DANA-FARBER
CANCER INSTITUTE

Knowledge Systems @ DFCI

Ethan Cerami, Ph.D.
Director, Knowledge Systems Group

James Lindsay, Ph.D.
Associate Director, Knowledge Systems Group

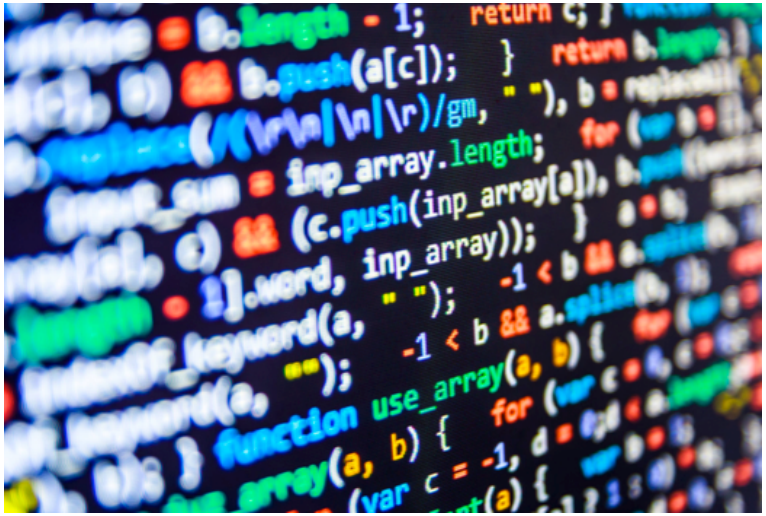
cBioCenter @ DFCI

DFCI Knowledge Systems Group, cBioCenter

Applied genomics software and data science group.

Building genomic software to enable:

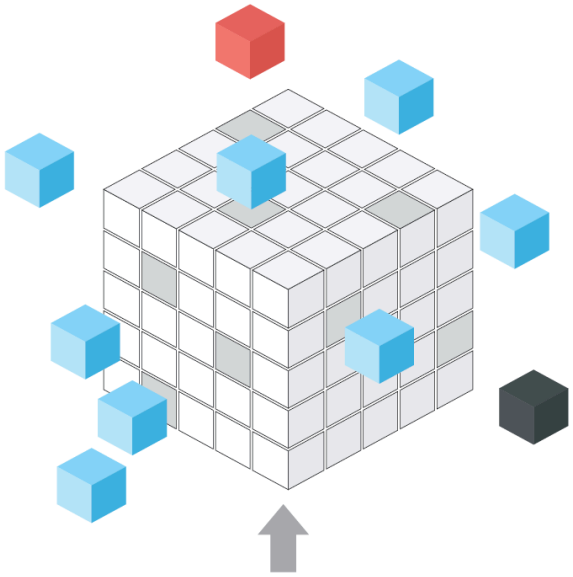
- cancer genomics research
- clinical application of genomic data.



Part of the new cBioCenter @ DFCI (Head: Chris Sander)

Enterprise Genomics and Data Science @ DFCI

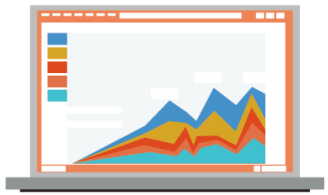
open source

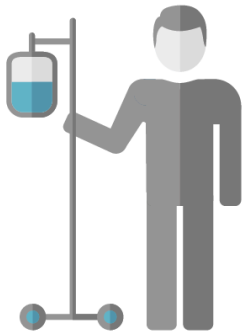


Data Sharing Initiatives

Data Mining Platform

Drive Genomic Discoveries





Clinical Decision Support Platform

Improve Patient Care

Data acquisition

Data utilization

Highlighted projects

Data utilization

- cBioPortal
- MatchMiner
- Insight engine

Data acquisition

- cBioOne
- AACR GENIE
- Intel CCC

DANA-FARBER CANCER INSTITUTE

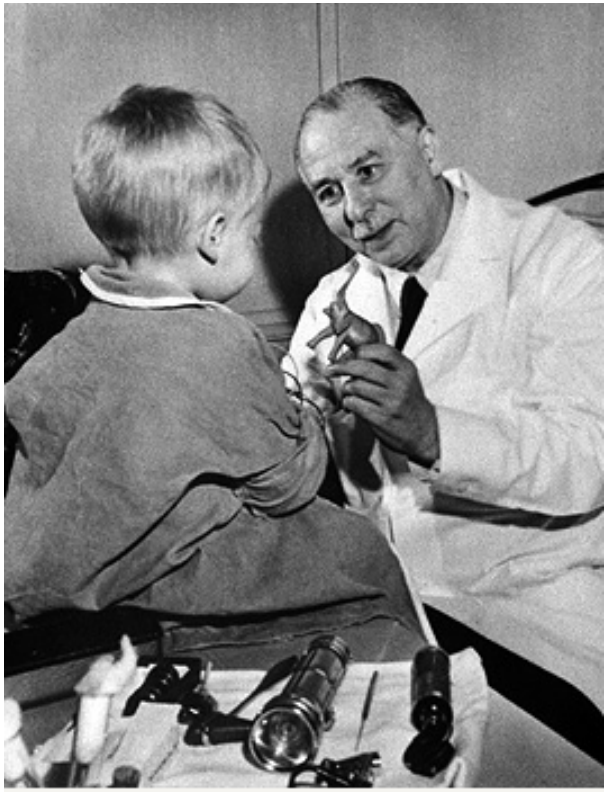
about

- Founded in 1947 by Sidney Farber

- > 4,500 employees
- > 450,000 patient visits / yr

- 900 clinical trials
  - > 500 treatment trials

- percent_research <- 0.5
- percent_clinical <- 0.5



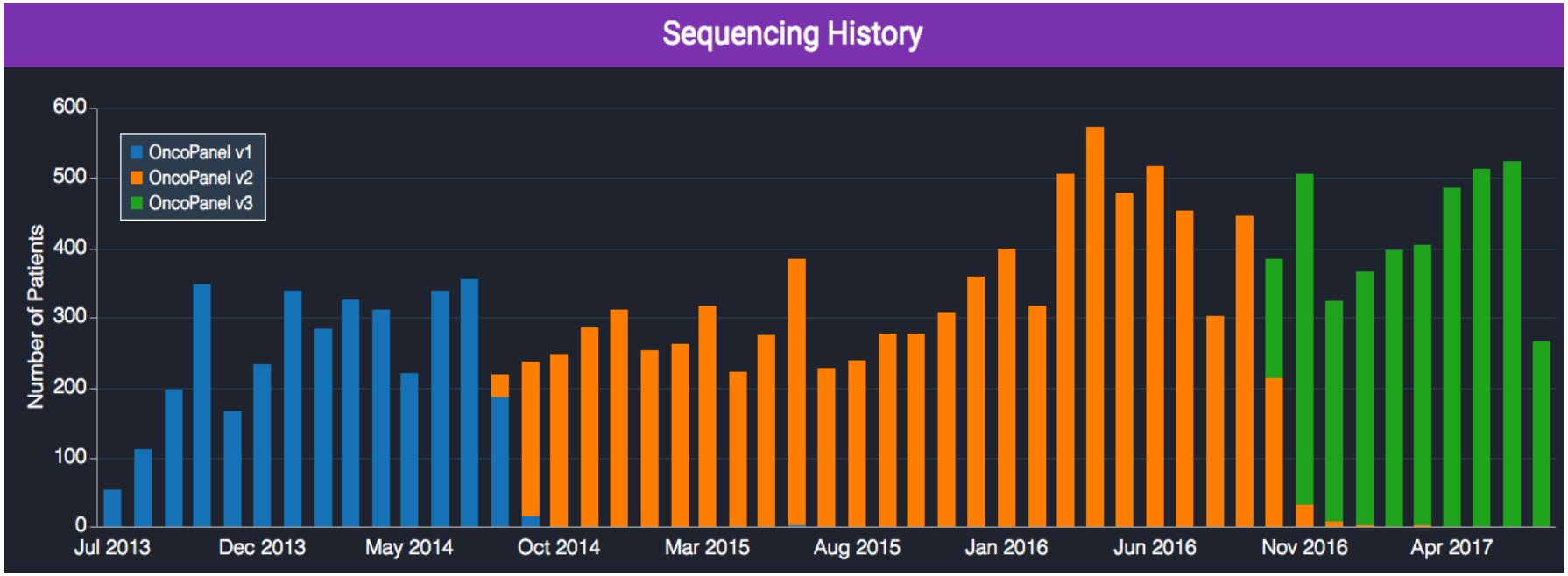
Sidney Farber, MD, with a young patient

Enterprise genomics: PROFILE

- 15,927 patients sequenced
- Targeted DNAseq
- 447 genes / regions targeted
- CLIA-certified

