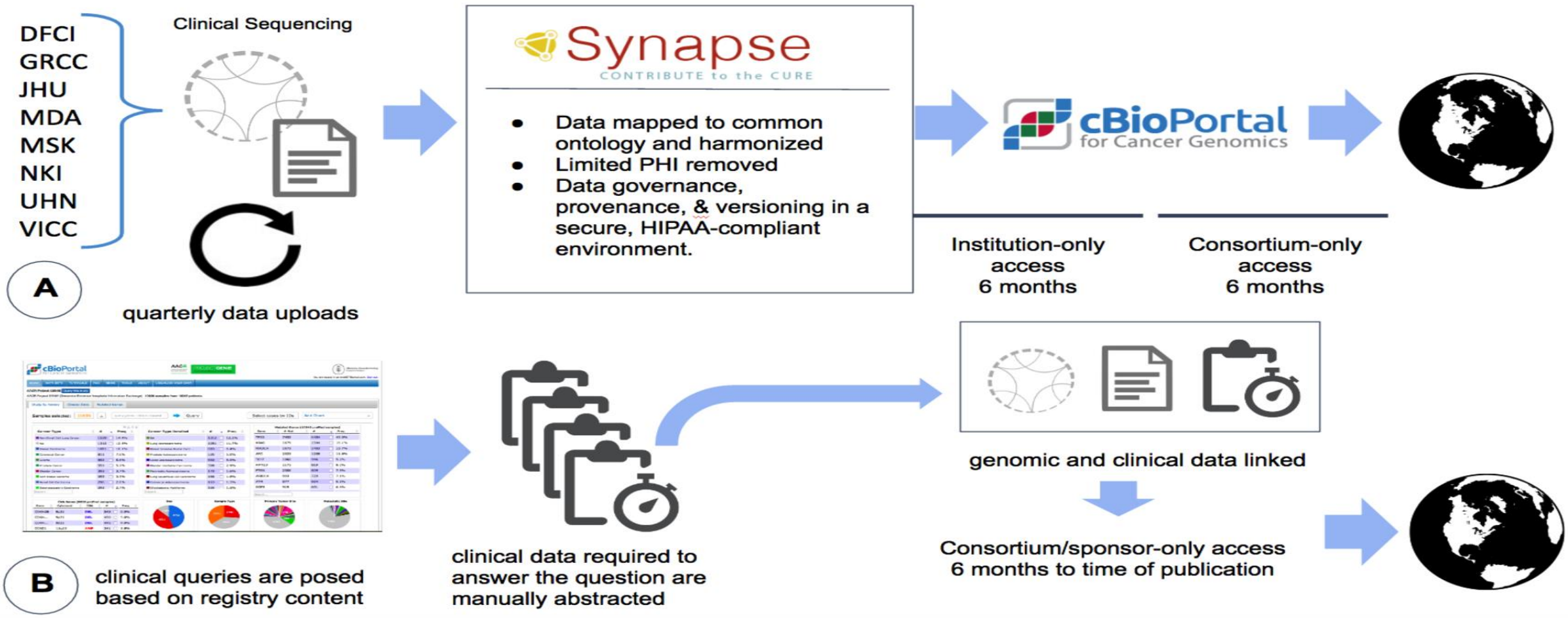
GENIE: Motivation



CANCER is largely a disease of acquired genetic mutations

~13	Genes are mutated in >5 % of patients
~78	Genes are mutated in 1-2 % of patients
100s	Genes are mutated in <1% of patients
	Most tumors have multiple mutations occurring even less frequently
	No single institution is likely to have enough data to make informed clinical decisions
	SHARING DATA IS THE SOLUTION

