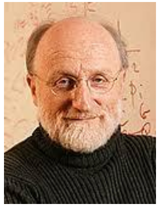
cBioPortal
FOR CANCER GENOMICS

Nikolaus Schultz, PhD
Director, Cancer Data Science Initiative
Attending Computational Oncologist, Department of Epidemiology & Biostatistics
Memorial Sloan Kettering Cancer Center

October 17, 2024

The cBioPortal for Cancer Genomics

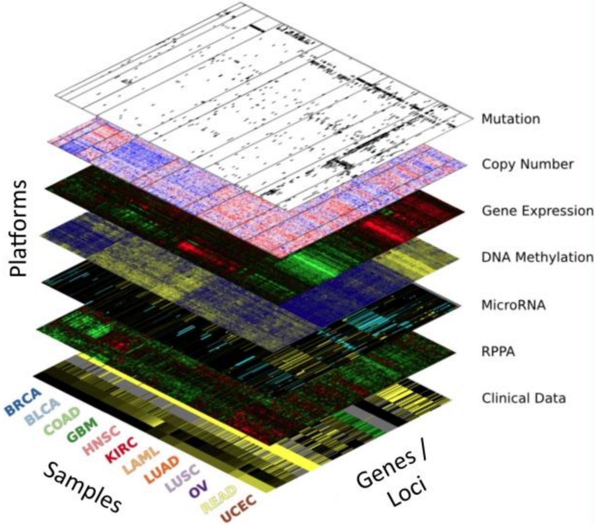
*Goal: Make complex cancer genomic data **accessible** and **interpretable** for cancer biologists and clinicians.*



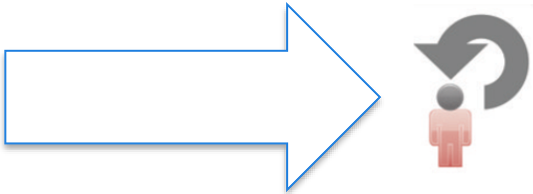
Chris Sander

2006

Start of TCGA



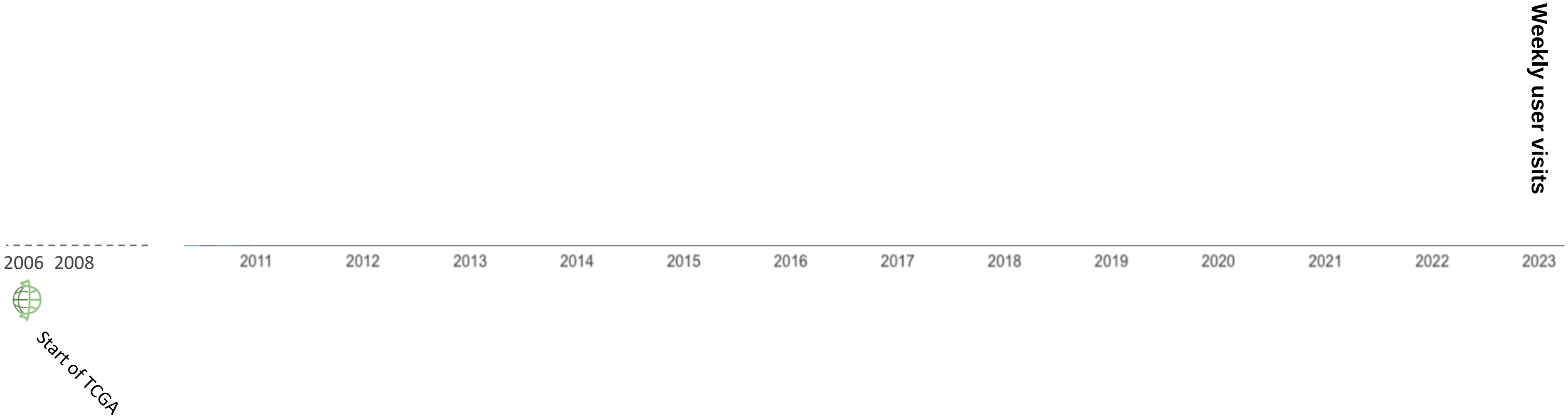
Adapted from TCGA Nat Genet 2014



Biological discovery

Clinical interpretation

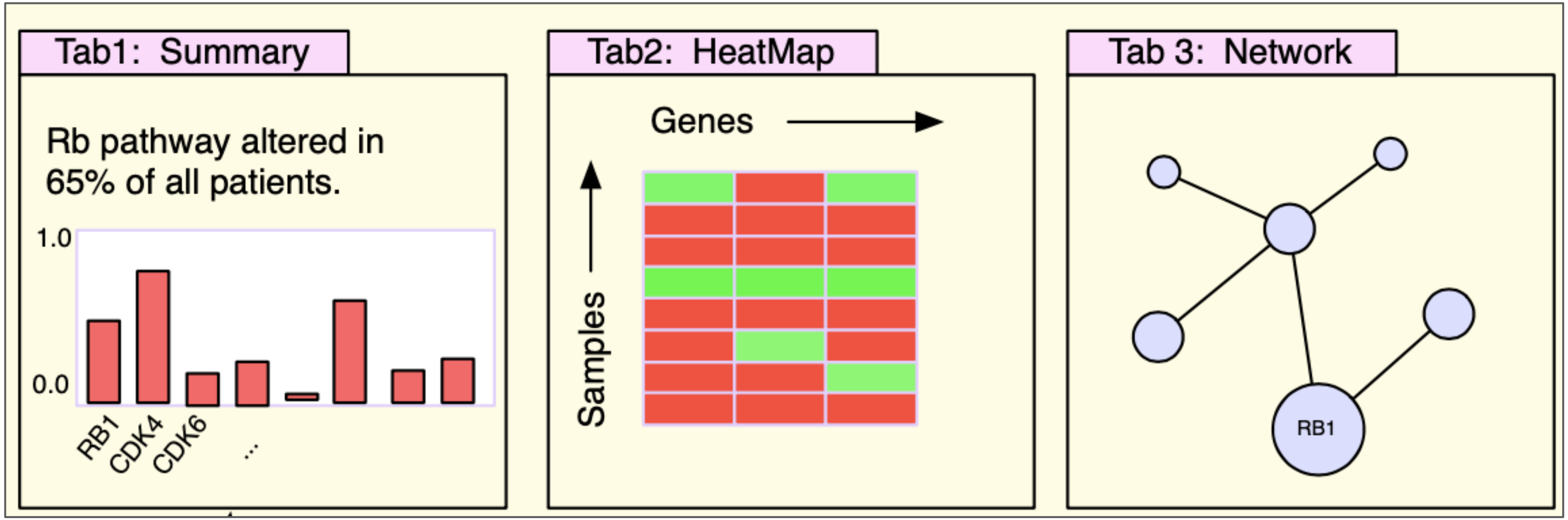
The cBioPortal for Cancer Genomics



The cBioPortal for Cancer Genomics

August 2008: First specs

Ethan Cerami



Tab1: Summary

Rb pathway altered in 65% of all patients.

Tab2: HeatMap

Tab 3: Network

2006 2008 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023

Start of TCGA

The cBioPortal for Cancer Genomics

October 2008: First live version

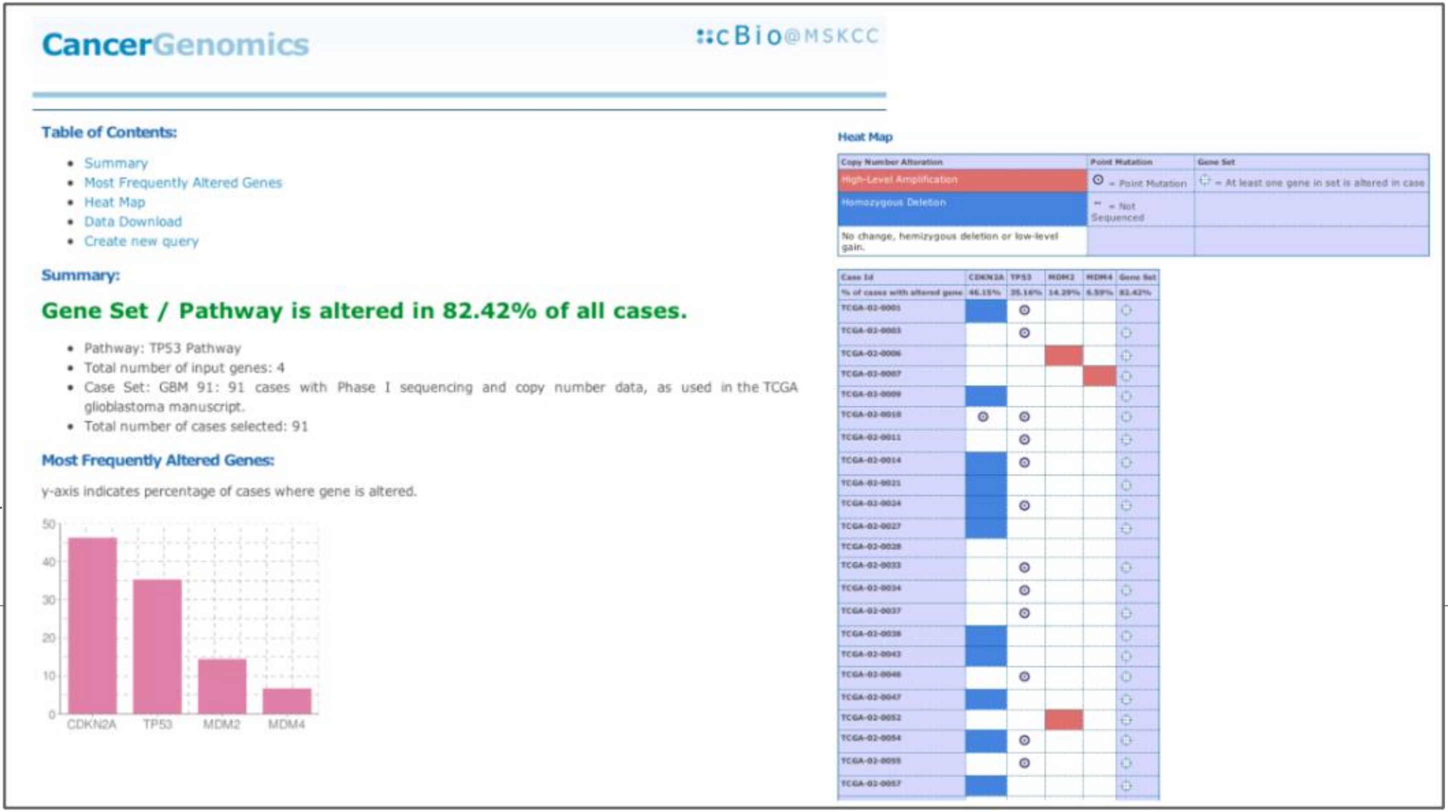


Table of Contents:

- Summary
- Most Frequently Altered Genes
- Heat Map
- Data Download
- Create new query

Summary:

Gene Set / Pathway is altered in 82.42% of all cases.

- Pathway: TP53 Pathway
- Total number of input genes: 4
- Case Set: GBM 91: 91 cases with Phase I sequencing and copy number data, as used in the TCGA glioblastoma manuscript.
- Total number of cases selected: 91

Most Frequently Altered Genes:

y-axis indicates percentage of cases where gene is altered.