DANA-FARBER/BRIGHAM AND WOMEN'S CANCER CENTER

Boston Children's Hospital
Until every child is well

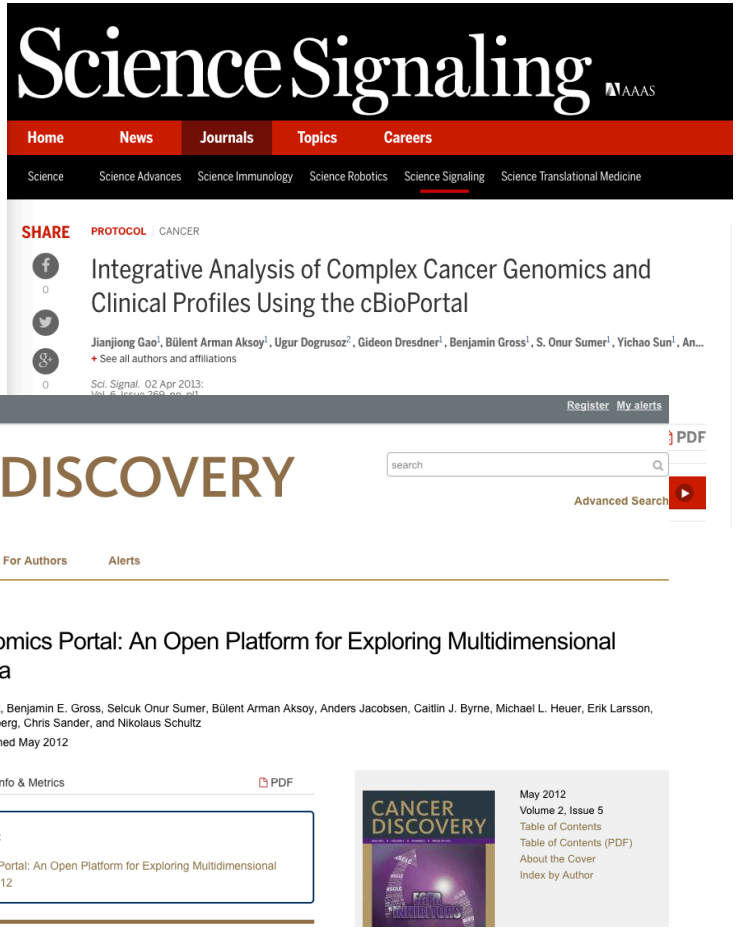


cBioPortal, Insight Engine, MatchMiner

DATA UTILIZATION

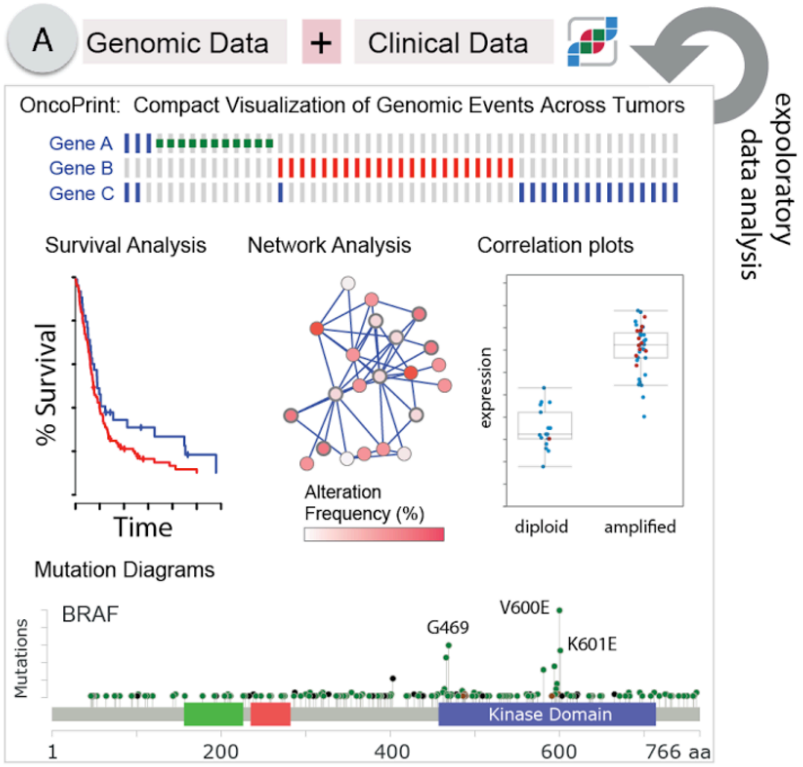
DANA-FARBER
CANCER INSTITUTE

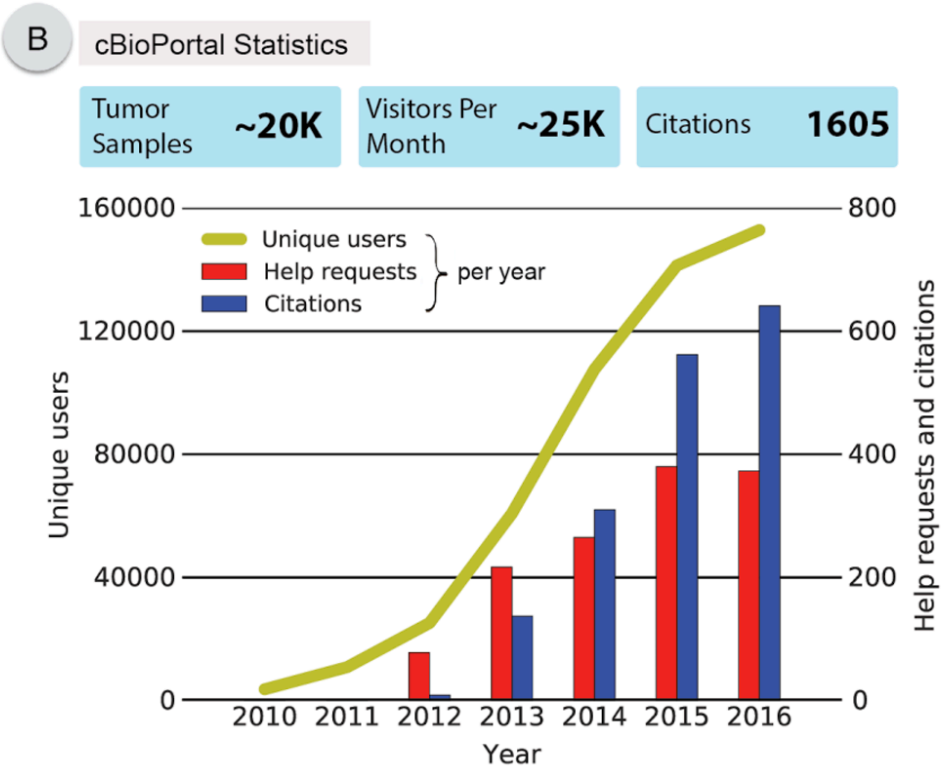
cBioPortal



cBioPortal for Cancer Genomics

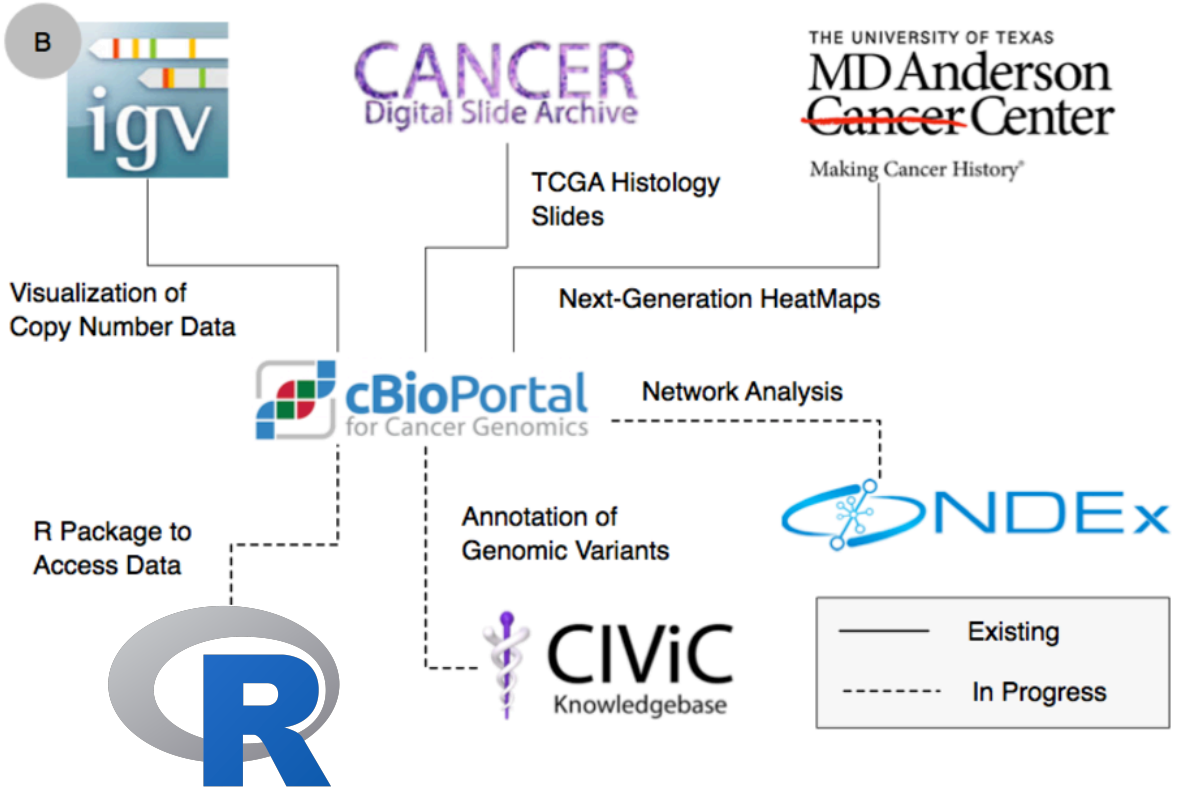
http://cbioportal.org





cBioPortal for Cancer Genomics

http://cbioportal.org



R package

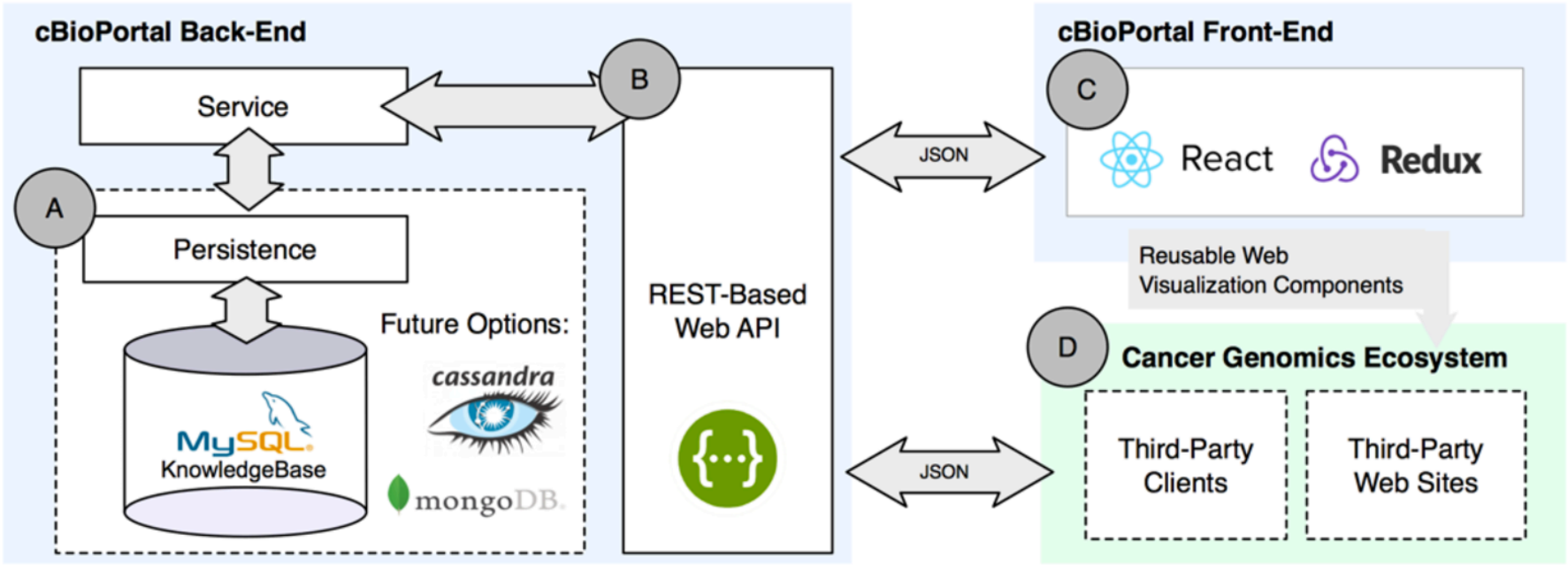
Installation

1. The CDGS-R package currently **only works with R Version 2.12 or higher**.
2. Then install the cgds-R package from within R: `install.packages('cgdsr')`