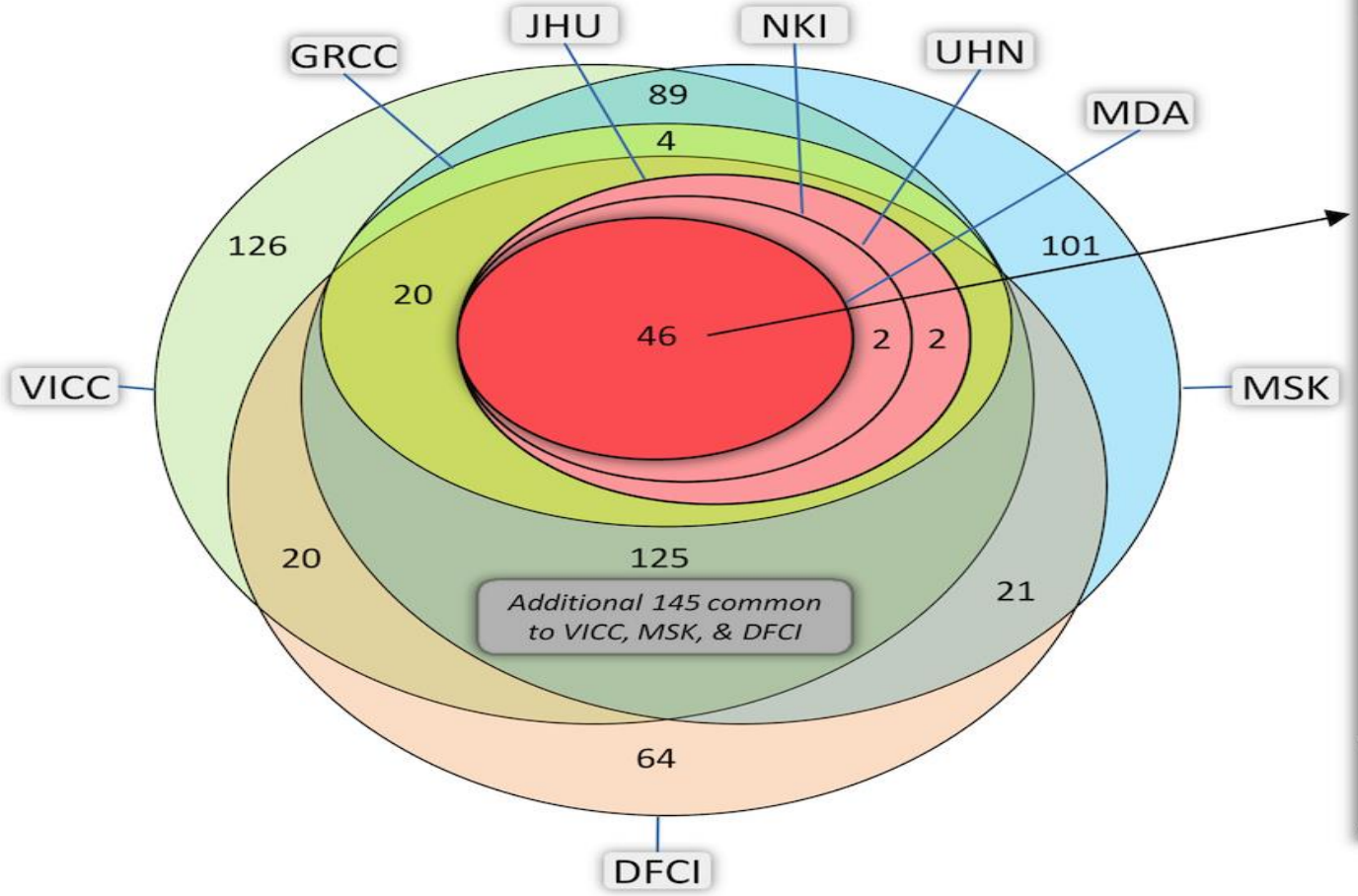
GENIE Consortium



First Data Release

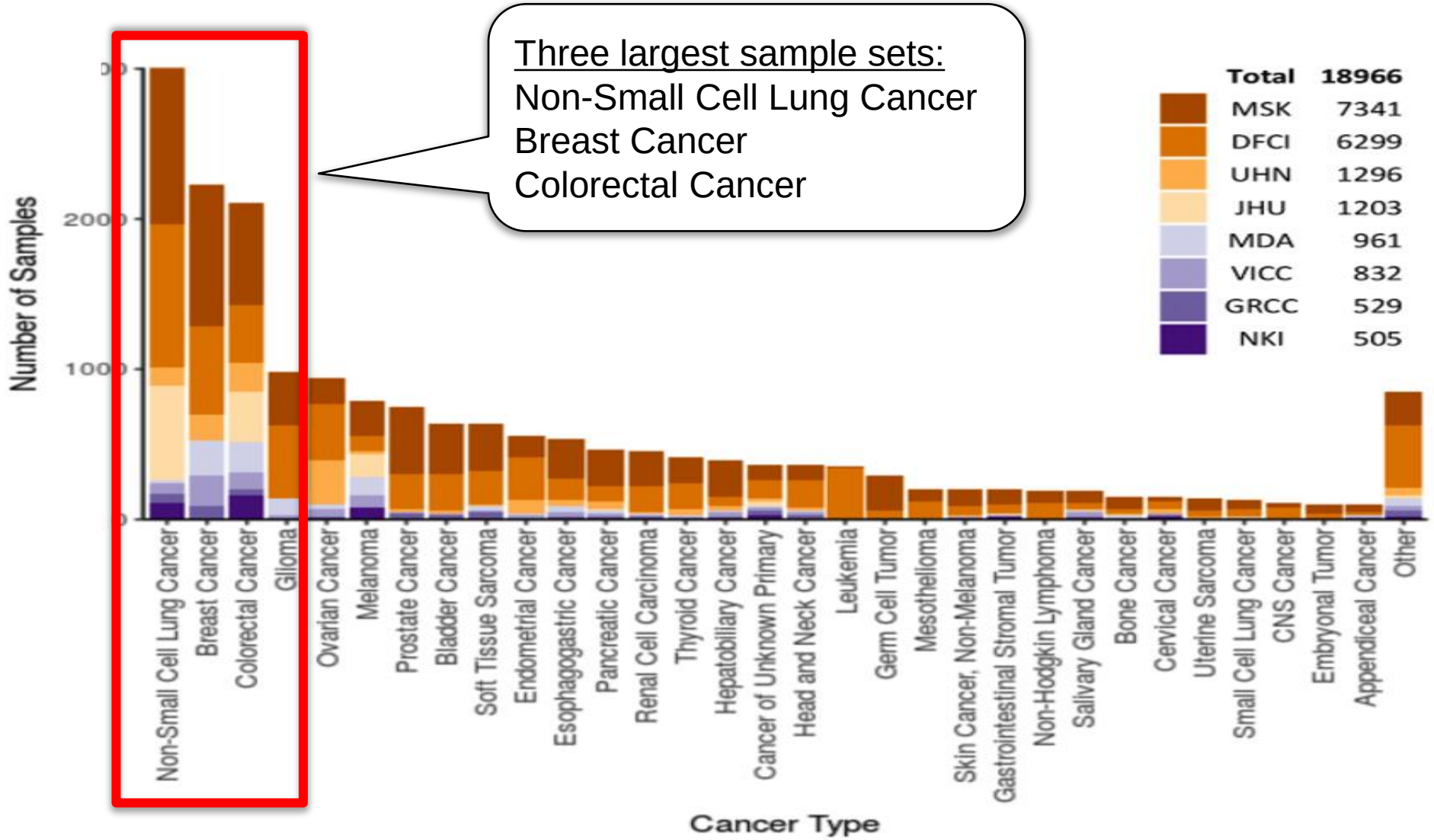
- Released January 5, 2017
- **~19,000 samples**
- Includes genomic data plus **Tier 1 Clinical Data**: cancer type, primary v. metastatic sample, gender, race, age at sequencing, etc.
- Data is now available at:
  - Sage Synapse Platform: http://synapse.org/genie
  - cBioPortal for Cancer Genomics: http://www.cbioportal.org/genie/
- Users are required to agree to **terms of access** at each site.