Heat Map

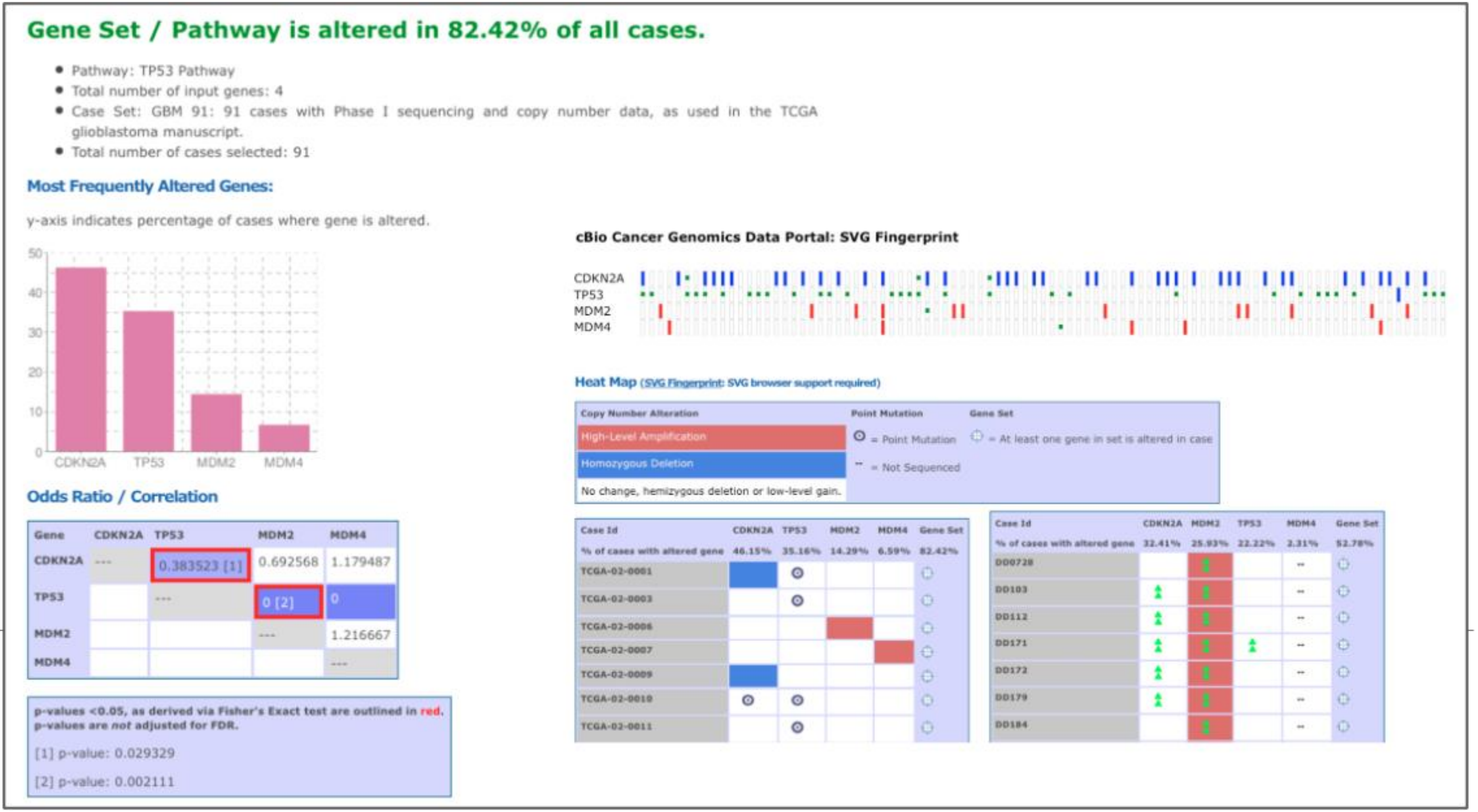
Copy Number Alteration	Point Mutation	Gene Set
🟥 High-Level Amplification	⊙ = Point Mutation	⌖ = At least one gene in set is altered in case
🟦 Homozygous Deletion	"" = Not Sequenced	
No change, hemizygous deletion or low-level gain.		

Case Id	CDKN2A	TP53	MDM2	MDM4	Gene Set
% of cases with altered gene	46.15%	35.16%	14.29%	6.59%	82.42%
TCGA-02-0001	🟦	⊙			⌖
TCGA-02-0003		⊙			⌖
TCGA-02-0006			🟥		⌖
TCGA-02-0007				🟥	⌖
TCGA-02-0009	🟦				⌖
TCGA-02-0010	⊙	⊙			⌖
TCGA-02-0011		⊙			⌖
TCGA-02-0014	🟦	⊙			⌖
TCGA-02-0021	🟦				⌖
TCGA-02-0024	🟦	⊙			⌖
TCGA-02-0027	🟦				⌖
TCGA-02-0028					
TCGA-02-0033		⊙			⌖
TCGA-02-0034		⊙			⌖
TCGA-02-0037		⊙			⌖
TCGA-02-0038	🟦				⌖
TCGA-02-0043	🟦				⌖
TCGA-02-0046		⊙			⌖
TCGA-02-0047	🟦				⌖
TCGA-02-0052			🟥		⌖
TCGA-02-0054	🟦	⊙			⌖
TCGA-02-0055		⊙			⌖
TCGA-02-0057	🟦				⌖

2023

The cBioPortal for Cancer Genomics

January 2009: First Oncoprints



Gene Set / Pathway is altered in 82.42% of all cases.

- Pathway: TP53 Pathway
- Total number of input genes: 4
- Case Set: GBM 91: 91 cases with Phase I sequencing and copy number data, as used in the TCGA glioblastoma manuscript.
- Total number of cases selected: 91

Most Frequently Altered Genes:

y-axis indicates percentage of cases where gene is altered.

Odds Ratio / Correlation

Gene	CDKN2A	TP53	MDM2	MDM4
CDKN2A	---	0.383523 [1]	0.692568	1.179487
TP53		---	0 [2]	0
MDM2			---	1.216667
MDM4				---

p-values <0.05, as derived via Fisher's Exact test are outlined in red.
p-values are *not* adjusted for FDR.

[1] p-value: 0.029329

[2] p-value: 0.002111

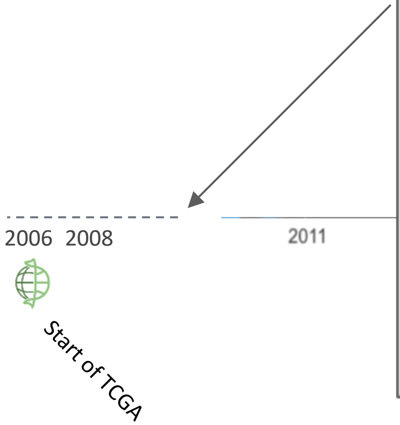
cBio Cancer Genomics Data Portal: SVG Fingerprint

Heat Map (SVG Fingerprint: SVG browser support required)

Copy Number Alteration	Point Mutation	Gene Set
🟥 High-Level Amplification	⊙ = Point Mutation	⊕ = At least one gene in set is altered in case
🟦 Homozygous Deletion	-- = Not Sequenced	
No change, hemizygous deletion or low-level gain.		

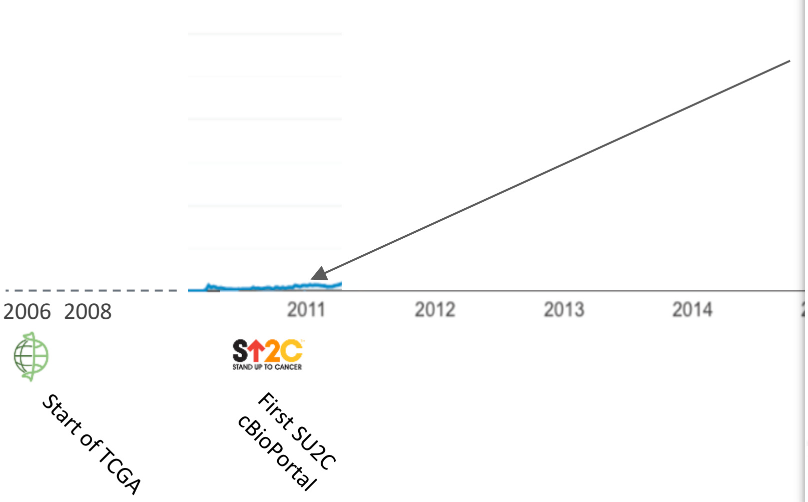
Case Id	CDKN2A	TP53	MDM2	MDM4	Gene Set
% of cases with altered gene	46.15%	35.16%	14.29%	6.59%	82.42%
TCGA-02-0001	🟦	⊙			⊕
TCGA-02-0003		⊙			⊕
TCGA-02-0006			🟥		⊕
TCGA-02-0007				🟥	⊕
TCGA-02-0009	🟦				⊕
TCGA-02-0010	⊙	⊙			⊕
TCGA-02-0011		⊙			⊕

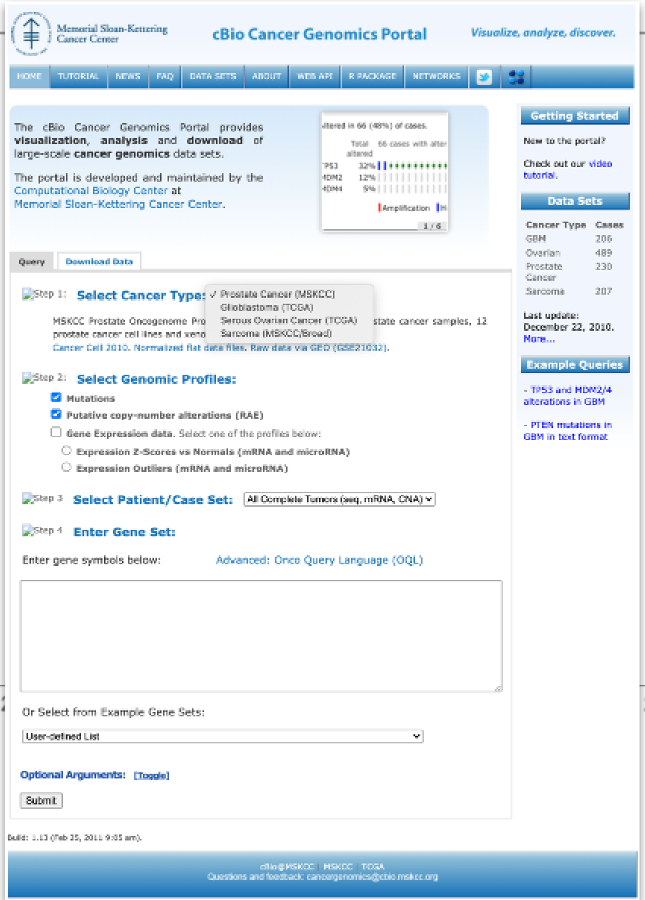
Case Id	CDKN2A	MDM2	TP53	MDM4	Gene Set
% of cases with altered gene	32.41%	25.93%	22.22%	2.31%	52.78%
DD0728		🟥 ▲		--	⊕
DD103	▲	🟥 ▲		--	⊕
DD112	▲	🟥 ▲		--	⊕
DD171	▲	🟥 ▲	▲	--	⊕
DD172	▲	🟥 ▲		--	⊕
DD179	▲	🟥 ▲		--	⊕
DD184		🟥 ▲		--	⊕



The cBioPortal for Cancer Genomics

2011: Four studies, 600 weekly visits





cBio Cancer Genomics Portal

Visualize, analyze, discover.

The cBio Cancer Genomics Portal provides **visualization**, **analysis** and **download** of large-scale **cancer genomics** data sets.

The portal is developed and maintained by the Computational Biology Center at Memorial Sloan-Kettering Cancer Center.

Step 1: **Select Cancer Type:**
- ✓ Prostate Cancer (MSKCC)
- Glioblastoma (TCGA)
- Serous Ovarian Cancer (TCGA)
- Sarcoma (MSKCC/Broad)

MSKCC Prostate Oncogenome Pro… …state cancer samples, 12 prostate cancer cell lines and xeno… Cancer Cell 2010. Normalized flat data files. Raw data via GEO (GSE21032).

Step 2: **Select Genomic Profiles:**
- ☑ **Mutations**
- ☑ **Putative copy-number alterations (RAE)**
- ☐ **Gene Expression data.** Select one of the profiles below:
  - ○ **Expression Z-Scores vs Normals (mRNA and microRNA)**
  - ○ **Expression Outliers (mRNA and microRNA)**

Step 3 **Select Patient/Case Set:** All Complete Tumors (seq, mRNA, CNA)

Step 4 **Enter Gene Set:**

Enter gene symbols below: Advanced: Onco Query Language (OQL)

Or Select from Example Gene Sets:

User-defined List

Optional Arguments: [Toggle]

Submit

Build: 1.13 (Feb 25, 2011 9:05 am).

Getting Started

New to the portal?

Check out our video tutorial.