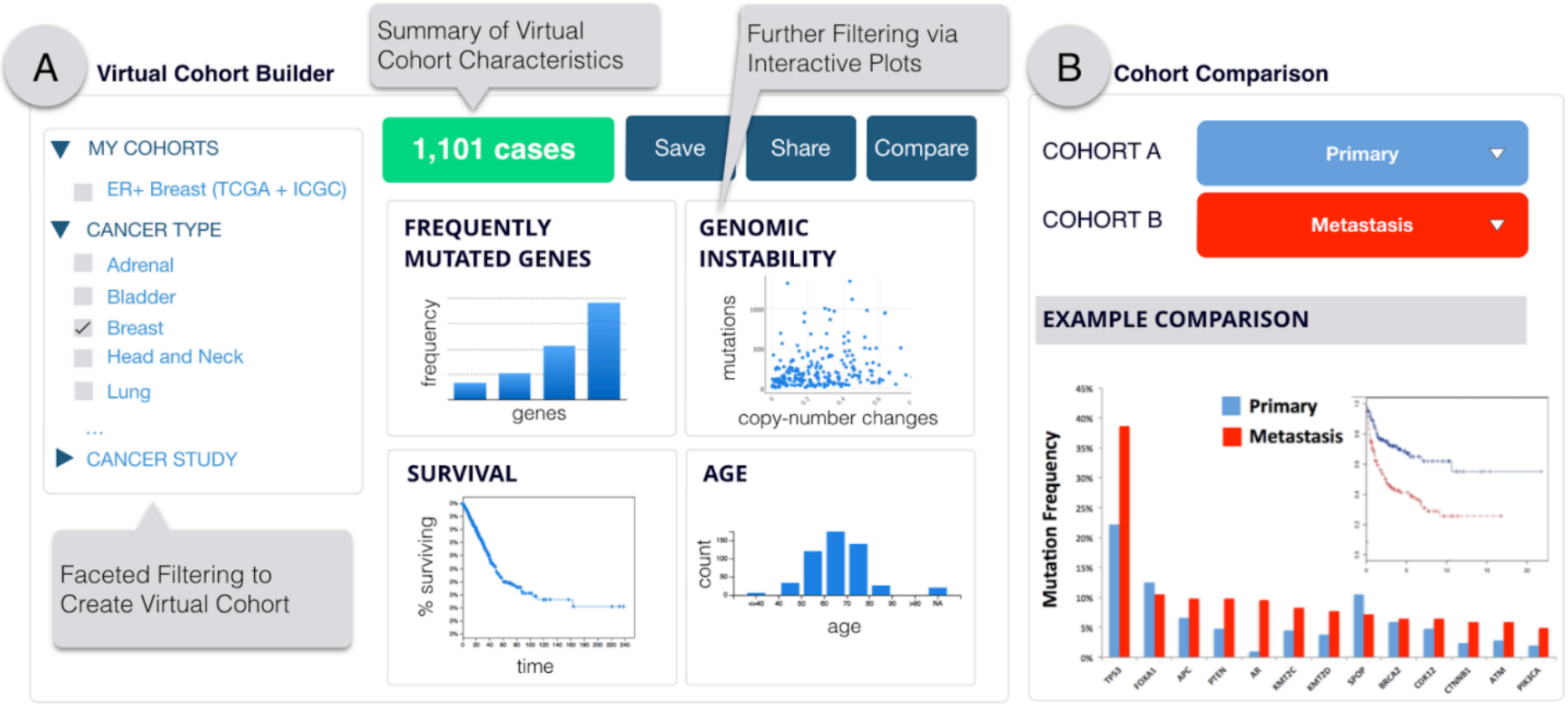
Example usage

```
# Create CGDS object
mycgds = CGDS("http://www.cbioportal.org/public-portal/")

test(mycgds)

# Get list of cancer studies at server
getCancerStudies(mycgds)

# Get available case lists (collection of samples) for a given cancer study
mycancerstudy = getCancerStudies(mycgds)[2,1]
mycaselist = getCaseLists(mycgds,mycancerstudy)[1,1]

# Get available genetic profiles
mygeneticprofile = getGeneticProfiles(mycgds,mycancerstudy)[4,1]

# Get data slices for a specified list of genes, genetic profile and case list
getProfileData(mycgds,c('BRCA1','BRCA2'),mygeneticprofile,mycaselist)

# Get clinical data for the case list
myclinicaldata = getClinicalData(mycgds,mycaselist)

# documentation
help('cgdsr')
help('CGDS')
```

New Architecture

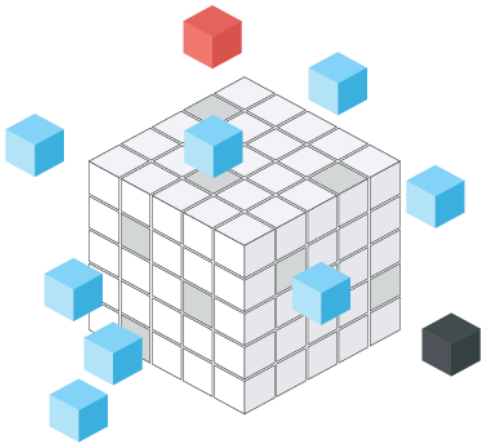


Virtual Cohorts



DFCI Insight Engine

http://insight.dfci.harvard.edu

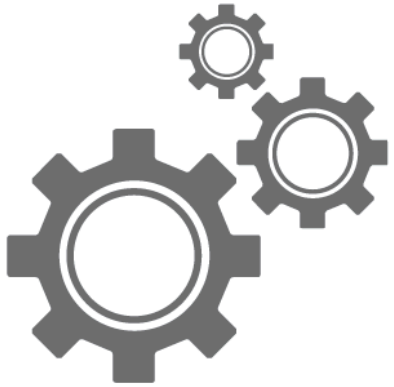


Profile + RHP Data

De-identified clinical data

GENIE Data

Weekly Updates



Reproducible Pipelines + Jupyter Notebooks

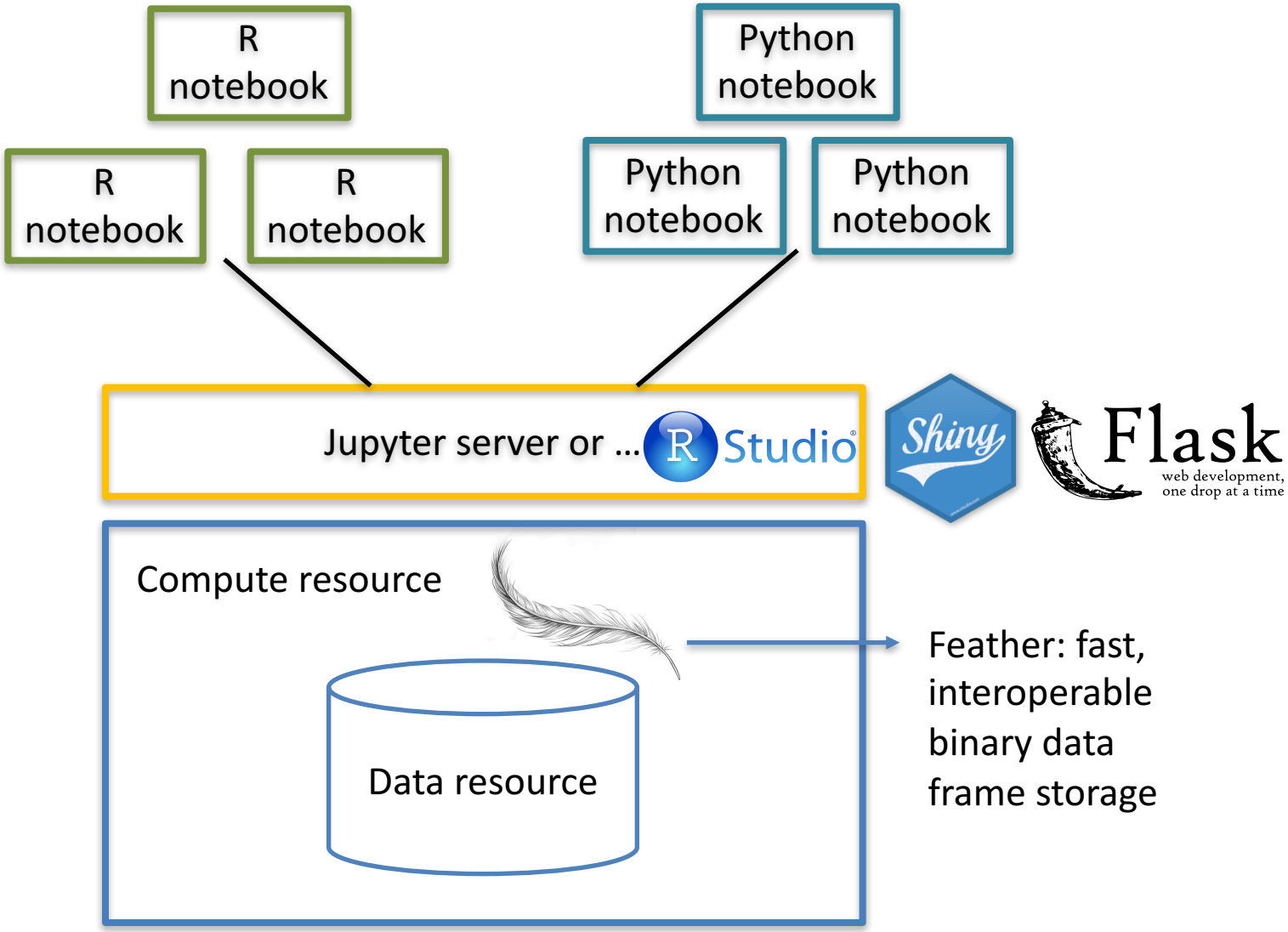
Community Driven

Reports

- Mutation Hotspots
- Clinical Actionability
- Trial Enrollment
- Germline Analysis
- Patient Similarity
- Network Analysis
- Mutational Load

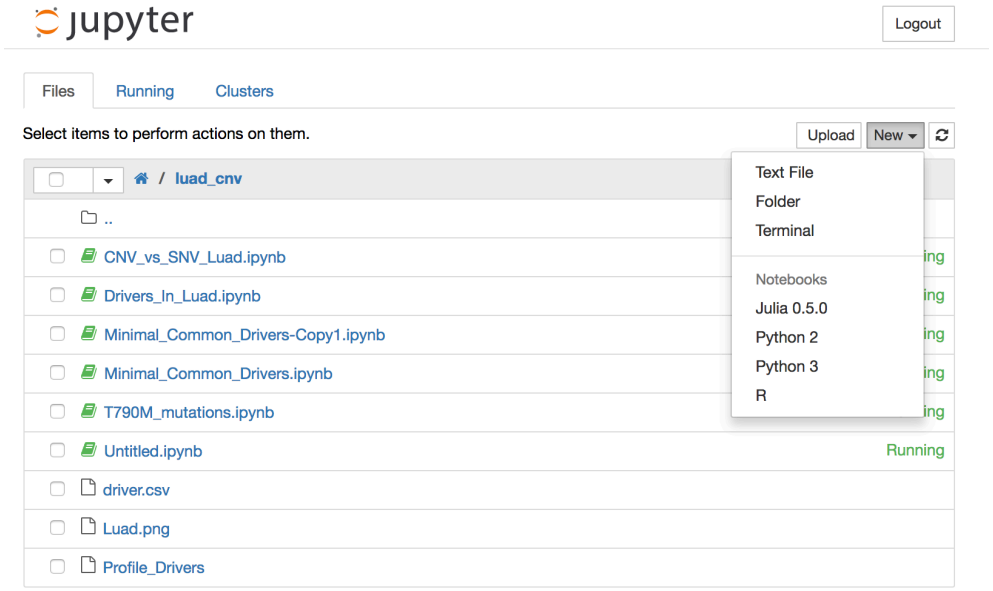
Future Profile Papers

Compute, data and code in same place



Remove barriers for data science

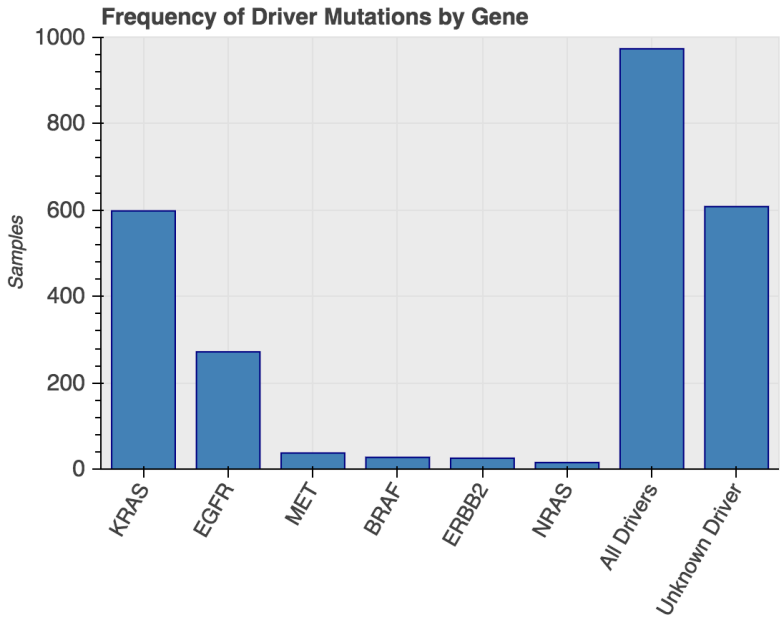
1) Create new notebook in language of choice