Multiple Gene Panels

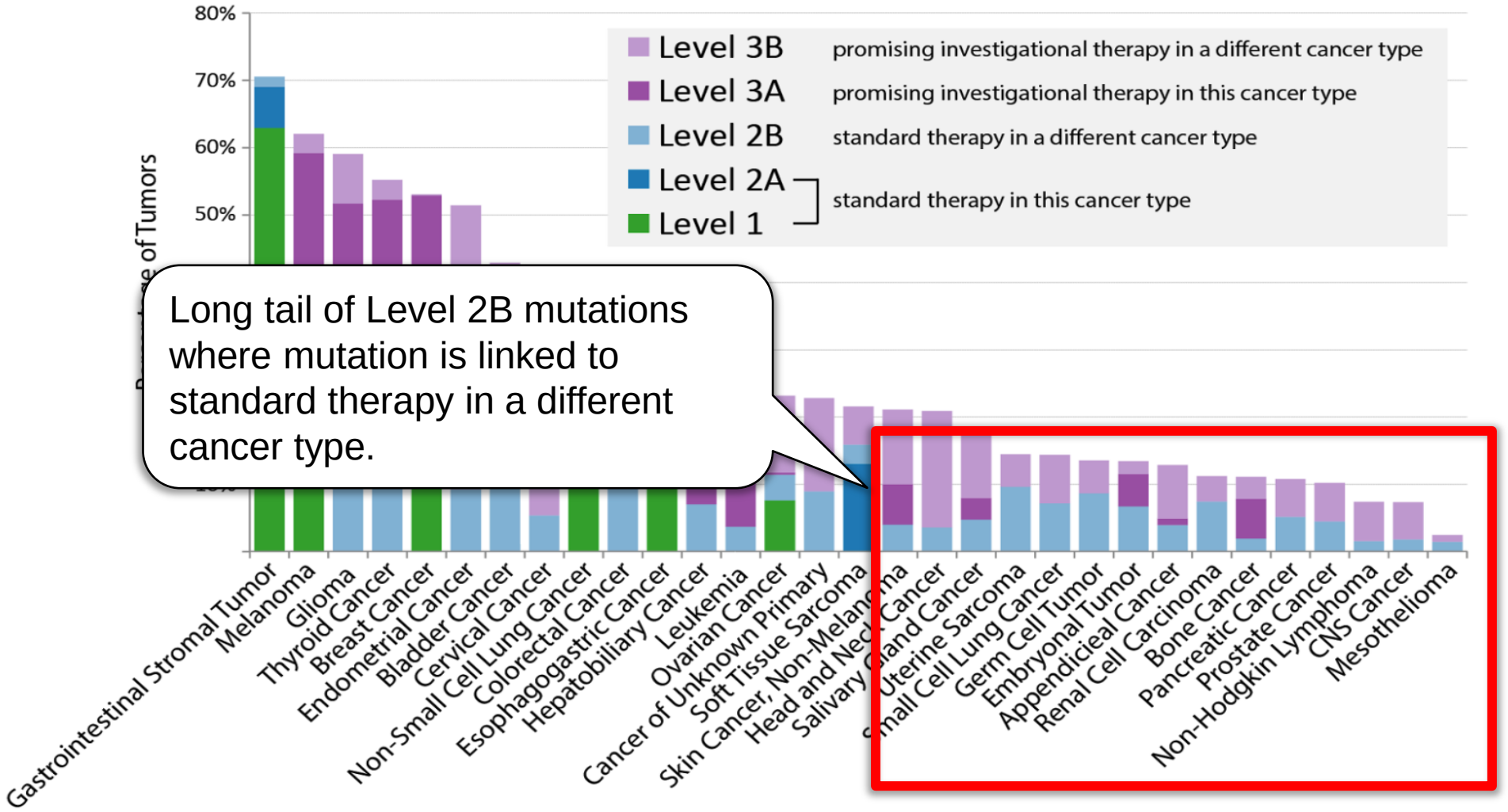


Core 44 Genes		
ABL1	FLT3	PIK3CA
AKT1	GNAS	PTEN
ALK	HRAS	PTPN11
APC	IDH1	RB1
ATM	JAK2	RET
BRAF	JAK3	SMAD4
CDH1	KDR	SMARCB1
CDKN2A	KIT	SMO
CTNNB1	KRAS	SRC
EGFR	MET	STK11
ERBB2	MLH1	TP53
ERBB4	MPL	VHL
FBXW7	NOTCH1	
FGFR1	NPM1	
FGFR2	NRAS	
FGFR3	PDGFRA	
	HNF1A CSF1A	

GENIE Landscape



Landscape of Clinical Actionability



GENIE’s future

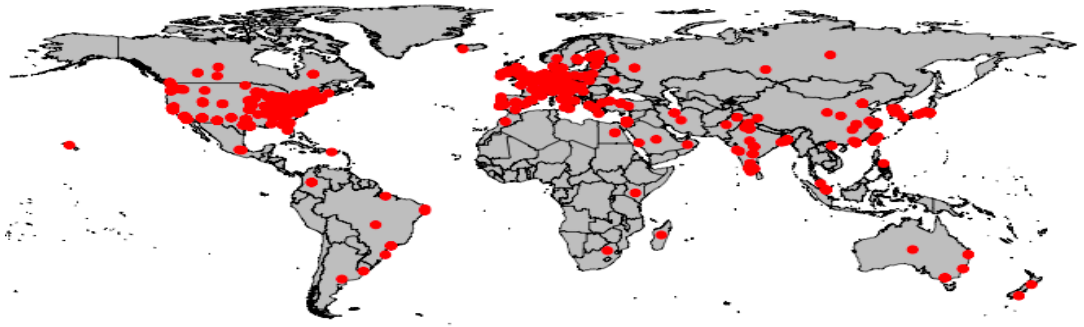
- 2nd release scheduled for end of 2017
- Expect to double current database size: over 40k samples!!
- More extensive clinical annotation, including patient outcomes, staging, and treatments
- **In process of moving GENIE data to GDC!**