Data Sets

Cancer Type	Cases
GBM	206
Ovarian	489
Prostate Cancer	230
Sarcoma	207

Last update: December 22, 2010. More...

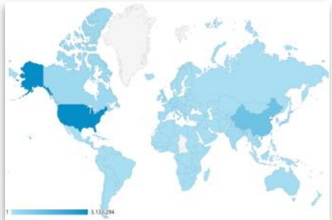
Example Queries

- TP53 and MDM2/4 alterations in GBM
- PTEN mutations in GBM in text format

cBio@MSKCC | MSKCC | TCGA
Questions and feedback: cancergenomics@cbio.mskcc.org

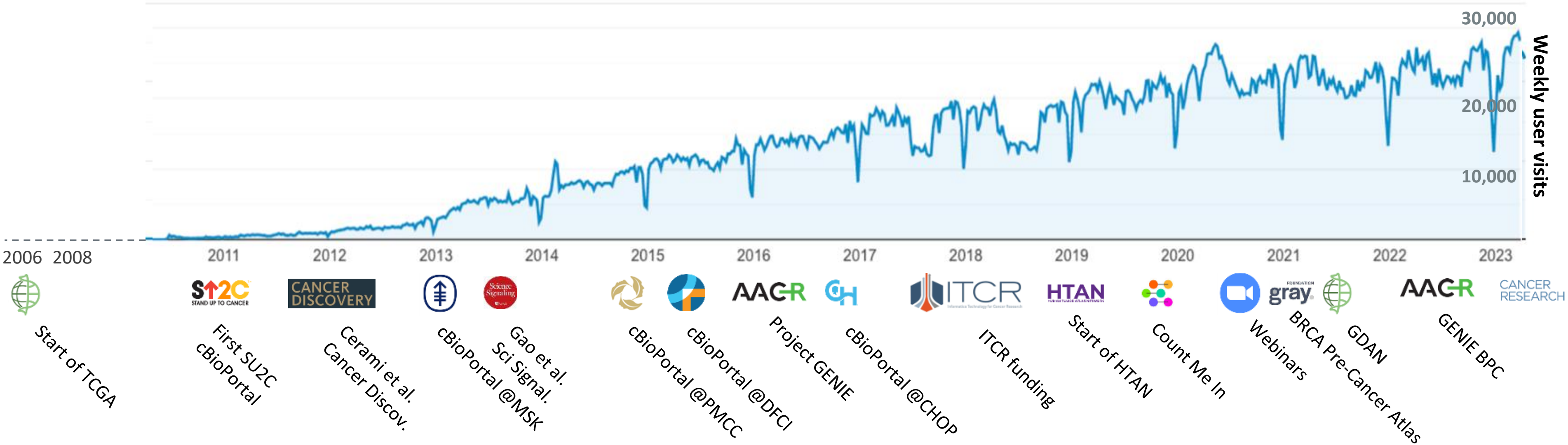
cBioPortal over time

Monthly Users	~38K
Total Citations	>30k
Curated Studies	468

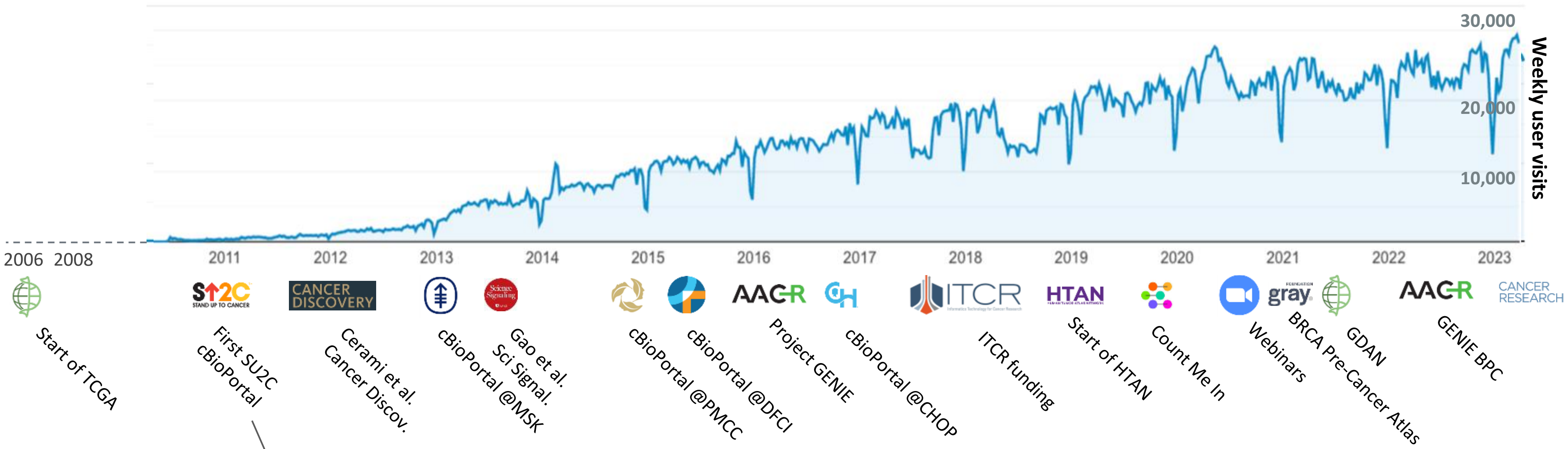


> 8,000,000 visits to cBioPortal.org since 2011

>93 institutional instances globally



cBioPortal over time



Start of TCGA

First SU2C cBioPortal

Cerami et al. Cancer Discov.

cBioPortal @MSK

Gao et al. Sci Signal.

cBioPortal @PMCC

cBioPortal @DFCI

Project GENIE

cBioPortal @CHOP

ITCR funding

Start of HTAN

Count Me In

Webinars

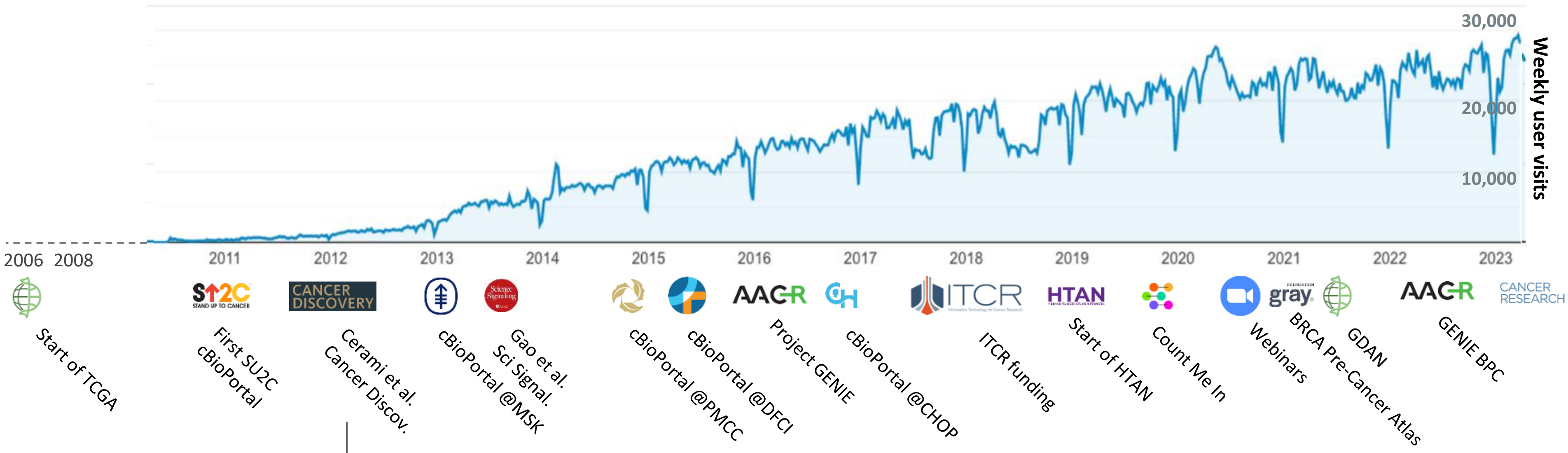
BRCA Pre-Cancer Atlas

GDAN

GENIE BPC

Use cBioPortal for sharing **Stand Up to Cancer (SU2C) data**

cBioPortal over time



First paper on cBioPortal!


Ethan


JJ

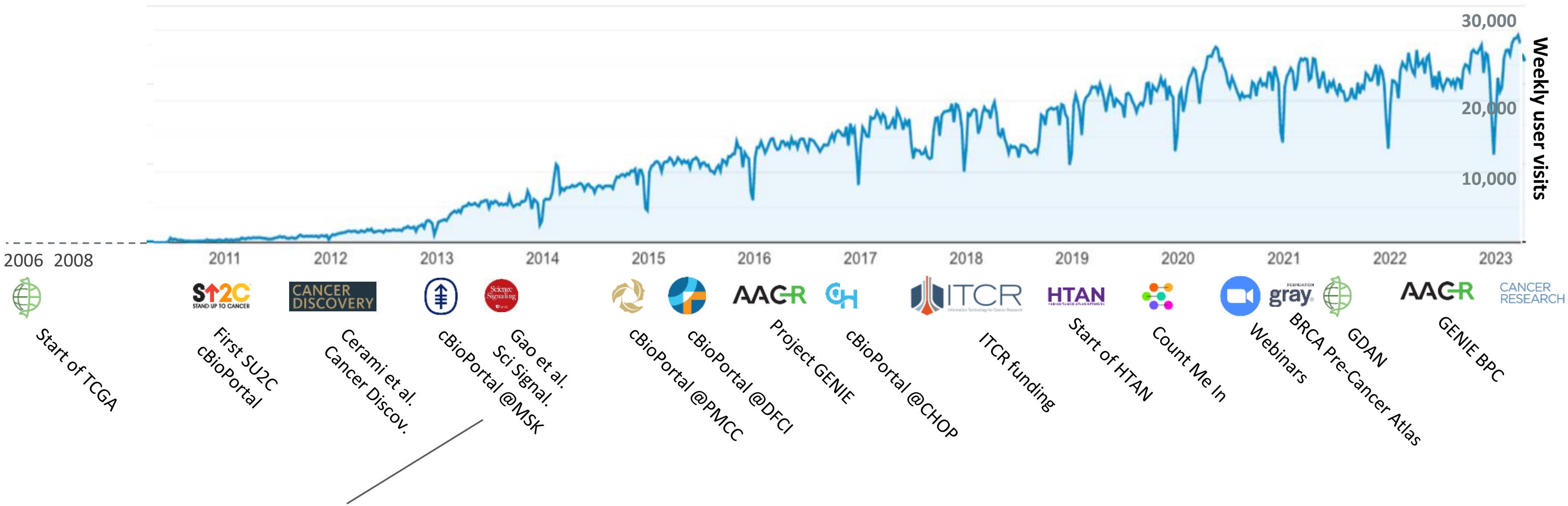

Ben


Onur


Chris


Niki

cBioPortal over time



Use cBioPortal for sharing **MSK Clinical Sequencing Data**

cBioPortal over time

Start of TCGA

First SU2C cBioPortal

Cerami et al. Cancer Discov.

cBioPortal @MSK

Gao et al. Sci Signal.

cBioPortal @PMCC

cBioPortal @DFCI

Project GENIE

cBioPortal @CHOP

ITCR funding

Start of HTAN

Count Me In

Webinars

BRCA Pre-Cancer Atlas

GDAN

GENIE BPC

Second paper on cBioPortal!