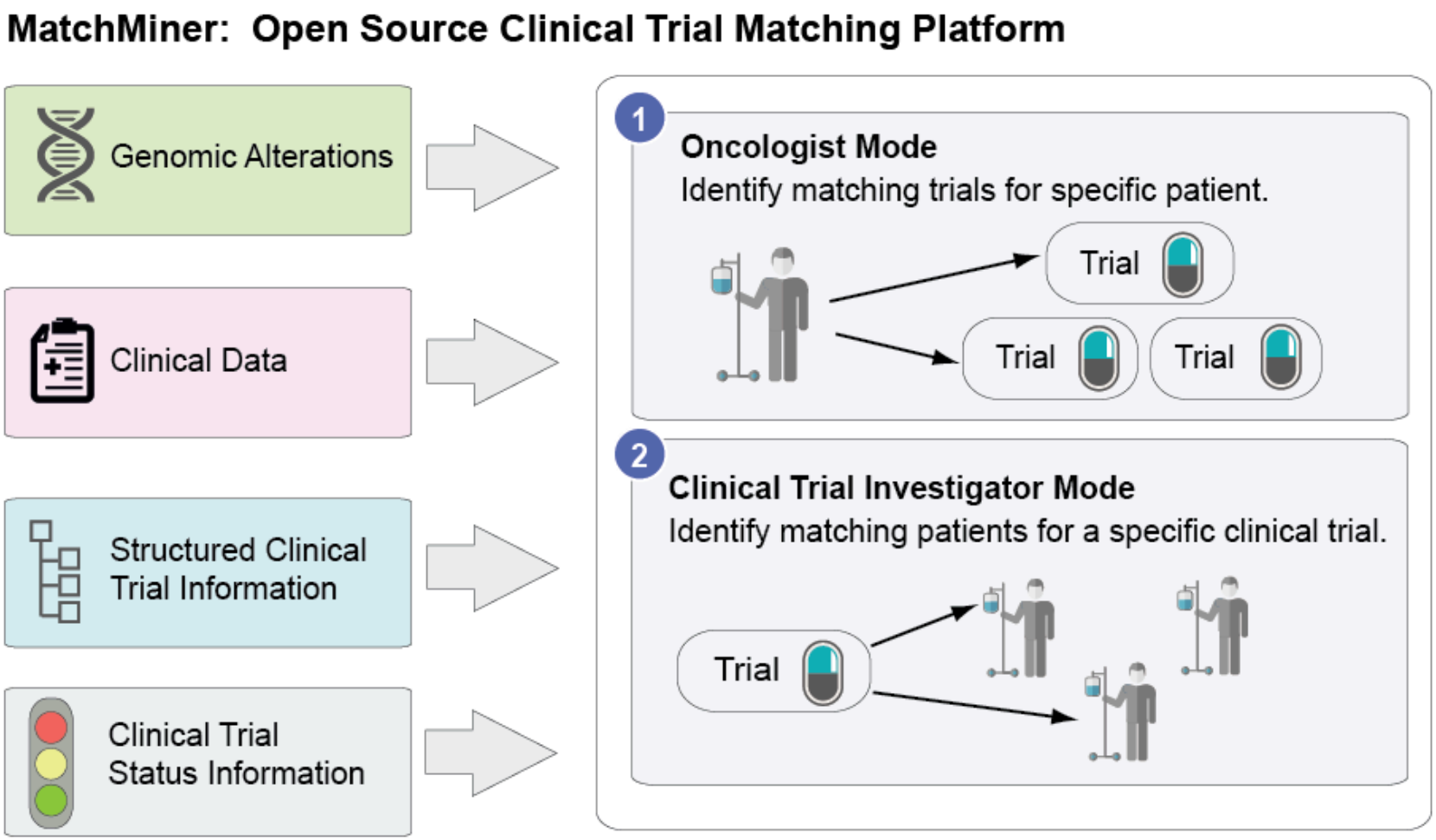
2) Pull in data via programmatic interface

```
genomic_df = data.get_genomic_df()
clinical_df = data.get_clinical_df()
cna_df = data.get_cna_df()
cna_df=cna_df.set_index('SAMPLE_ID')
merged_df = genomic_df.join(clinical_df, on="Tumor_Sample_Barcode")
```

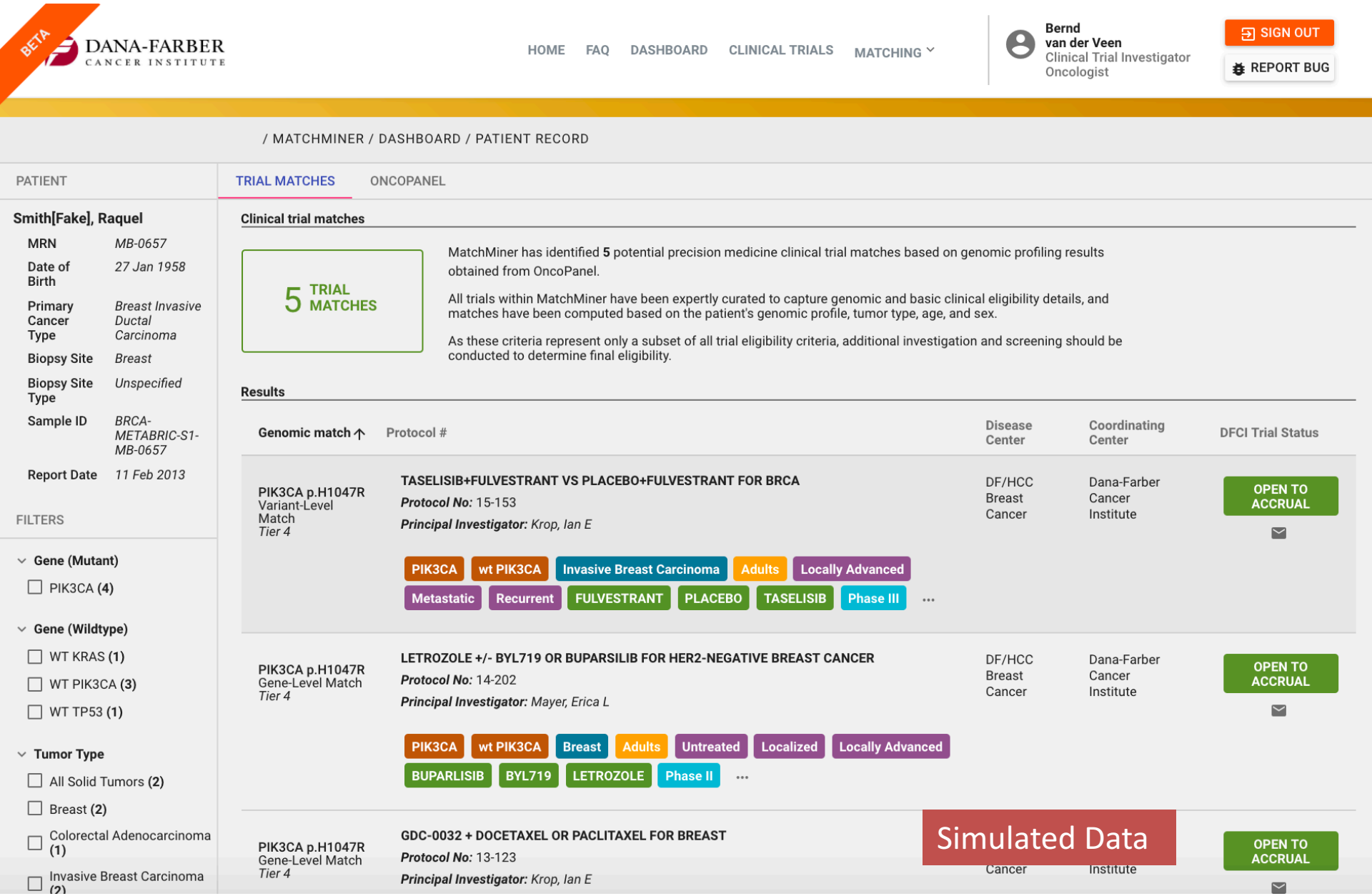
3) … science?