DREAM
CHALLENGES
powered by Sage Bionetworks

DREAM CHALLENGES

powered by Sage Bionetworks

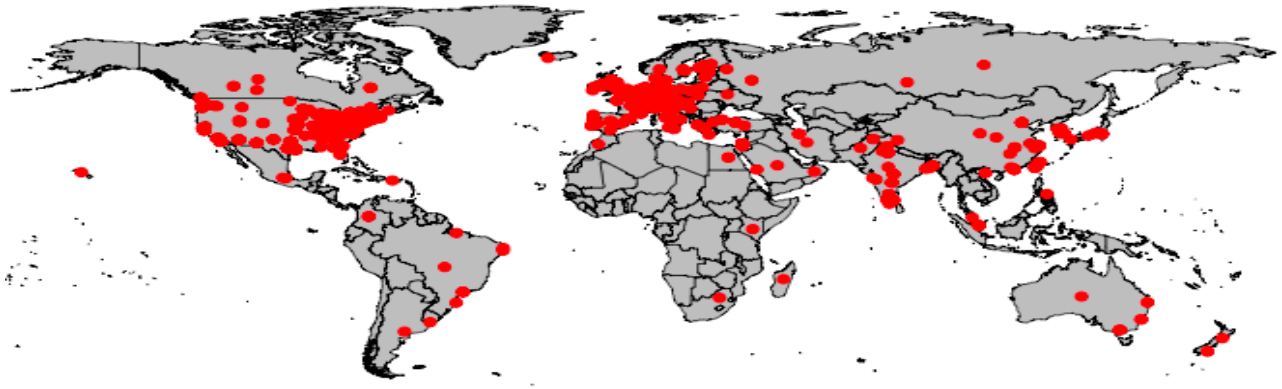
- A crowdsourcing effort that poses quantitative challenges in biomedicine.
- Our mission is
  - to contribute to the solution of important biomedical problems
  - to foster collaboration between research groups
  - to democratize data
  - to accelerate research
  - to objectively assess and benchmark algorithms



DREAM CHALLENGES

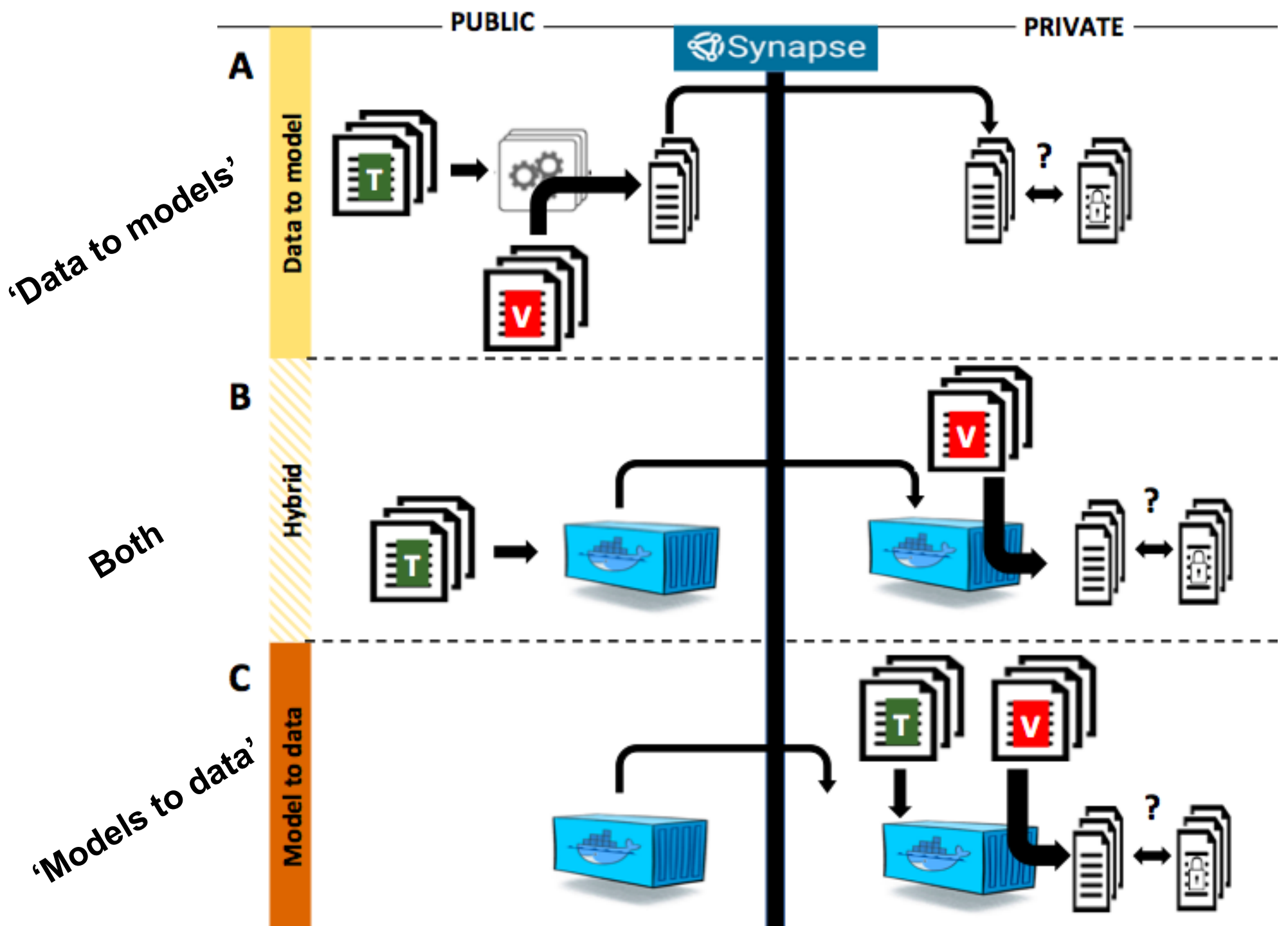
powered by Sage Bionetworks

- Over last 10 years, we have run Challenges on:
  - Breast cancer prognosis
  - Prostate cancer prognosis
  - Somatic variant detection
  - Drug sensitivity prediction
  - Drug combination prediction
  - Drug toxicity prediction
  - ALS
  - Alzheimer’s
  - Many others…