JJ


Ugur


Ethan

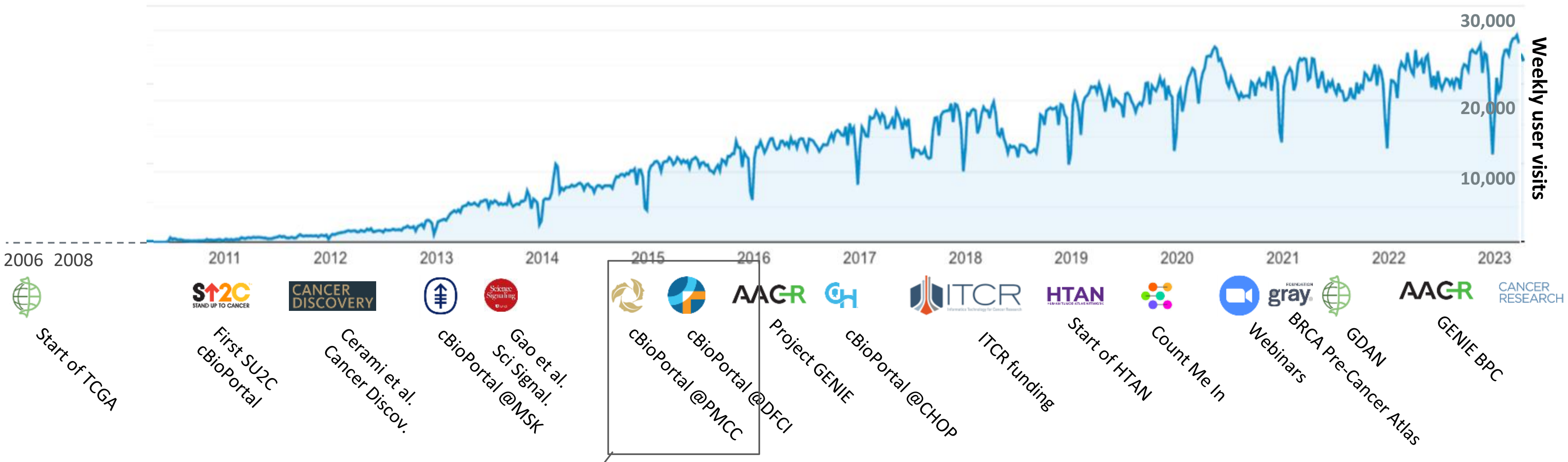

Onur


Ben


Chris


Niki

cBioPortal over time



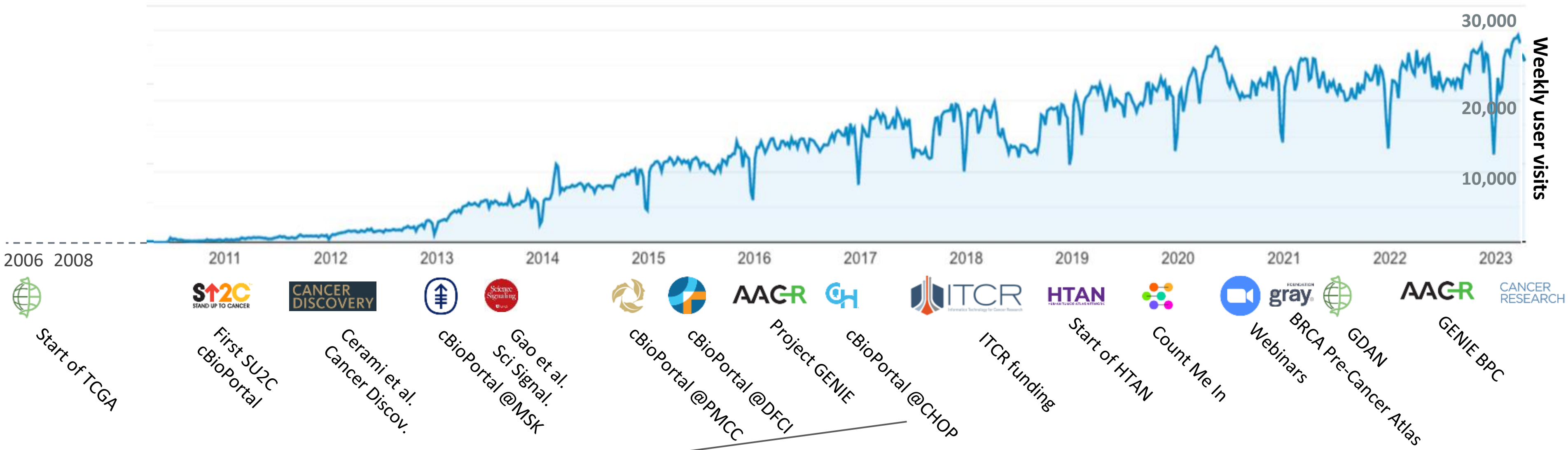
cBioPortal goes **Open Source!**

cBioPortal over time

- Start of TCGA
- First SU2C cBioPortal
- Cerami et al. Cancer Discov.
- cBioPortal @MSK
- Gao et al. Sci Signal.
- cBioPortal @PMCC
- cBioPortal @DFCI
- Project GENIE
- cBioPortal @CHOP
- ITCR funding
- Start of HTAN
- Count Me In
- Webinars
- BRCA Pre-Cancer Atlas
- GDAN
- GENIE BPC

Use cBioPortal for sharing **AACR Project GENIE** clinical sequencing data

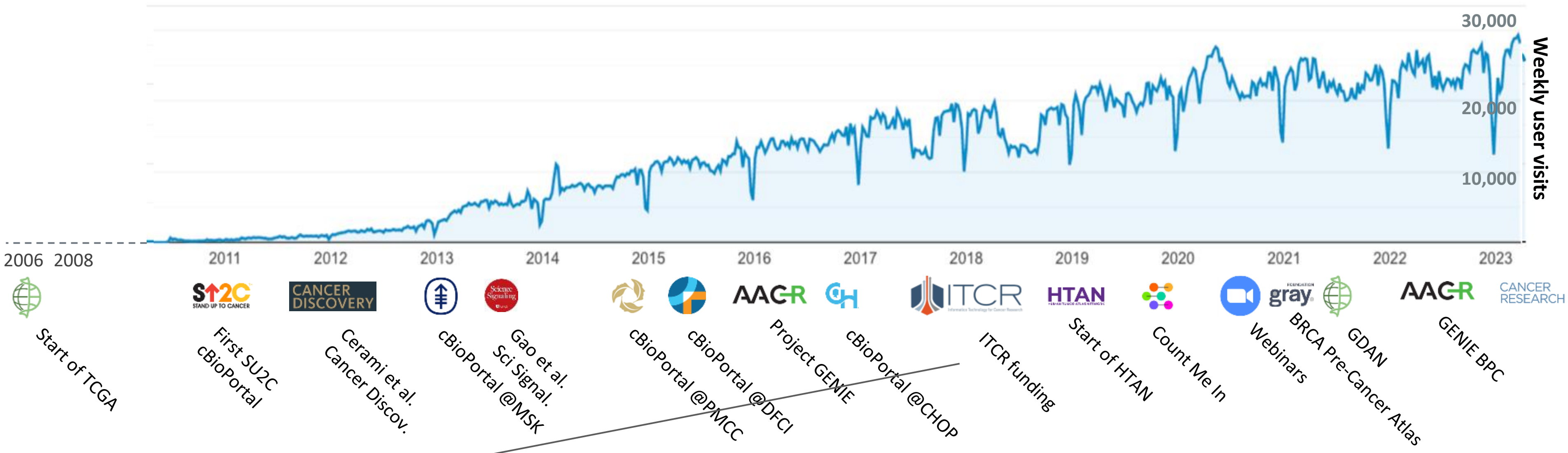
cBioPortal over time



CHOP decides to use cBioPortal and joins contributor community

cBioPortal over time

cBioPortal
FOR CANCER GENOMICS

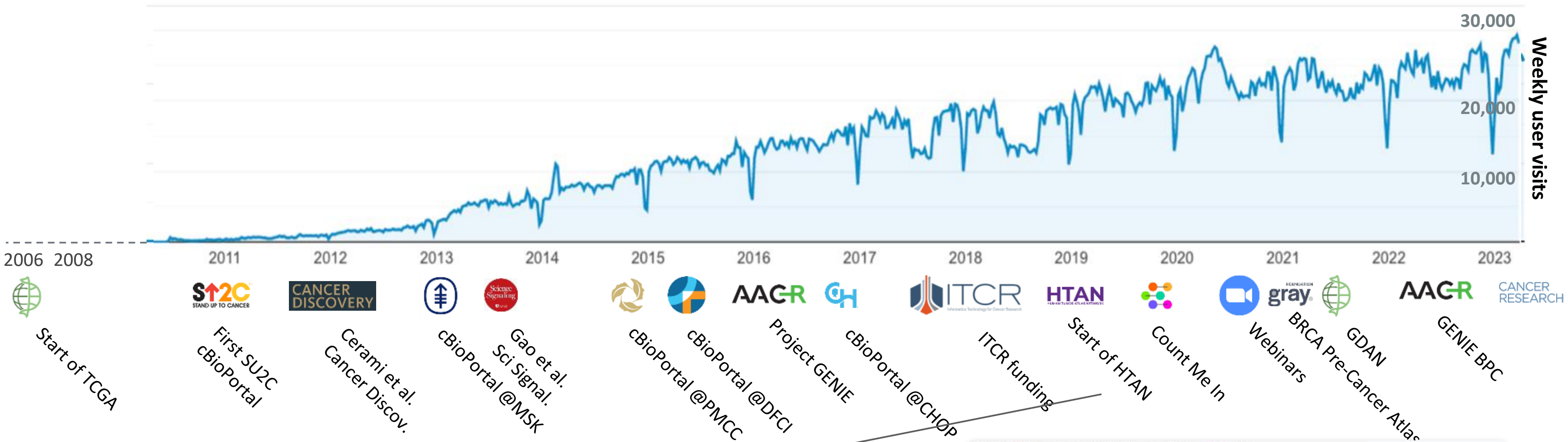


ITCR funding enables many enhancements to cBioPortal:

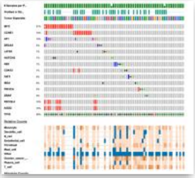
Evolve Core Platform (Scalability, Stability, Interoperability)
New Features (New Molecular Data Types, New Analyses)
Build Open Source Community (Governance, Local Installations)
Outreach and Training

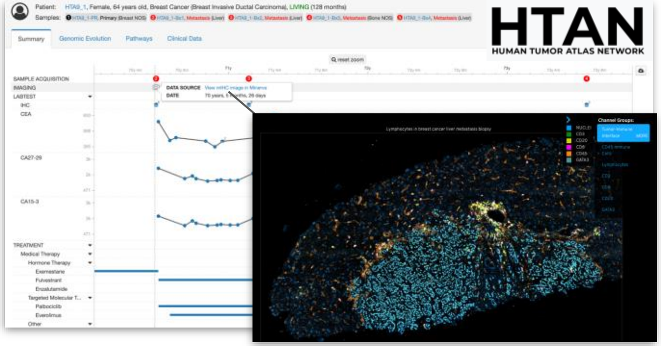
cBioPortal over time

cBioPortal FOR CANCER GENOMICS



Use cBioPortal for sharing **HTAN Data** with focus on single cell and imaging



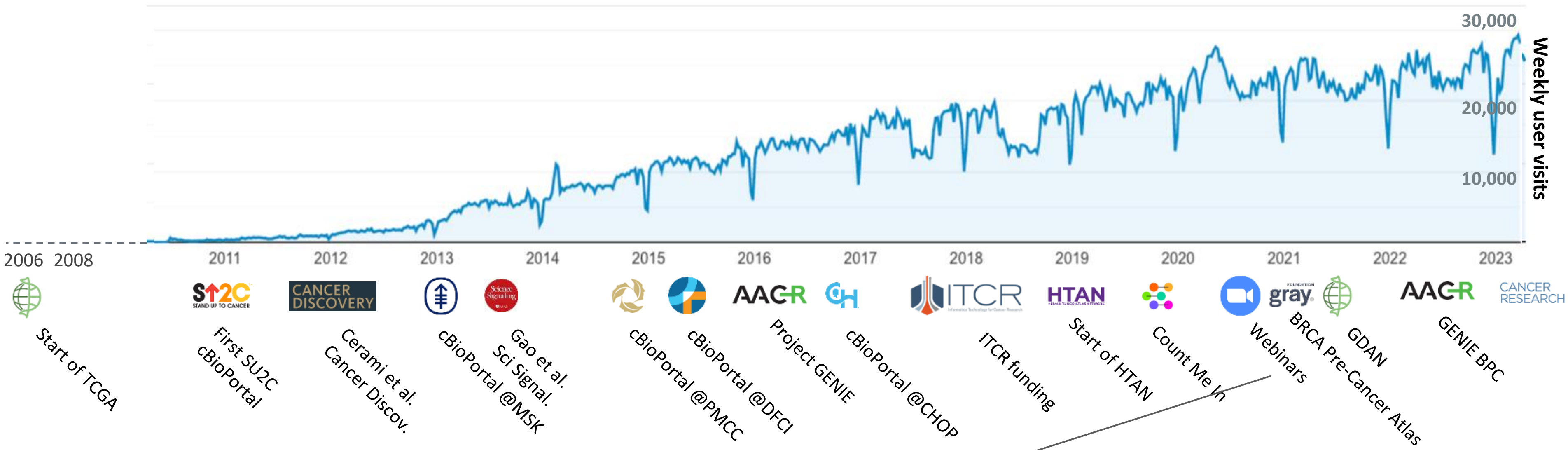


Patient: HTA9_1, Female, 64 years old, Breast Cancer (Breast Invasive Ductal Carcinoma), LIVING (128 months)