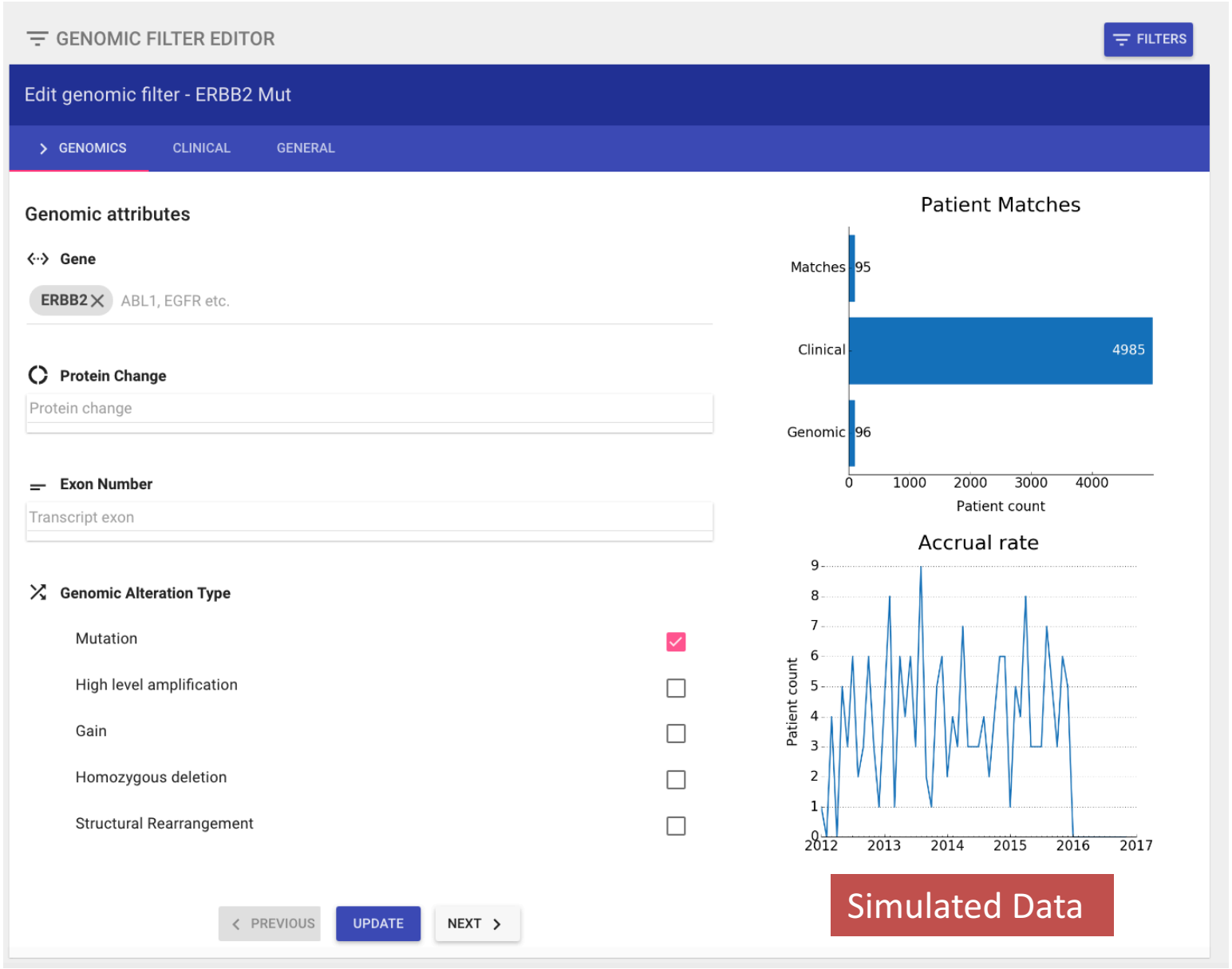
MatchMiner


MatchMiner: Open Source Clinical Trial Matching Platform
Genomic Alterations
Clinical Data
Structured Clinical Trial Information
Clinical Trial Status Information
1
Oncologist Mode
Identify matching trials for specific patient.
Trial
Trial
Trial
2
Clinical Trial Investigator Mode
Identify matching patients for a specific clinical trial.
Trial

Patient View – Trial Matches



Trial centric – Create filter



Clinical Trial Markup Language

Clinical Trial Details

```
nct_id: NCT02097225
nct_purpose: This phase I trial studies the side effects ...
phase: I
protocol_id: 6534
protocol_no: 14-186
protocol_target_accrual: 32
protocol_type: Treatment
short_title: AT13387 W/ DABRAFENIB + TRAMETINIB IN BRAF-MUTANT MELANOMA
status: Open to Accrual
treatment_list:
  step:
  - arm:
    match:
    - and:
      - and:
        - or:
          - genomic:
              hugo_symbol: BRAF
              protein_change: p.V600E
              variant_category: Mutation
          - genomic:
              hugo_symbol: BRAF
              protein_change: p.V600K
              variant_category: Mutation
        - genomic:
            hugo_symbol: KRAS
            wildtype: true
        - genomic:
            hugo_symbol: NRAS
            wildtype: true
      - clinical:
          age_numerical: '>=18'
          oncotree_primary_diagnosis: _SOLID_
```

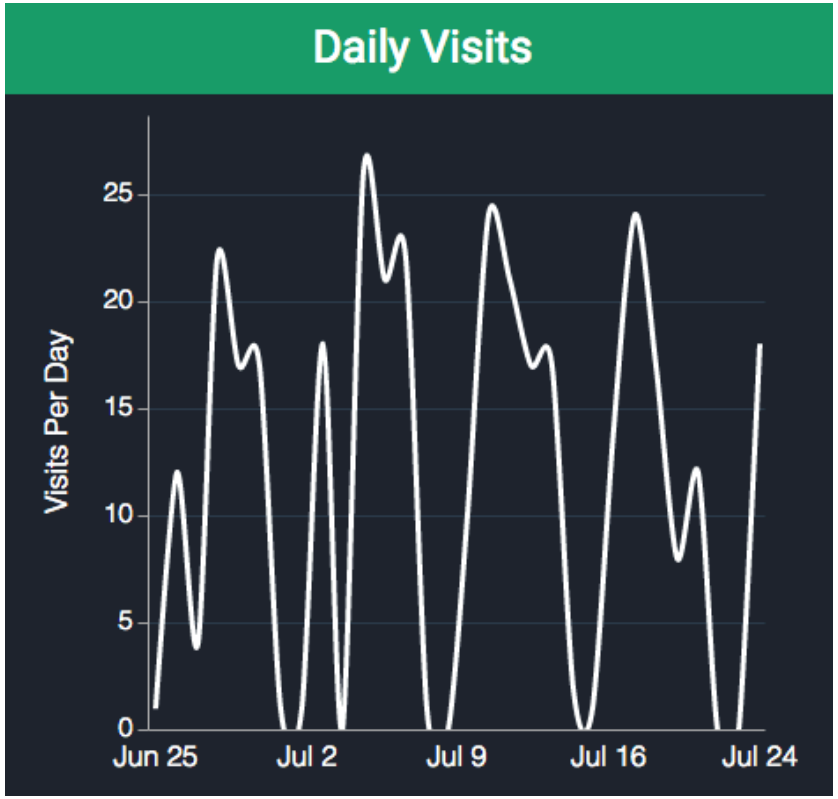
Eligibility Criteria

Genomic Criteria

Clinical Criteria

MatchMiner usage stats

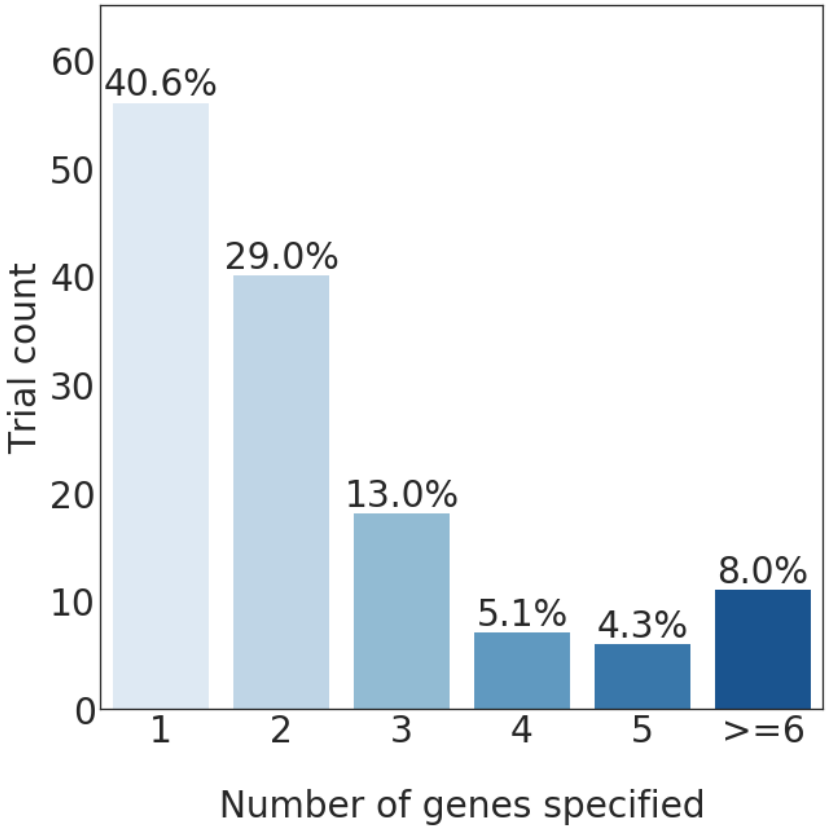
138 Trials, 675 register oncologists, 125 clinical trial investigators

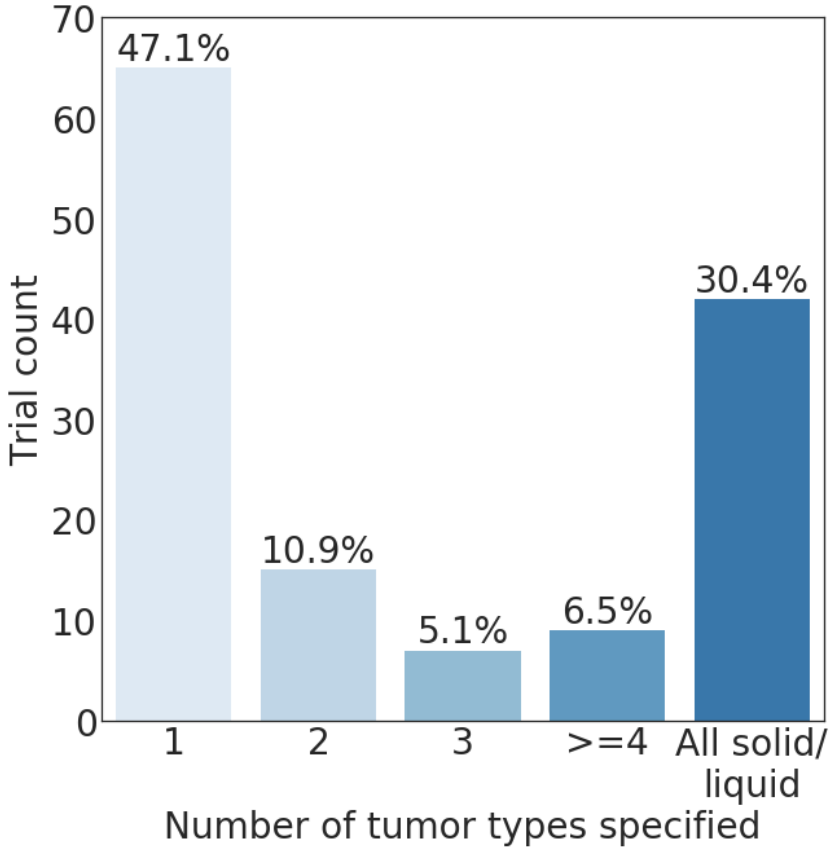


Search Terms (July)