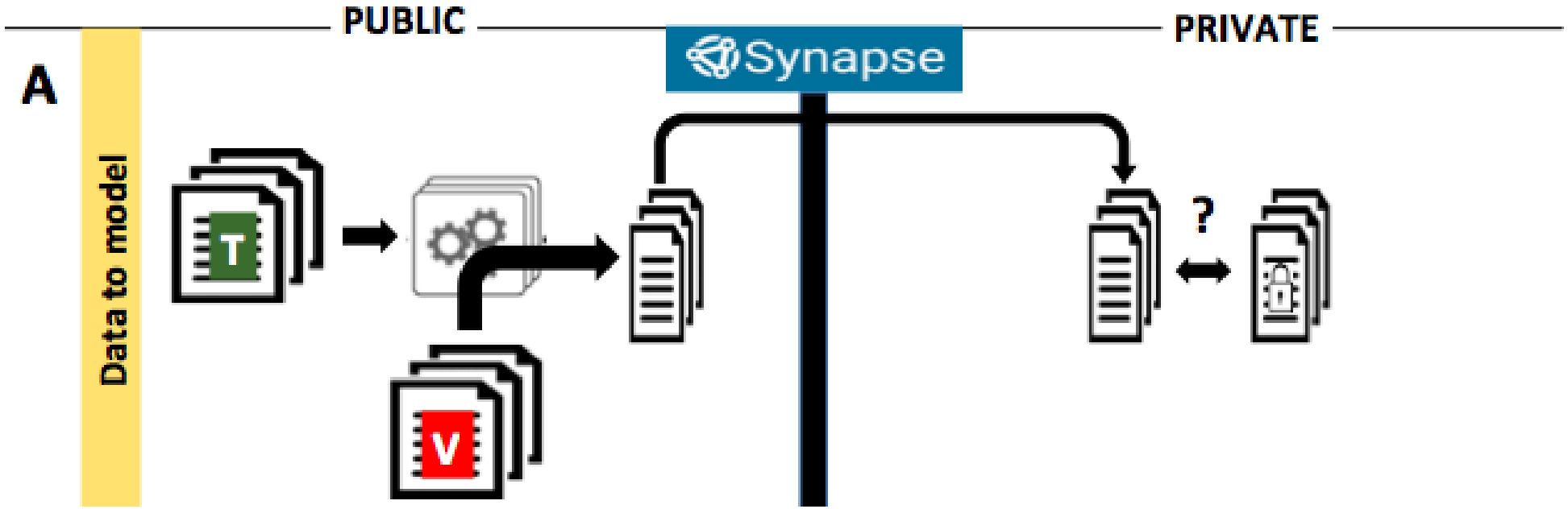
DREAM CHALLENGES
powered by Sage Bionetworks

Models of sharing

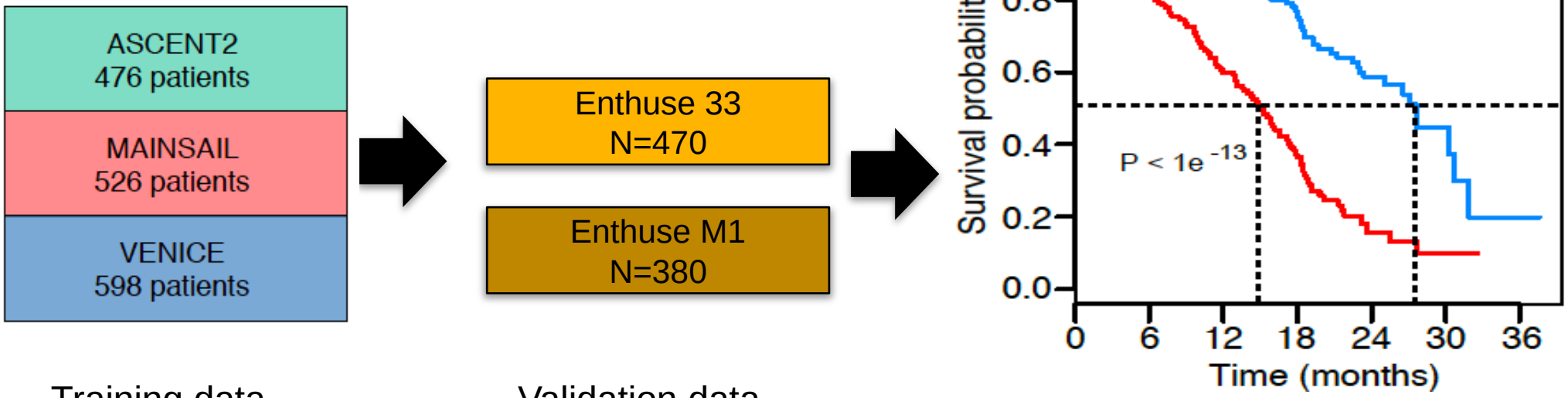


DREAM CHALLENGES
powered by Sage Bionetworks

‘Data to model(ers)’



Goal: Predict overall survival in patients with metastatic castration resistant prostate cancer



Training data

Validation data

Guinney, et al, Lancet Oncology, 2017

DREAM CHALLENGES
powered by Sage Bionetworks

How can we improve model reproducibility?