Samples: HTA9_1-PR, Primary (Breast NOS) HTA9_1-Bx1, Metastasis (Liver) HTA9_1-Bx2, Metastasis (Liver) HTA9_1-Bx3, Metastasis (Bone NOS) HTA9_1-Bx4, Metastasis (Liver)

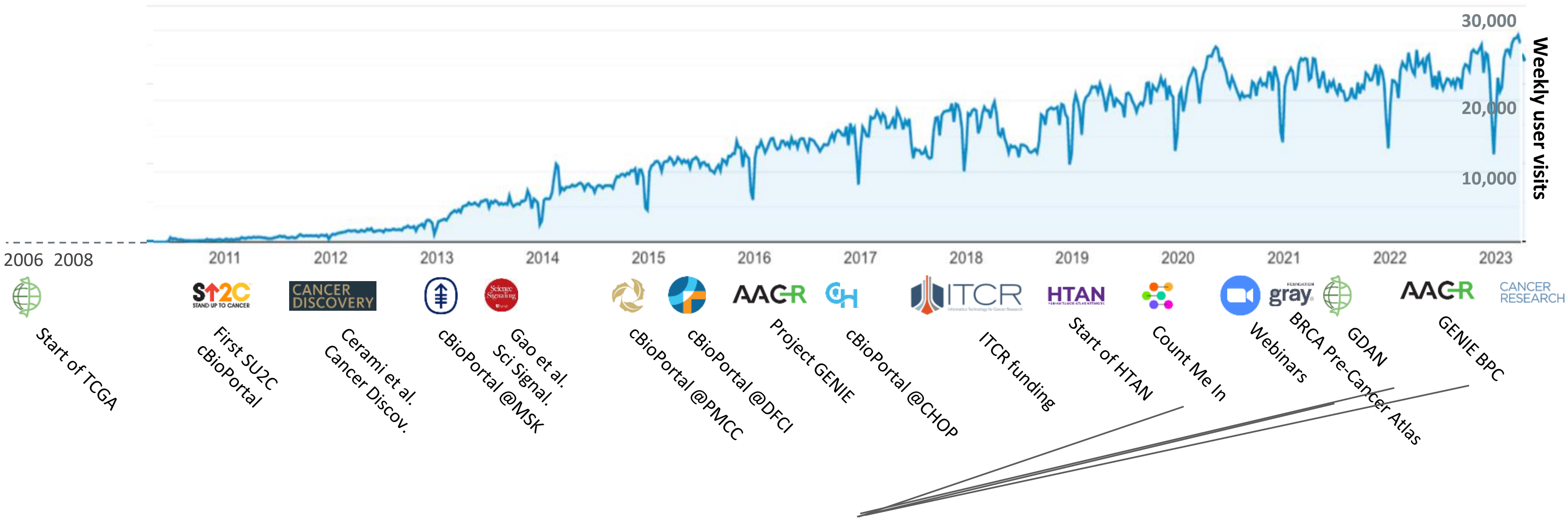
DATA SOURCE View mIHC image in Minerva
DATE 70 years, 5 months, 26 days

cBioPortal over time



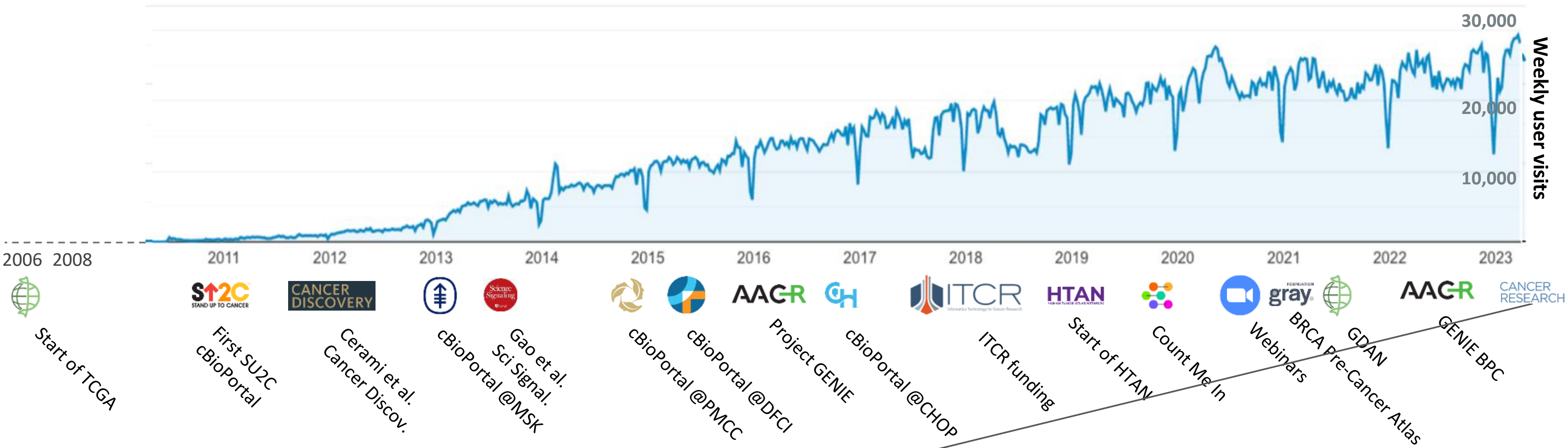
Webinars with >1K attendees
Videos viewed >100K times

cBioPortal over time



Use of cBioPortal for sharing data by several different **consortia**

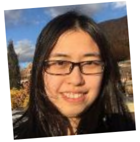
cBioPortal over time



New cBioPortal article with focus on AACR Project GENIE BPC
Joint Longitudinal Clinical and Genomic Data Analysis
Cancer Res. 2023.

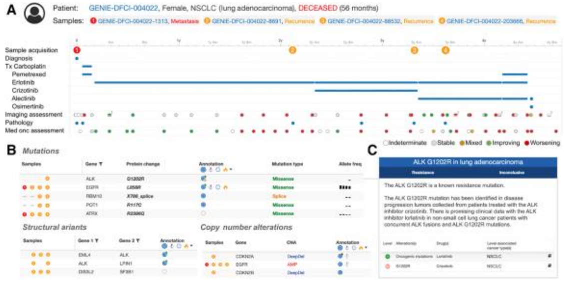

Ino


Gaofei


Xiang


Onur


JJ

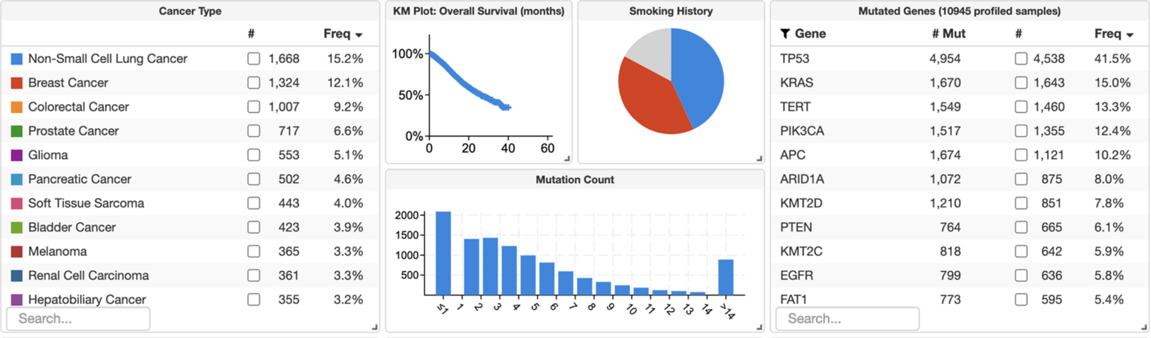


The ALK G1202R is a known resistance mutation.

The ALK G1202R mutation has been identified in disease progression tumors collected from patients treated with the ALK inhibitor crizotinib. There is promising clinical data with the ALK inhibitor lorlatinib in non-small cell lung cancer patients with concurrent ALK fusions and ALK G1202R mutations.

Four main cBioPortal views

Study view: Cohort exploration



Cancer Type	#	Freq ▾
🟦 Non-Small Cell Lung Cancer	☐ 1,668	15.2%
🟥 Breast Cancer	☐ 1,324	12.1%
🟧 Colorectal Cancer	☐ 1,007	9.2%
🟩 Prostate Cancer	☐ 717	6.6%
🟪 Glioma	☐ 553	5.1%
🟦 Pancreatic Cancer	☐ 502	4.6%
🟥 Soft Tissue Sarcoma	☐ 443	4.0%
🟩 Bladder Cancer	☐ 423	3.9%
🟥 Melanoma	☐ 365	3.3%
🟦 Renal Cell Carcinoma	☐ 361	3.3%
🟪 Hepatobiliary Cancer	☐ 355	3.2%

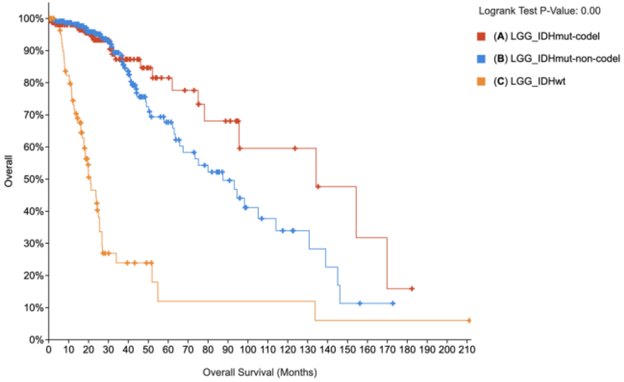
Search...

Mutated Genes (10945 profiled samples)

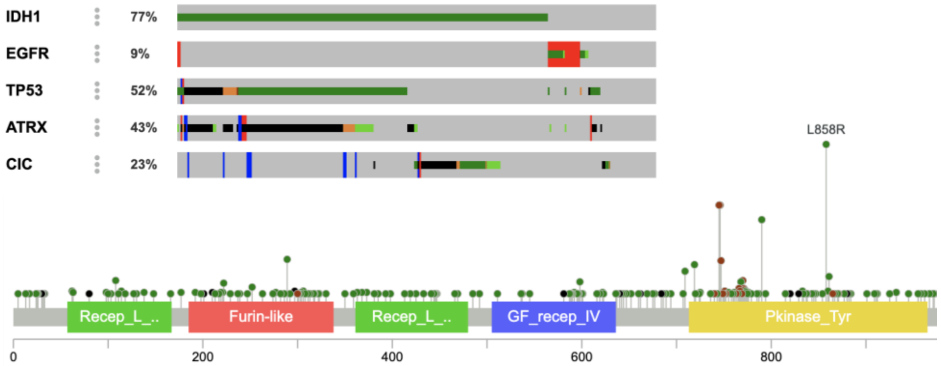
Gene	# Mut	#	Freq ▾
TP53	4,954	☐ 4,538	41.5%
KRAS	1,670	☐ 1,643	15.0%
TERT	1,549	☐ 1,460	13.3%
PIK3CA	1,517	☐ 1,355	12.4%
APC	1,674	☐ 1,121	10.2%
ARID1A	1,072	☐ 875	8.0%
KMT2D	1,210	☐ 851	7.8%
PTEN	764	☐ 665	6.1%
KMT2C	818	☐ 642	5.9%
EGFR	799	☐ 636	5.8%
FAT1	773	☐ 595	5.4%

Search...