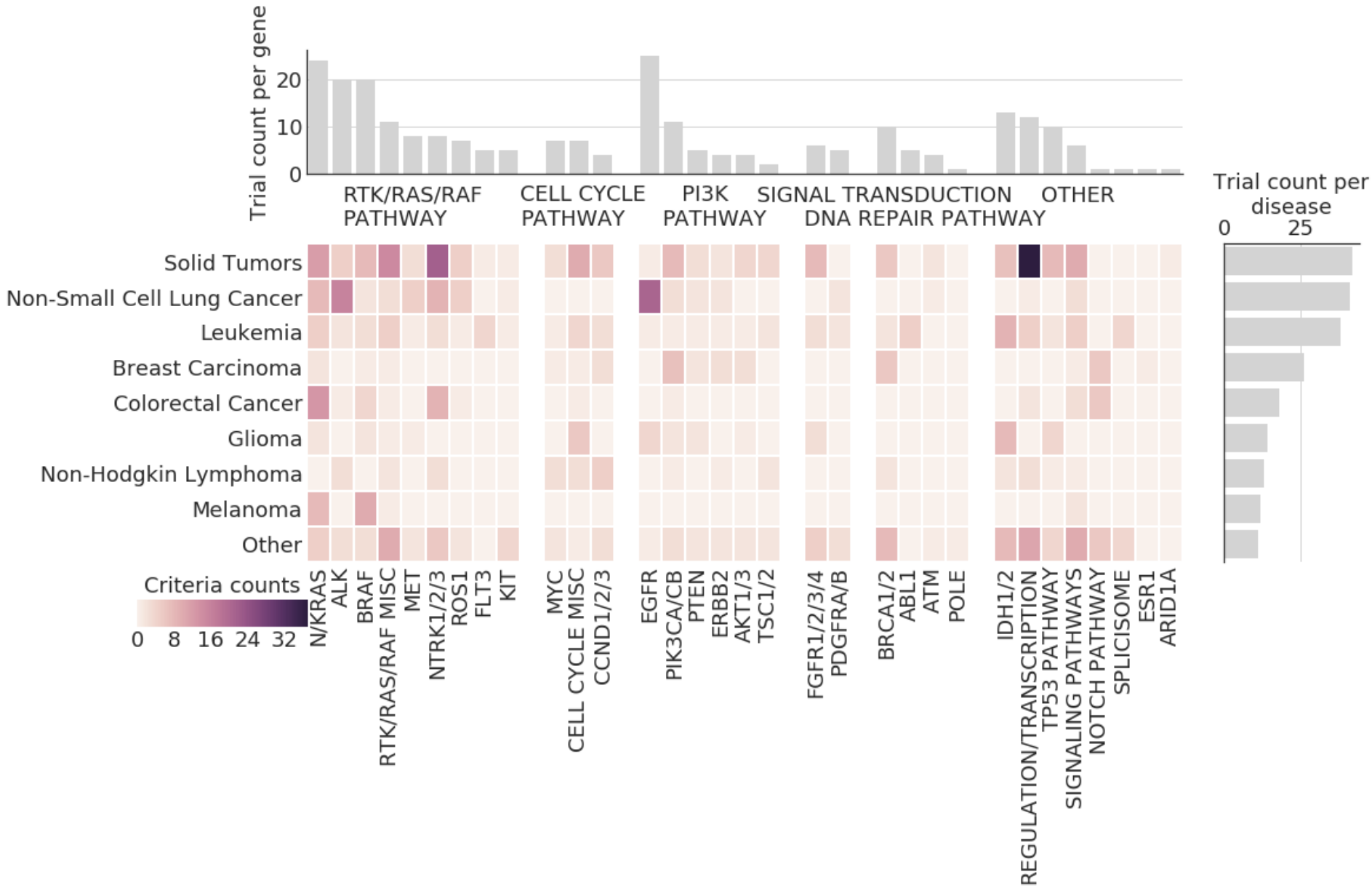
Term	Count
EGFR	6
16-494	4
16-711	3
demetri	3
(cdkn2a)	2
(cdkn2a,cdkn2b)	2
(cdkn2b)	2
(pik3ca)	2
(pik3ca,All solid tumors)	2
(pik3ca,All solid tumors,Open to Accrual)	2

Trial complexity of 138 curated trials





Genomic criteria overview



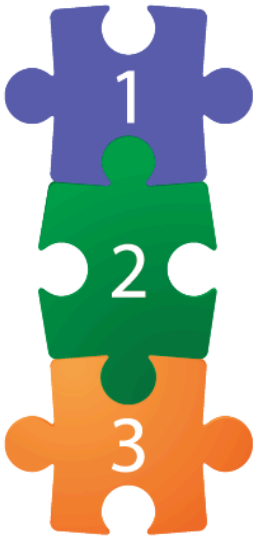
cBioOne, Intel CCC, GENIE

DATA ACQUISITION

cBioOne

cBioPortal “as a service”