How can we improve utilization of restricted data?

docker

Cheap and scalable data storage and computing

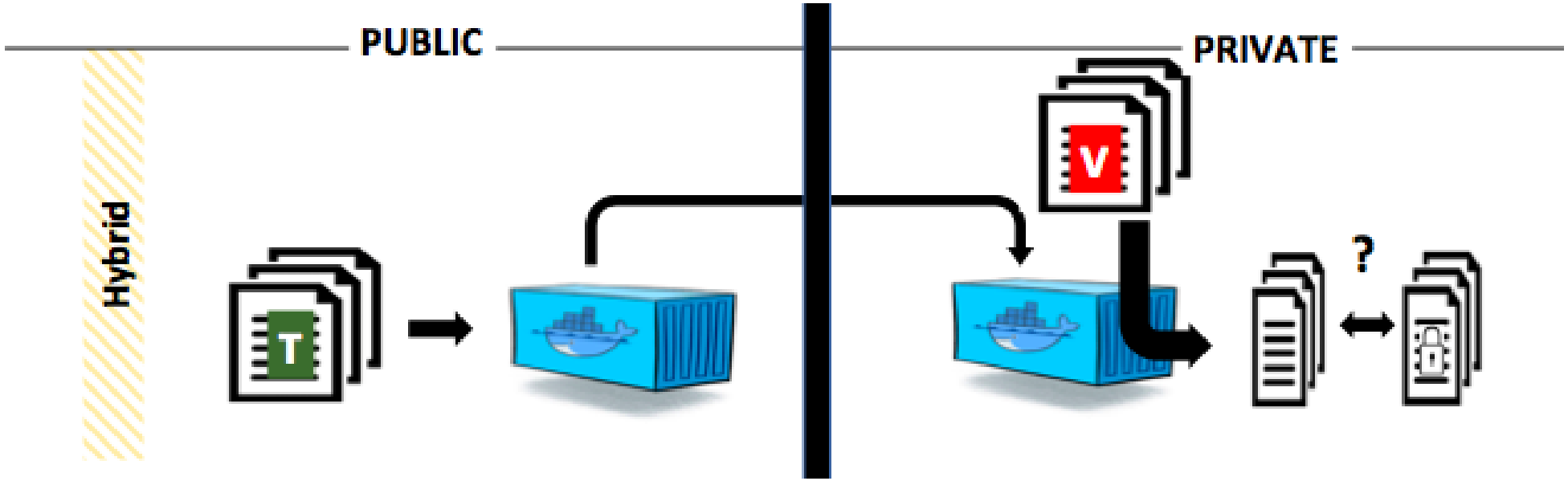
amazon web services

Google Cloud Platform

Microsoft Azure

Virtualization and container technologies: platform agnostic application and model portability

Hybrid: ‘Data to model(ers)’ ‘Models to data’



Goal: Improve identification of “high-risk” patients with newly diagnosed multiple myeloma

Public

MM RF MULTIPLE MYELOMA Research Foundation

GEO Gene Expression Omnibus

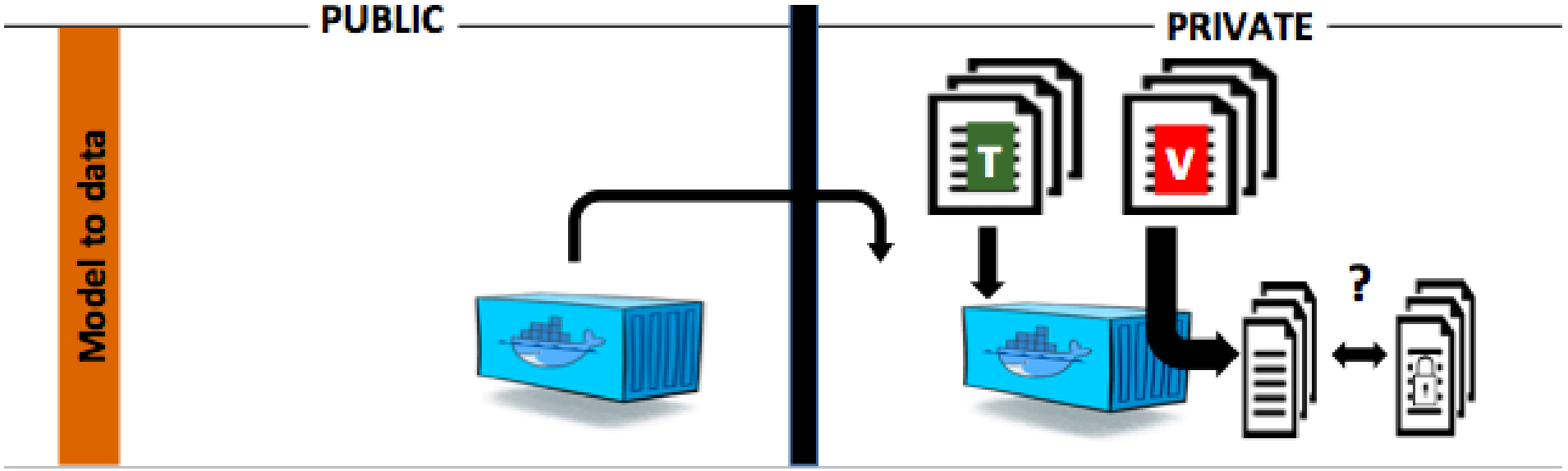
Private

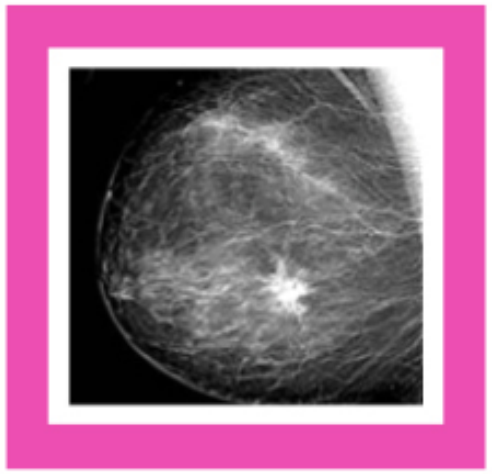
UAMS UNIVERSITY OF ARKANSAS FOR MEDICAL SCIENCES

DANA-FARBER CANCER INSTITUTE

MOFFITT CANCER CENTER

‘Models to data’





The Digital Mammography DREAM Challenge

Build a model to help reduce the recall rate for breast cancer screening

Learn more & register to participate here: www.synapse.org/Digital_Mammography_DREAM_Challenge