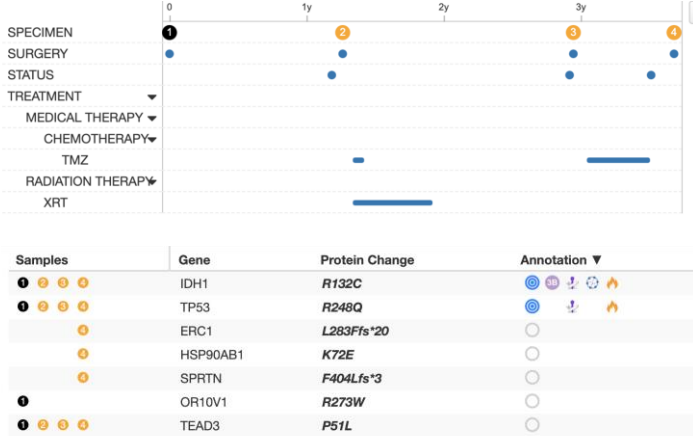
Group comparison



Results View: Gene-centric queries



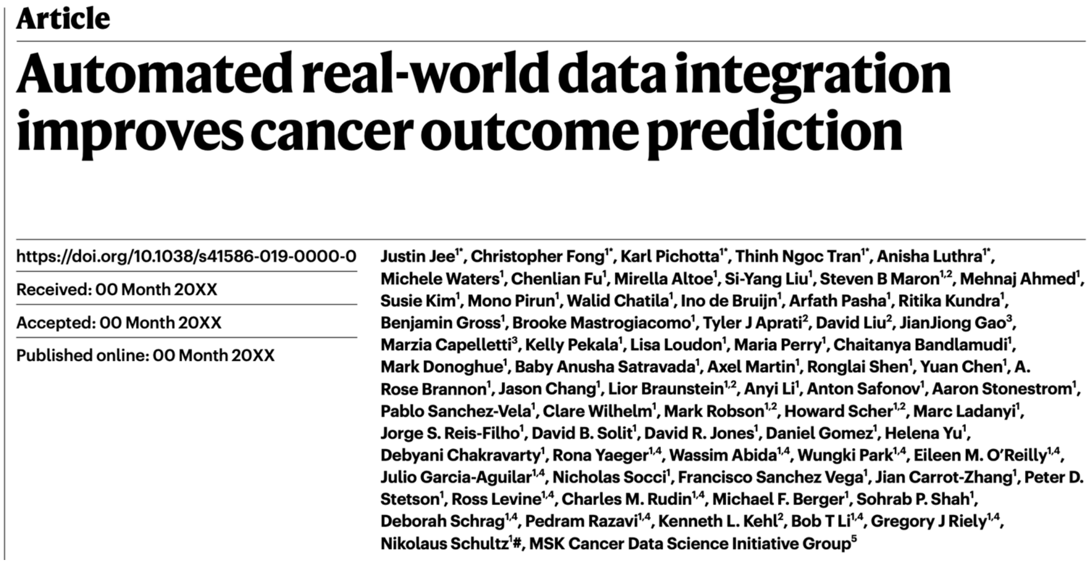
Patient View: Genomic and clinical timeline



Samples	Gene	Protein Change	Annotation ▼
❶ ② ③ ④	IDH1	***R132C***	◎ 3B
❶ ② ③ ④	TP53	***R248Q***	◎
④	ERC1	***L283Ffs*20***	○
④	HSP90AB1	***K72E***	○
④	SPRTN	***F404Lfs*3***	○
❶	OR10V1	***R273W***	○
❶ ② ③ ④	TEAD3	***P51L***	○

MSK-CHORD

Cohort of 25,000 patients, manuscript in press in *Nature*



Article

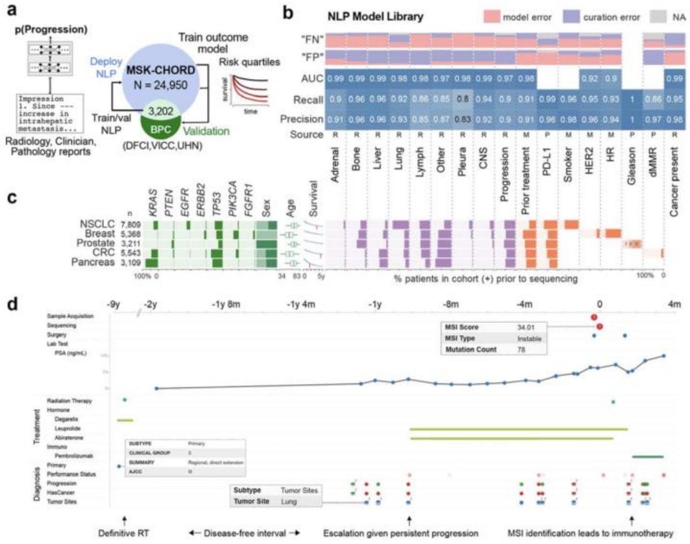
Automated real-world data integration improves cancer outcome prediction

https://doi.org/10.1038/s41586-019-0000-0

Received: 00 Month 20XX

Accepted: 00 Month 20XX

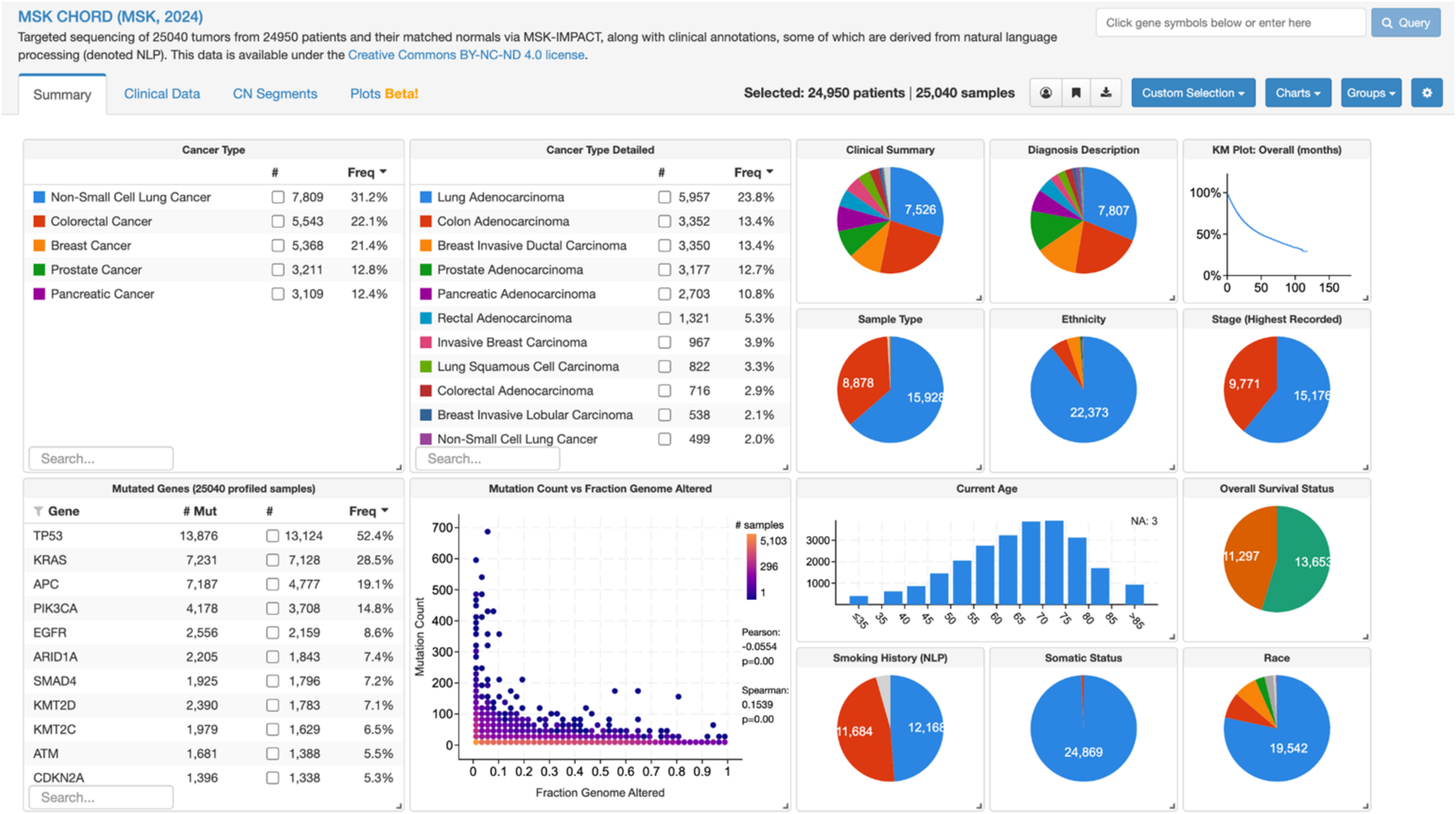
Justin Jee[1*], Christopher Fong[1*], Karl Pichotta[1*], Thinh Ngoc Tran[1*], Anisha Luthra[1*], Michele Waters[1], Chenlian Fu[1], Mirella Altoe[1], Si-Yang Liu[1], Steven B Maron[1,2], Mehnaj Ahmed[1], Susie Kim[1], Mono Pirun[1], Walid Chatila[1], Ino de Bruijn[1], Arfath Pasha[1], Ritika Kundra[1], Benjamin Gross[1], Brooke Mastrogiacomo[1], Tyler J Aprati[2], David Liu[2], JianJiong Gao[3], Marzia Capelletti[3], Kelly Pekala[1], Lisa Loudon[1], Maria Perry[1], Chaitanya Bandlamudi[1], Mark Donoghue[1], Baby Anusha Satravada[1], Axel Martin[1], Ronglai Shen[1], Yuan Chen[1], A. Rose Brannon[1], Jason Chang[1], Lior Braunstein[1,2], Anyi Li[1], Anton Safonov[1], Aaron Stonestrom[1], Pablo Sanchez-Vela[1], Clare Wilhelm[1], Mark Robson[1,2], Howard Scher[1,2], Marc Ladanyi[1], Jorge S. Reis-Filho[1], David B. Solit[1], David R. Jones[1], Daniel Gomez[1], Helena Yu[1], Debyani Chakravarty[1], Rona Yaeger[1,4], Wassim Abida[1,4], Wungki Park[1,4], Eileen M. O'Reilly[1,4], Julio Garcia-Aguilar[1,4], Nicholas Socci[1], Francisco Sanchez Vega[1], Jian Carrot-Zhang[1], Peter D. Stetson[1], Ross Levine[1,4], Charles M. Rudin[1,4], Michael F. Berger[1], Sohrab P. Shah[1], Deborah Schrag[1,4], Pedram Razavi[1,4], Kenneth L. Kehl[2], Bob T Li[1,4], Gregory J Riely[1,4], Nikolaus Schultz[1]#, MSK Cancer Data Science Initiative Group[5]



25,000 tumor samples across five cancer types, available via cBioPortal

Clinical data derived via NLP pipelines, outcome models

Within MSK: automated data updates, cohort of >90,000 patients



MSK CHORD (MSK, 2024)

Targeted sequencing of 25040 tumors from 24950 patients and their matched normals via MSK-IMPACT, along with clinical annotations, some of which are derived from natural language processing (denoted NLP). This data is available under the Creative Commons BY-NC-ND 4.0 license.