- for individual investigators
- for disease centers



Profile Data

Other Genomic Data

Clinical Data

cBioPortal for Cancer Genomics

Secure Live Link File Server

Validate, Merge + De-Identify

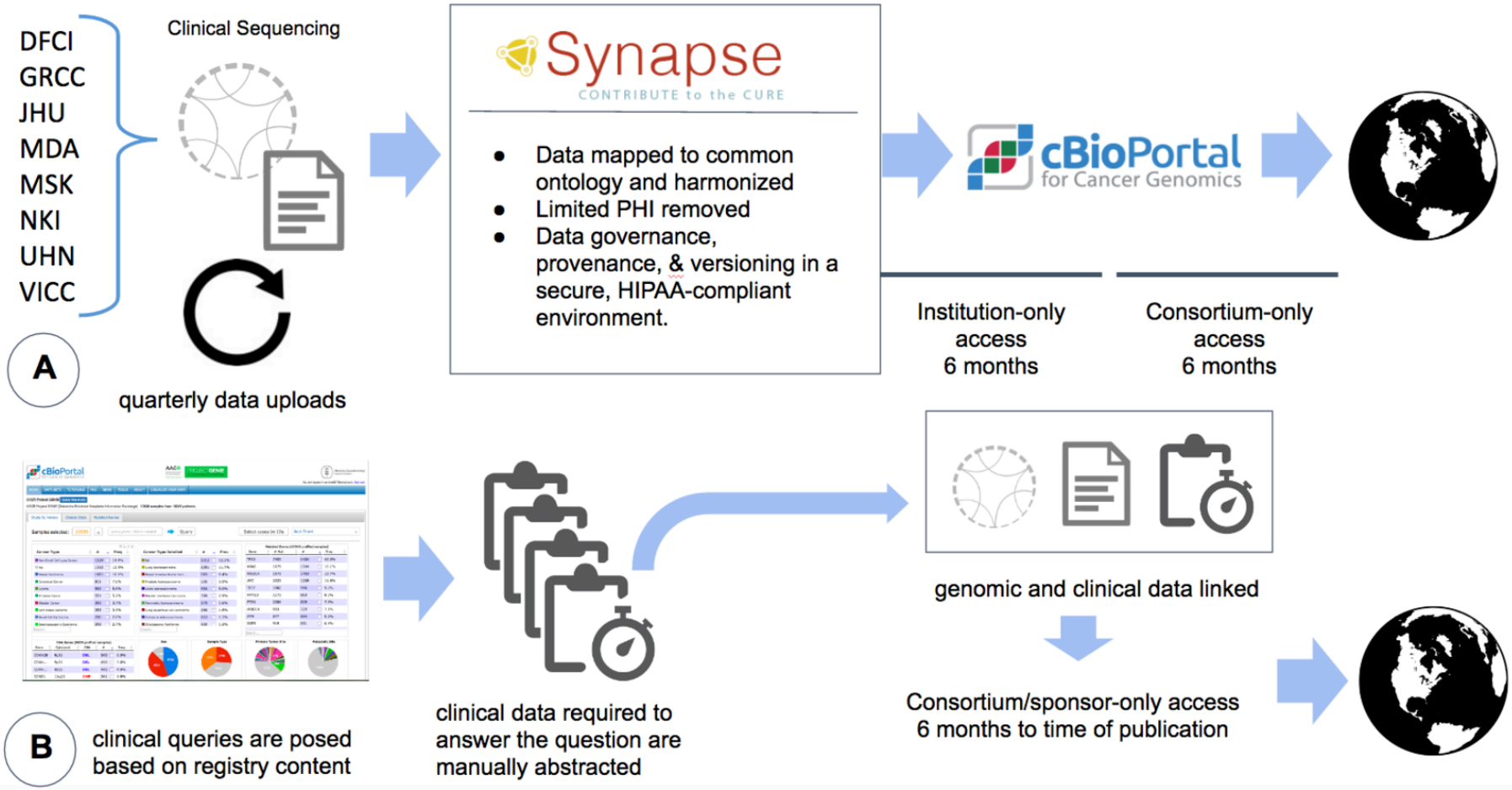
Private to Group

Shared @ DFCI

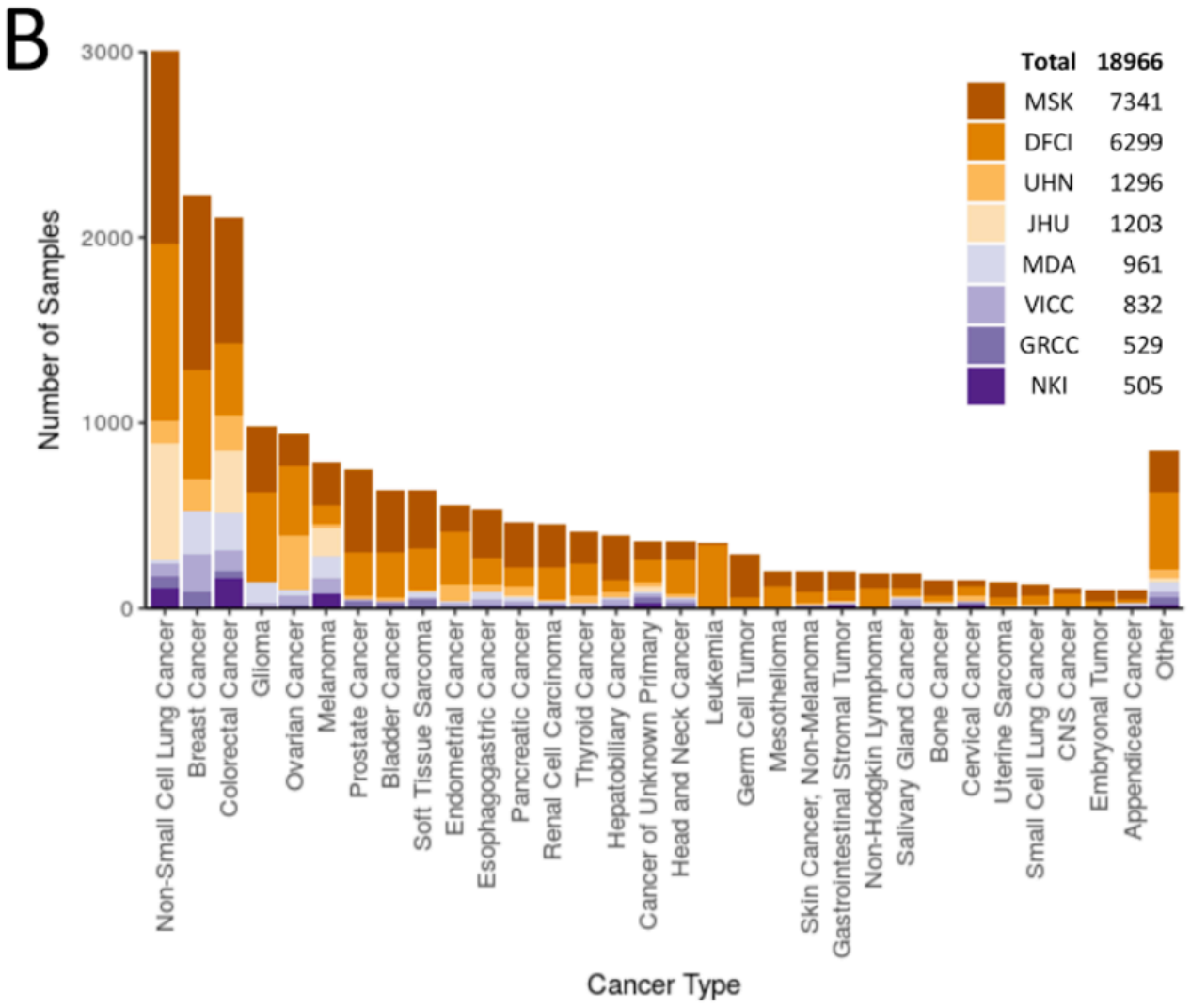
7 PIs / Groups Now Active

Priti Kumari

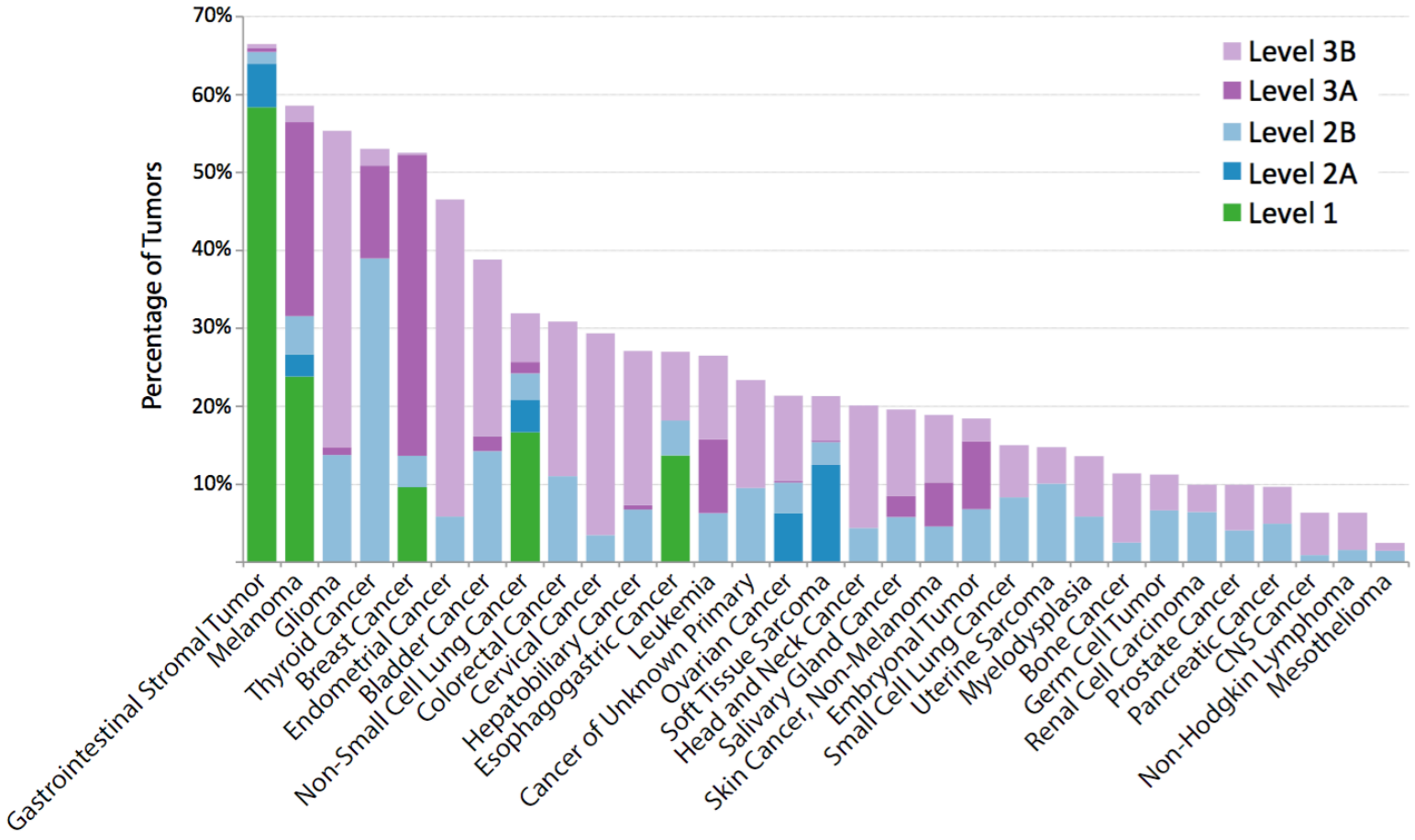
AACR GENIE: First Data Release!



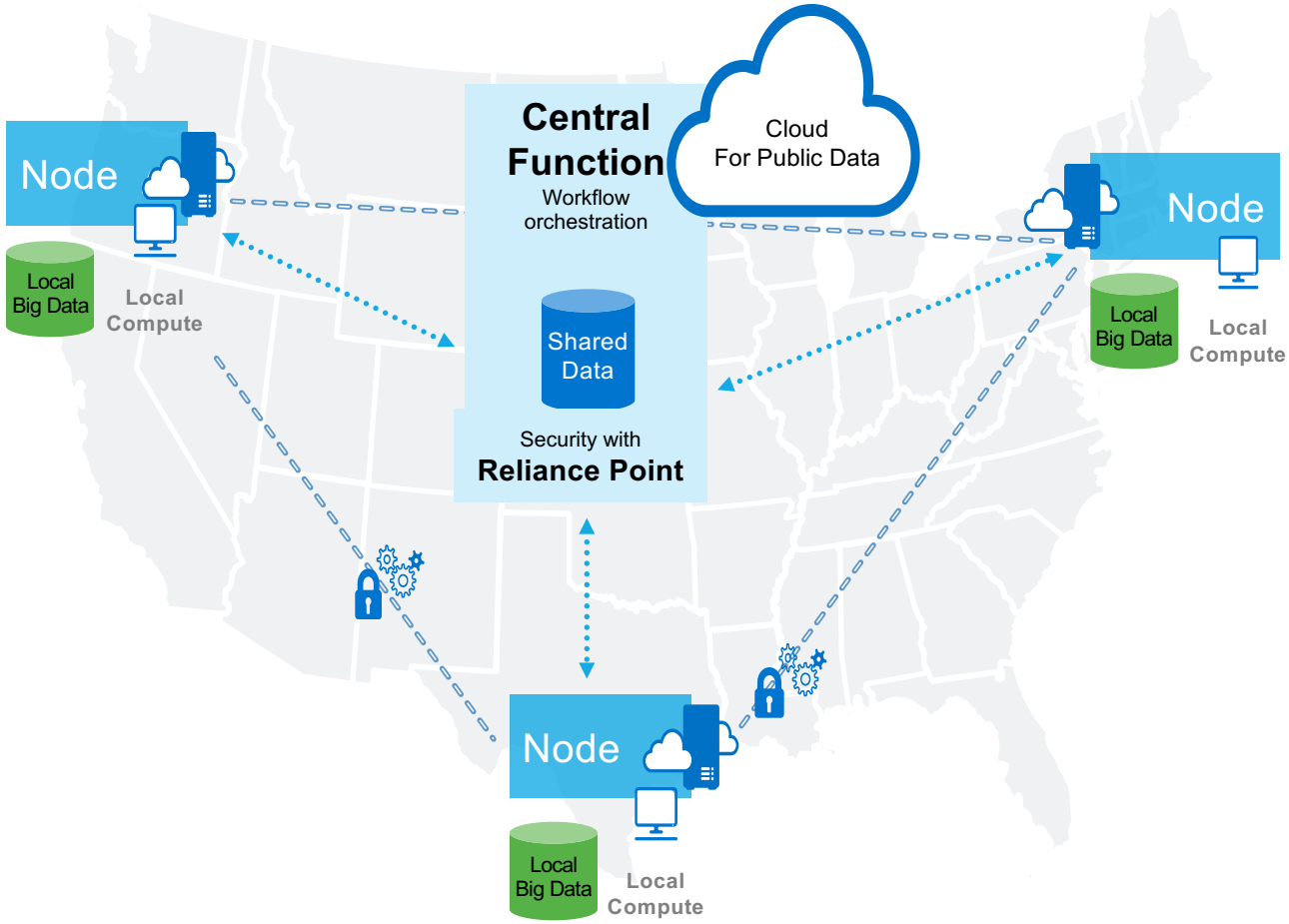
GENIE Landscape



Landscape of Clinical Actionability



Intel Collaborative Cancer Cloud

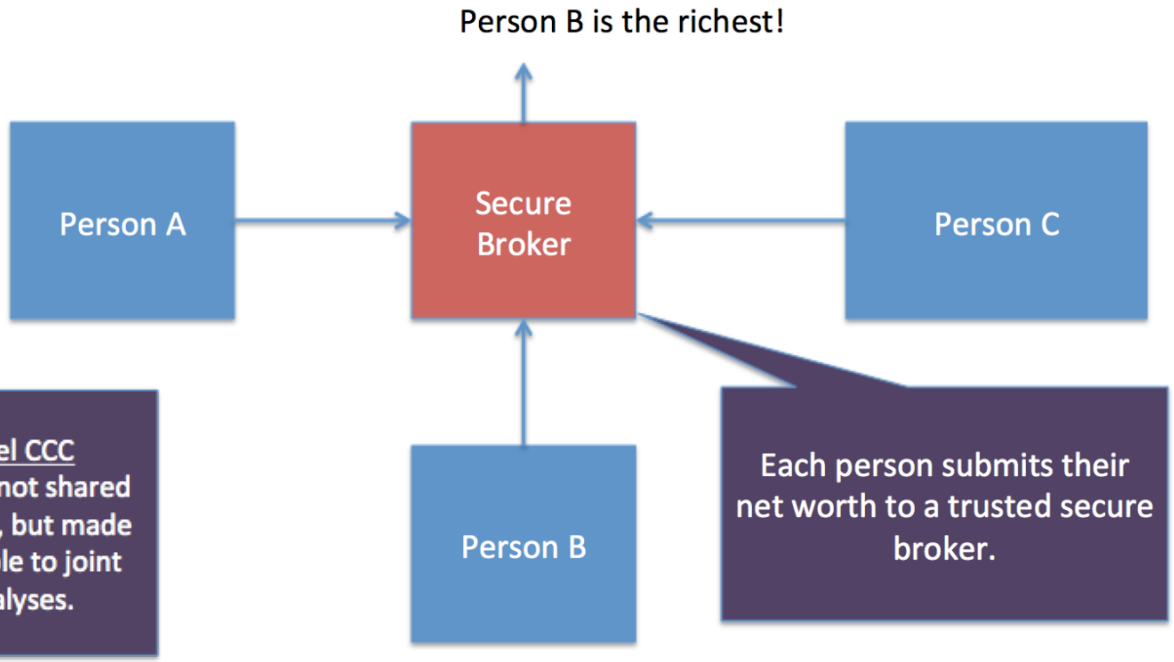


CCC is an open platform, enabling community best practice precision medicine analytics

OHSU | intel

Unique Aspect of Intel CCC: Secure joint computation

Millionaire's Dilemma

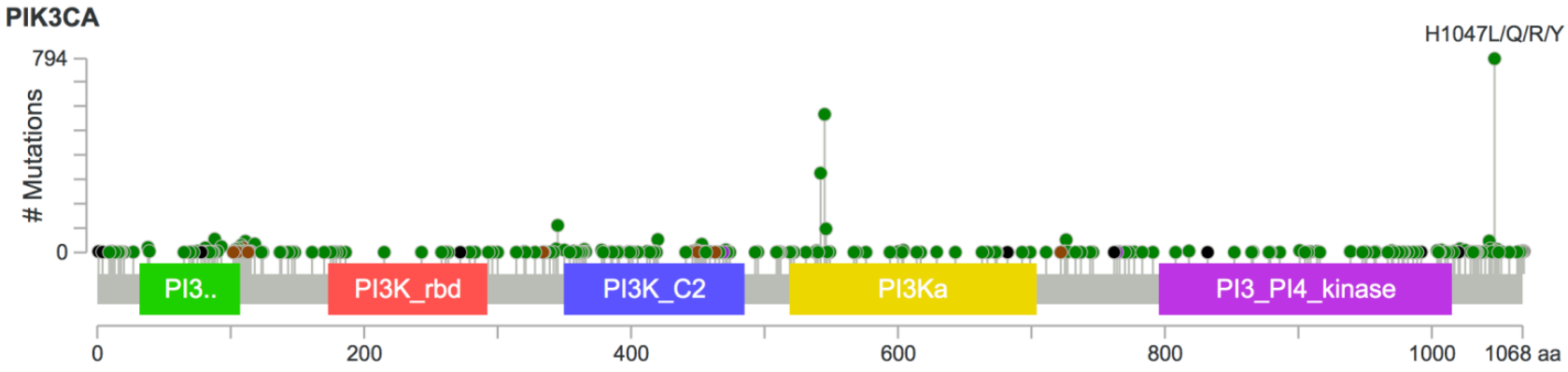


AACR GENIE:
Direct sharing of data between 8 cancer centers, but data is withheld for ~ 1 year.