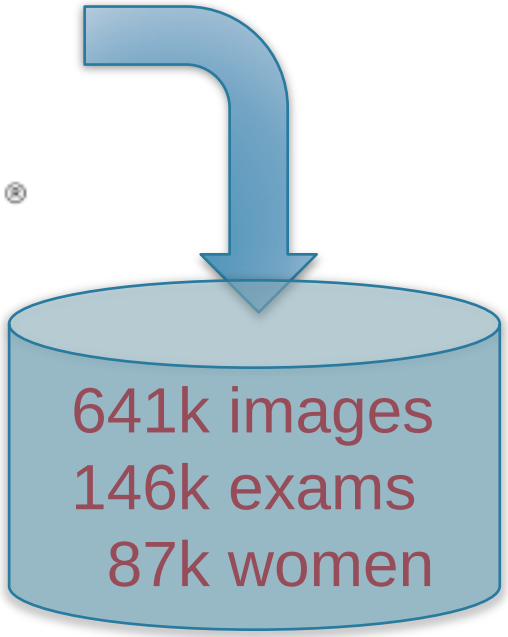
Funded by tjaf

Enabled by DREAM CHALLENGES, Sage, Coding4Cancer, GroupHealth Research Institute, IBM, BCSC, amazon web services, Icahn School of Medicine at Mount Sinai, Seattle Cancer Care Alliance, RADISH

Goal: Improve accuracy of digital mammograms screening by classifying images as low or high risk for breast cancer

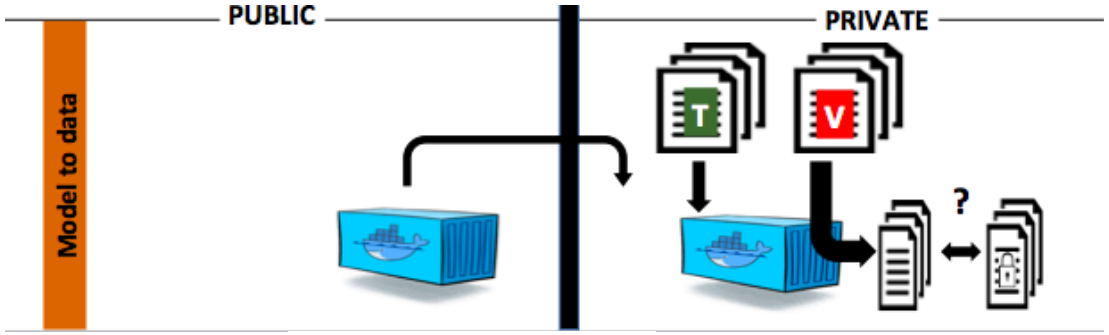
1 in 10 women are falsely diagnosed with breast cancer.

GroupHealth



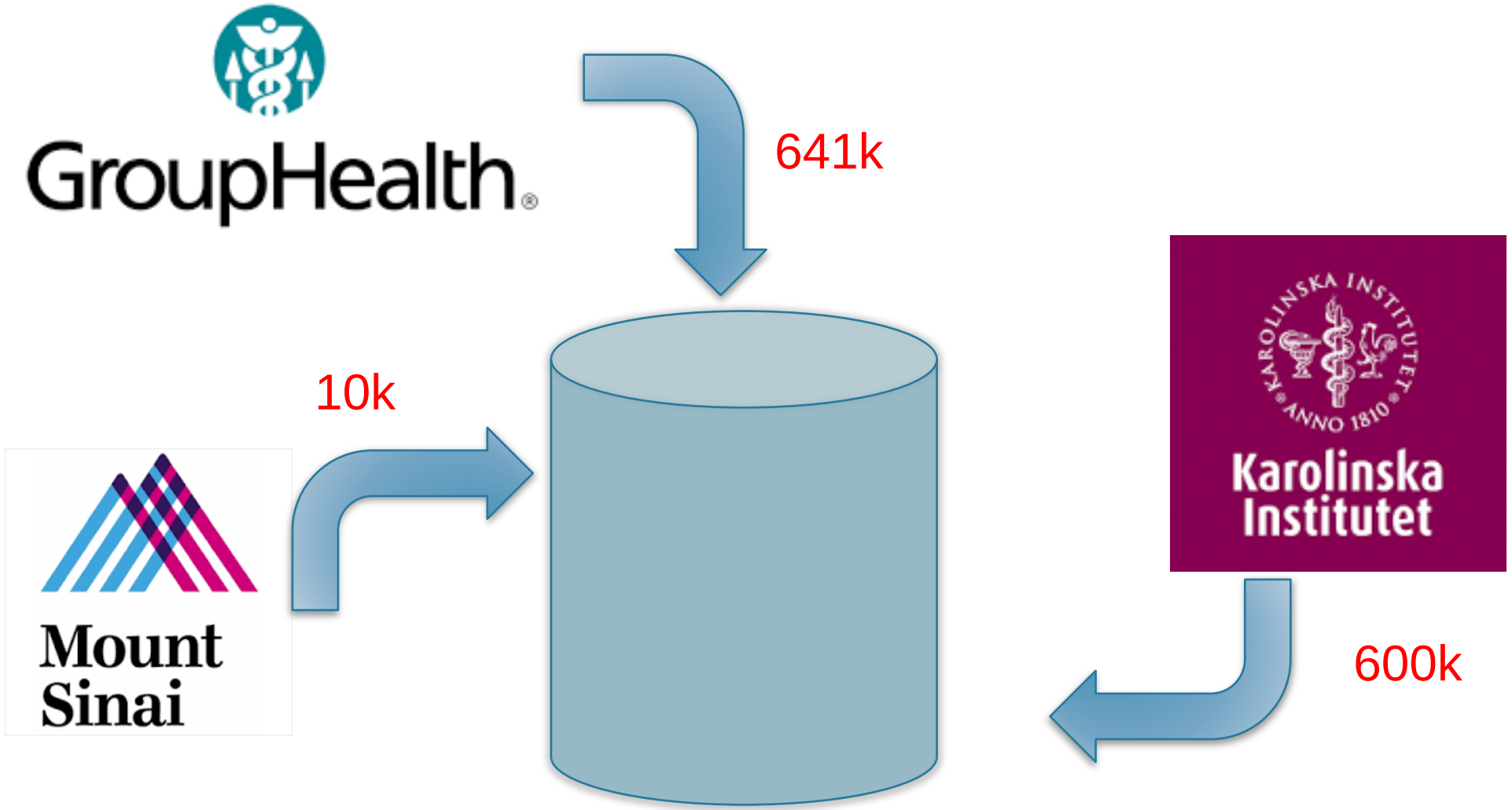
But…

Images not allowed to be directly accessed by participants.