Selected: 24,950 patients | 25,040 samples

Cancer Type

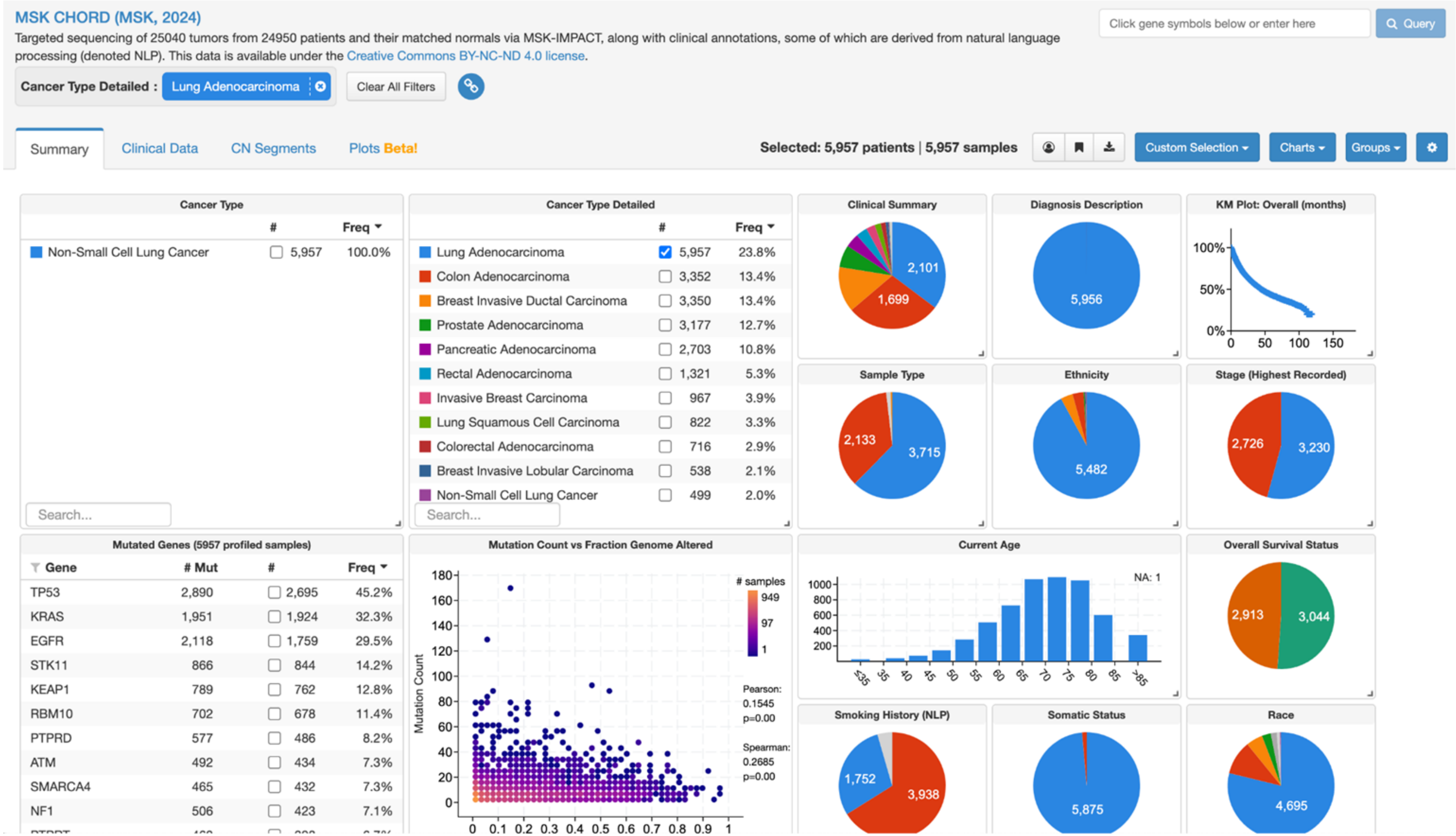
	#	Freq
🟦 Non-Small Cell Lung Cancer	☐ 7,809	31.2%
🟥 Colorectal Cancer	☐ 5,543	22.1%
🟧 Breast Cancer	☐ 5,368	21.4%
🟩 Prostate Cancer	☐ 3,211	12.8%
🟪 Pancreatic Cancer	☐ 3,109	12.4%

Cancer Type Detailed

	#	Freq
🟦 Lung Adenocarcinoma	☐ 5,957	23.8%
🟥 Colon Adenocarcinoma	☐ 3,352	13.4%
🟧 Breast Invasive Ductal Carcinoma	☐ 3,350	13.4%
🟩 Prostate Adenocarcinoma	☐ 3,177	12.7%
🟪 Pancreatic Adenocarcinoma	☐ 2,703	10.8%
🟦 Rectal Adenocarcinoma	☐ 1,321	5.3%
🟥 Invasive Breast Carcinoma	☐ 967	3.9%
🟩 Lung Squamous Cell Carcinoma	☐ 822	3.3%
🟥 Colorectal Adenocarcinoma	☐ 716	2.9%
🟦 Breast Invasive Lobular Carcinoma	☐ 538	2.1%
🟪 Non-Small Cell Lung Cancer	☐ 499	2.0%

Mutated Genes (25040 profiled samples)

Gene	# Mut	#	Freq
TP53	13,876	☐ 13,124	52.4%
KRAS	7,231	☐ 7,128	28.5%
APC	7,187	☐ 4,777	19.1%
PIK3CA	4,178	☐ 3,708	14.8%
EGFR	2,556	☐ 2,159	8.6%
ARID1A	2,205	☐ 1,843	7.4%
SMAD4	1,925	☐ 1,796	7.2%
KMT2D	2,390	☐ 1,783	7.1%
KMT2C	1,979	☐ 1,629	6.5%
ATM	1,681	☐ 1,388	5.5%
CDKN2A	1,396	☐ 1,338	5.3%



MSK CHORD (MSK, 2024)

Targeted sequencing of 25040 tumors from 24950 patients and their matched normals via MSK-IMPACT, along with clinical annotations, some of which are derived from natural language processing (denoted NLP). This data is available under the Creative Commons BY-NC-ND 4.0 license.

Cancer Type Detailed : Lung Adenocarcinoma

Selected: 5,957 patients | 5,957 samples

Cancer Type

	#	Freq
Non-Small Cell Lung Cancer	☐ 5,957	100.0%

Cancer Type Detailed

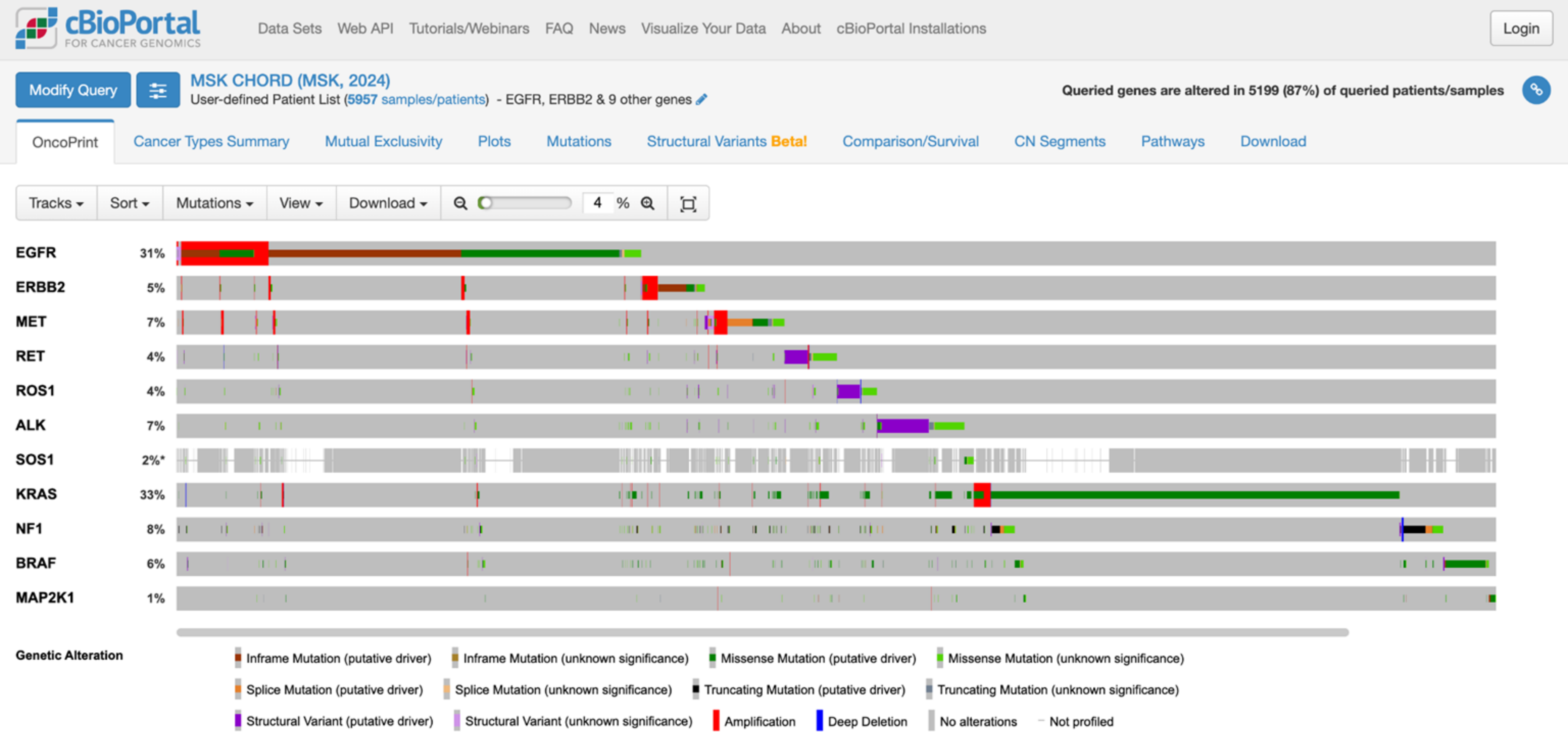
	#	Freq
Lung Adenocarcinoma	☑ 5,957	23.8%
Colon Adenocarcinoma	☐ 3,352	13.4%
Breast Invasive Ductal Carcinoma	☐ 3,350	13.4%
Prostate Adenocarcinoma	☐ 3,177	12.7%
Pancreatic Adenocarcinoma	☐ 2,703	10.8%
Rectal Adenocarcinoma	☐ 1,321	5.3%
Invasive Breast Carcinoma	☐ 967	3.9%
Lung Squamous Cell Carcinoma	☐ 822	3.3%
Colorectal Adenocarcinoma	☐ 716	2.9%
Breast Invasive Lobular Carcinoma	☐ 538	2.1%
Non-Small Cell Lung Cancer	☐ 499	2.0%

Mutated Genes (5957 profiled samples)

Gene	# Mut	#	Freq
TP53	2,890	☐ 2,695	45.2%
KRAS	1,951	☐ 1,924	32.3%
EGFR	2,118	☐ 1,759	29.5%
STK11	866	☐ 844	14.2%
KEAP1	789	☐ 762	12.8%
RBM10	702	☐ 678	11.4%
PTPRD	577	☐ 486	8.2%
ATM	492	☐ 434	7.3%
SMARCA4	465	☐ 432	7.3%
NF1	506	☐ 423	7.1%