Intel CCC
Data is not shared directly, but made available to joint analyses.

Motivation: Mutation Hotspots

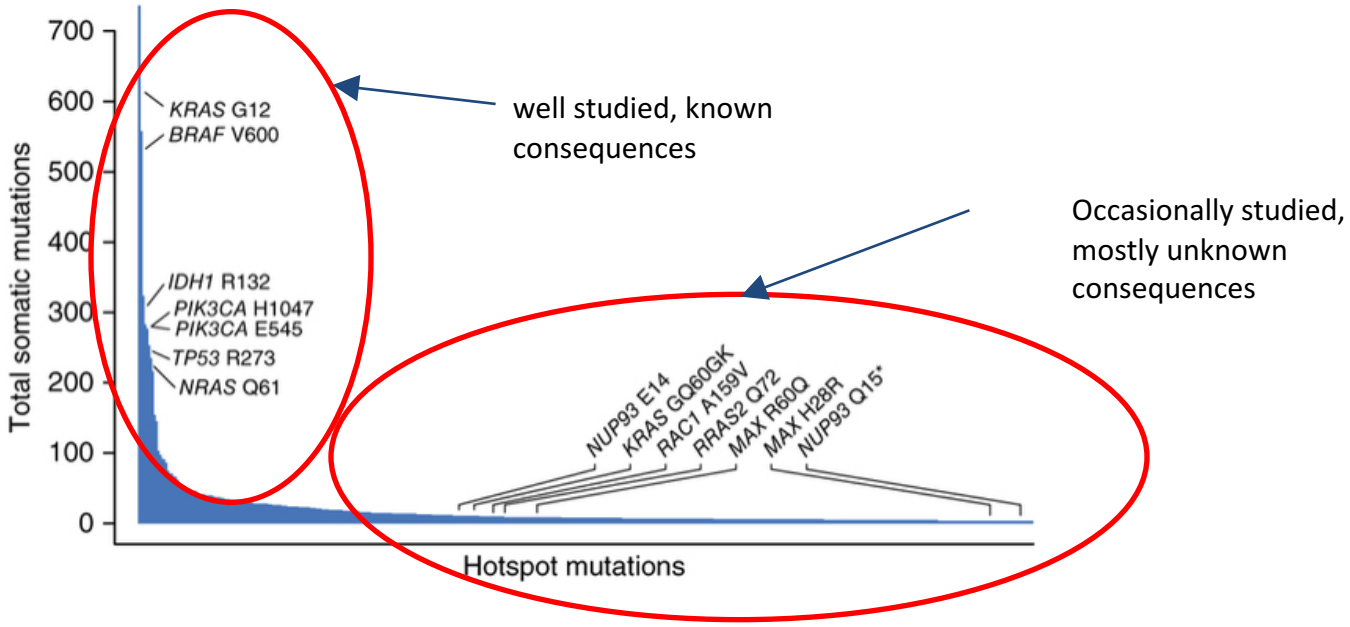
- **Mutation Hotspot**
  - a specific amino acid position that is mutated more frequently than expected by chance.
  - likely indicative of oncogenic activity.



Example Mutation Hotspots in PIK3CA (Pan-Cancer)

Motivation: Long Tail of Mutation Hotspots

"85% of all hotspots identified were mutated in less than 5% of tumors"



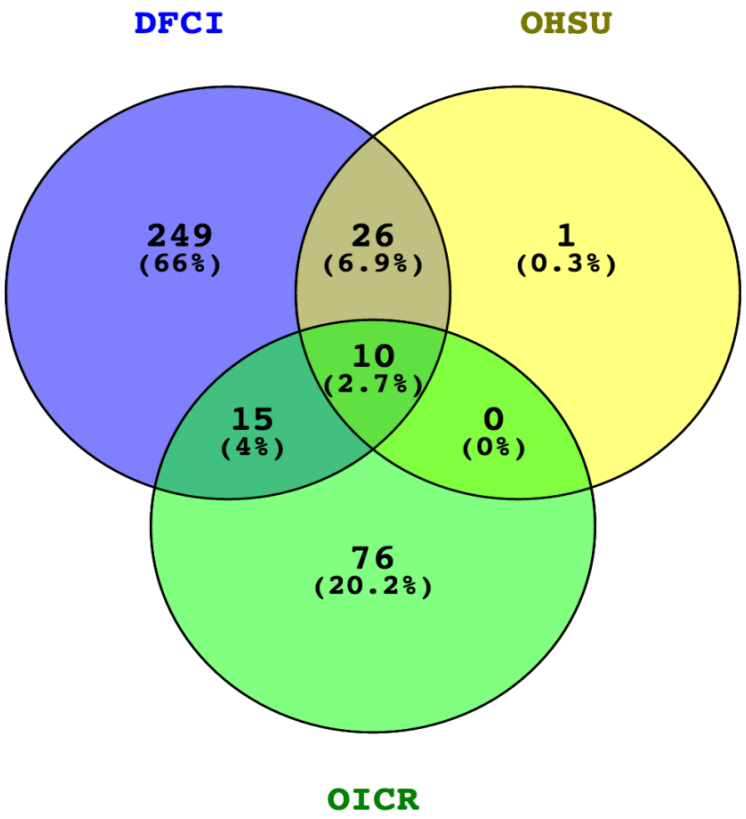
"Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity". Chang et al. 2016

Can be identified within a single cancer type or via pan-cancer analysis.

Gene Overlap of Private Data Sets

All Three Centers

AKT1
ERBB2
KRAS
MET
NF1
PIK3CA
PIK3R1
PTEN
RB1
TP53



DFCI/OHSU

AKT2
AKT3
ALK
BRAF
CDK4
CDKN2A
DDR2
EGFR
FGFR1
FGFR3
GNA11
GNAQ
GNAS
HRAS
KDR
KIT
MAP2K1
NRAS
NTRK2
NTRK3
NOTCH1
RET
STK11
TSC1
TSC2
VHL

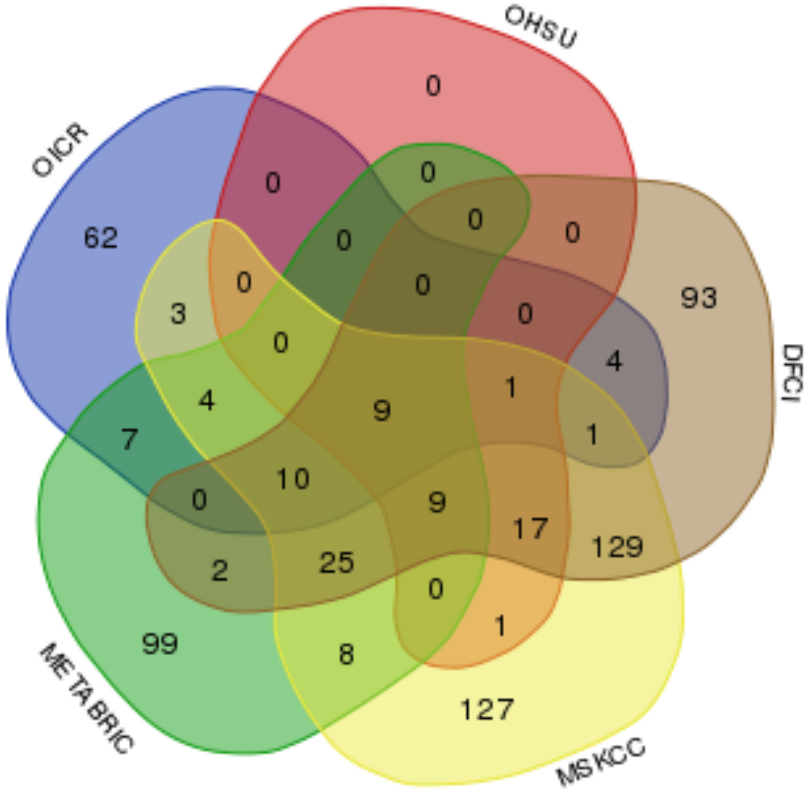
DFCI/OICR

ARID1A
BRCA1
BRCA2
CBLB
CDH1
CDKN1B
EPHA5
GATA3
MAP2K4
MAP3K1
MYB
RUNX1
SF3B1
STAG1
TLR4

Full Gene Set Overlap

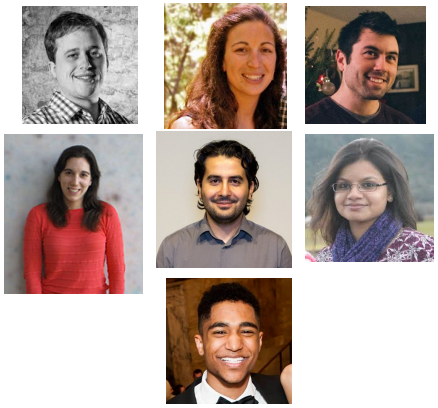
Present at all Centers

AKT1
ERBB2
KRAS
NF1
PIK3CA
PIK3R1
PTEN
RB1
TP53



Acknowledgements

Knowledge Systems



cBio @ MSKCC

Nikolaus Schultz
JianJiong Gao
Benjamin Gross

The Hyve

Sjoerd van Hagen
Pieter Lukasse
Sander de Ridder
Fedde Schaeffer
Bernd van der Veen

MatchMiner

Bruce Johnson
Drew Memmott
Geoffrey Shapiro
George Demetri
Khanh Do
Steve DuBois
Erica Woulf
Adem Albayrak
Susan Barry

DFCI / BWH

Barrett Rollins
Laura MacConaill
Jane Song
Matt Ducar
Priyanka Shivdasani
Lynette Sholl
Neal Lindeman
Stacy Gray
Eliezer Van Allen