SOFTLAYER an IBM Company

amazon web services


GroupHealth
641k
10k
Mount Sinai
KAROLINSKA INSTITUTET
ANNO 1810
Karolinska Institutet
600k

Key statistics: DM Challenge

~ 1k participants
~ 10k model submissions
~ 1k TB (1 Petabyte) data usage
~ 874k CPU-hours

Challenge summary

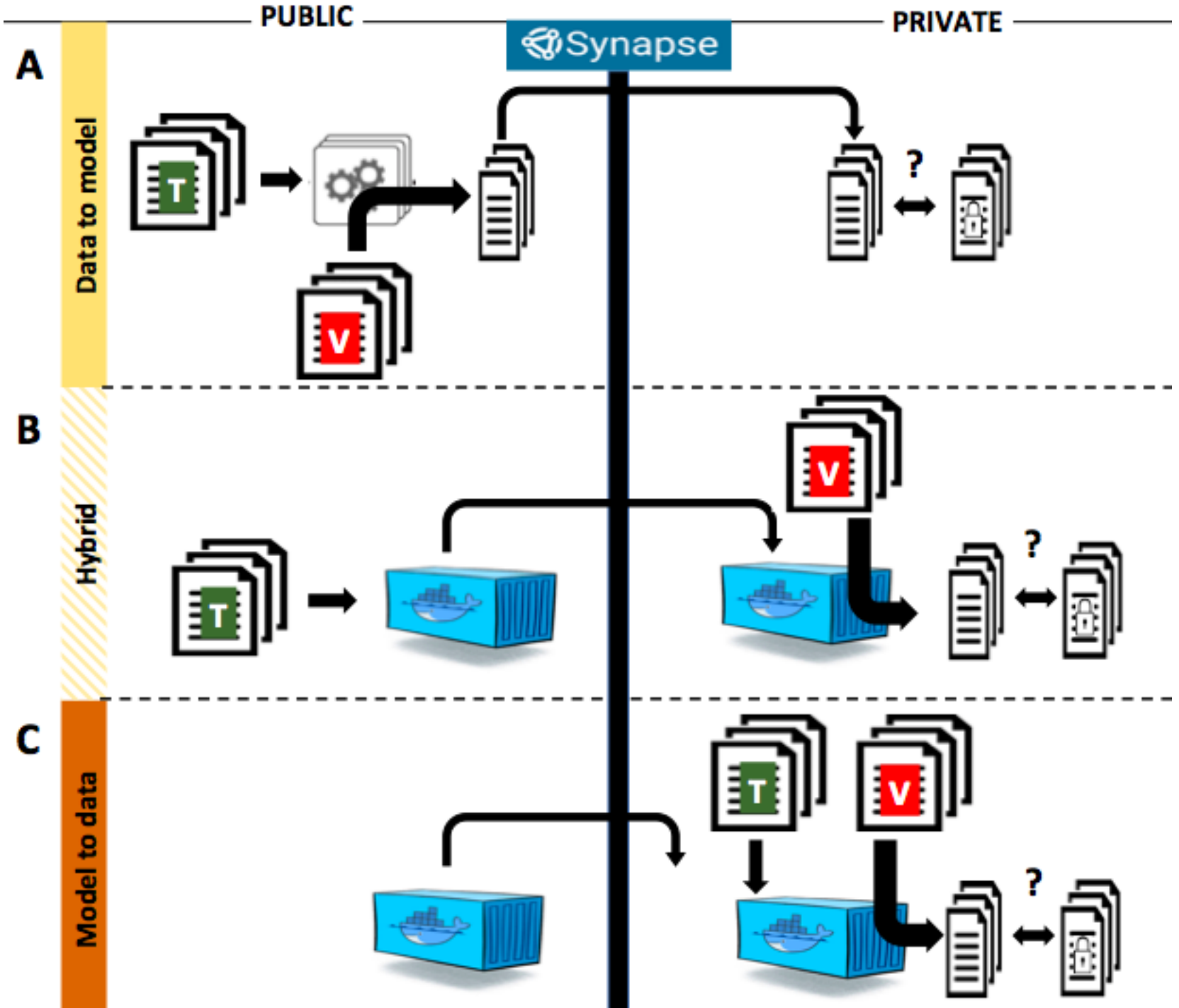
- Currently, in 3rd round of leaderboard phase
- Validation phase begins in April
- Currently, top models are performing as well as a radiologist (sensitivity + specificity)

Challenges

Prostate Cancer,
Drug Combination,
Toxicogenomics

Multiple Myeloma,
RNA fusion detection

Digital Mammography



Data sharing: what can you do?

- Play an active role in setting data sharing policies.
- Set clear guidelines and expectations on what is meant by data sharing.
- Put in place mechanisms for oversight and enforcement of data sharing practices.

Thank you


Sage
BIONETWORKS


DREAM
CHALLENGES
powered by Sage Bionetworks


CHILDREN'S
TUMOR
FOUNDATION
ENDING NF
THROUGH RESEARCH


NTAP


ljaf
laura and john arnold foundation®


NATIONAL
CANCER
INSTITUTE


MM
RF
SM
MULTIPLE MYELOMA
Research Foundation


AACR
American Association
for Cancer Research