Pearson: 0.1545 p=0.00

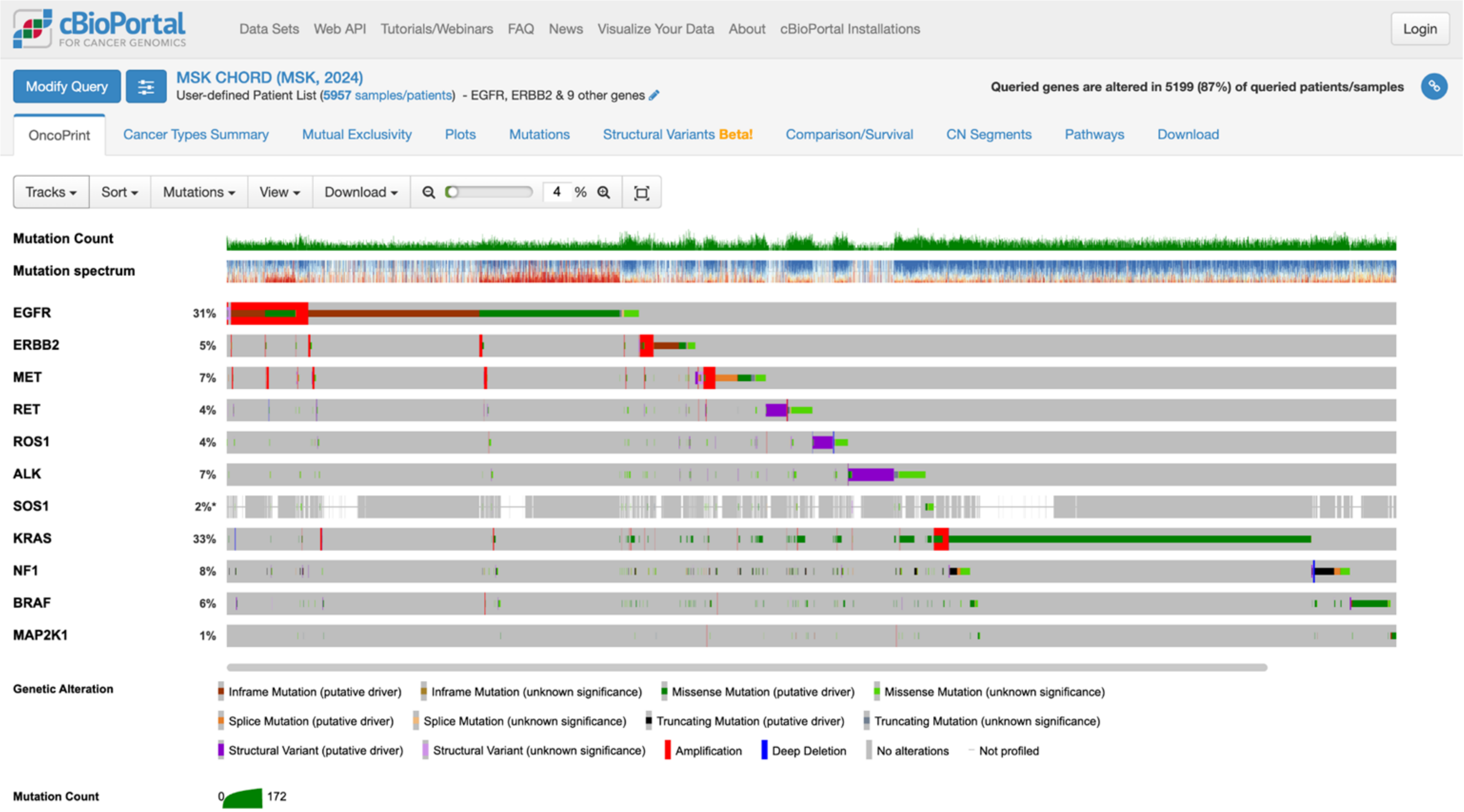
Spearman: 0.2685 p=0.00



MSK CHORD (MSK, 2024)

User-defined Patient List (5957 samples/patients) - EGFR, ERBB2 & 9 other genes

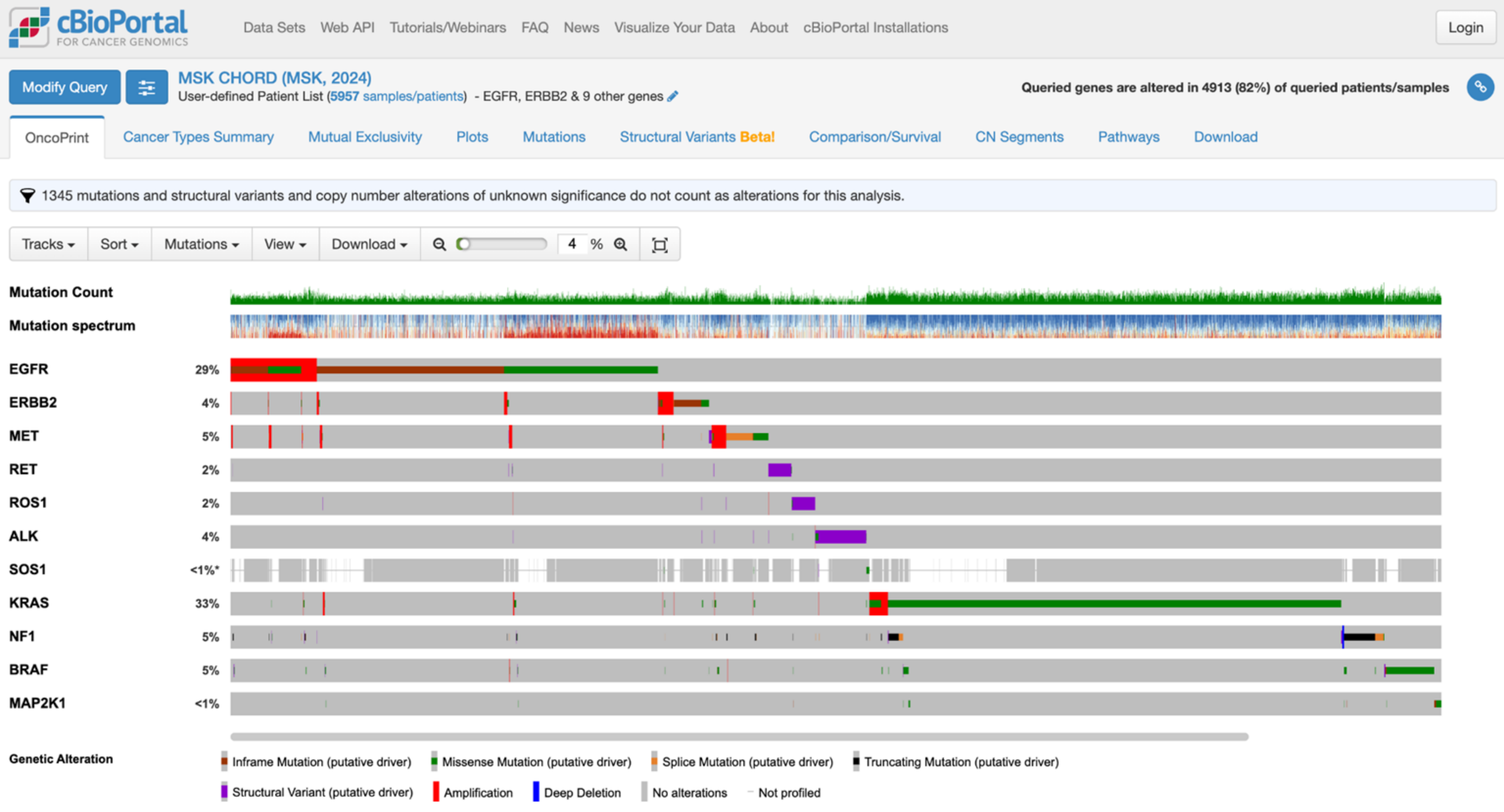
Queried genes are altered in 5199 (87%) of queried patients/samples



MSK CHORD (MSK, 2024)

User-defined Patient List (5957 samples/patients) - EGFR, ERBB2 & 9 other genes

Queried genes are altered in 5199 (87%) of queried patients/samples



MSK CHORD (MSK, 2024)

User-defined Patient List (5957 samples/patients) - EGFR, ERBB2 & 9 other genes

Queried genes are altered in 4913 (82%) of queried patients/samples

1345 mutations and structural variants and copy number alterations of unknown significance do not count as alterations for this analysis.

EGFR	29%
ERBB2	4%
MET	5%
RET	2%
ROS1	2%
ALK	4%
SOS1	<1%*
KRAS	33%
NF1	5%
BRAF	5%
MAP2K1	<1%

MSK CHORD (MSK, 2024)

User-defined Patient List (5957 samples/patients) - EGFR, ERBB2 & 9 other genes

Queried genes are altered in 4913 (82%) of queried patients/samples

168 mutations of unknown significance are shown below.

NM_005228 | ENST00000275493
CCDS5514 | EGFR_HUMAN

Somatic Mutation Frequency 29.7%

	Driver	VUS
	1951	186
Missense	1038	143
Truncating	0	11
Inframe	912	0
Splice	0	14
Fusion	1	18

2137 Mutations: includes 1 duplicate mutation in patients with multiple samples (page 1 of 86)

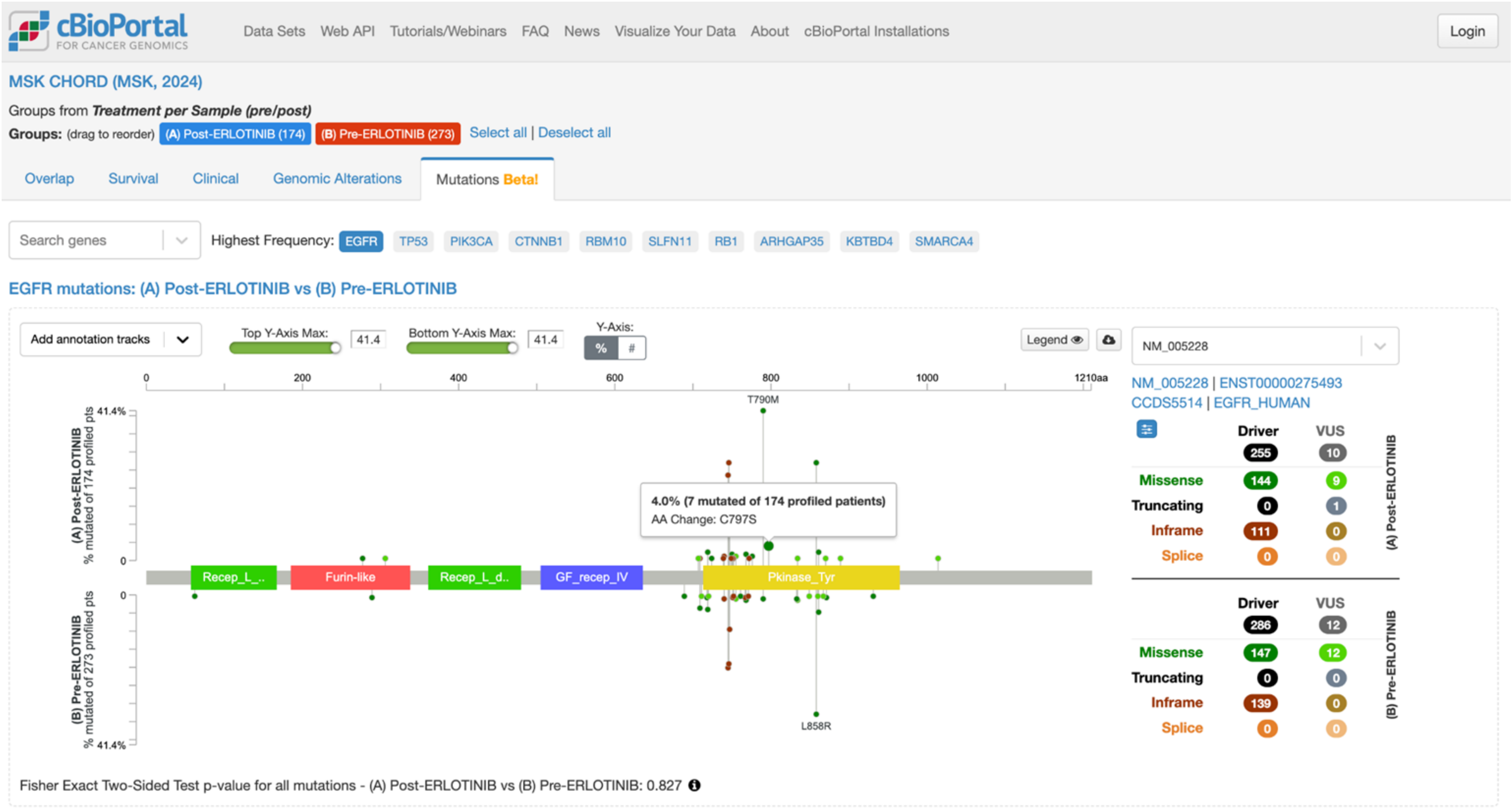
Sample ID	Protein Change	Annotation	Mutation Type	Copy #	COSMIC	Allele Freq (T)	# Mut in Sample
P-0000583-T01-IM3	T790M		Missense	Diploid	355	0.46	10
P-0000580-T01-IM3	T790M		Missense	Diploid	355	0.13	10
P-0000774-T01-IM3	T790M		Missense	Diploid	355	0.23	10
P-0003425-T01-IM5	T790M		Missense	Amp	355	0.04	5

Example patient with lung adenocarcinoma

Patient: P-0015921, FEMALE, Non-Small Cell Lung Cancer (Lung Adenocarcinoma), DECEASED (71 months)

Samples: ① P-0015921-T02-IM6, Metastasis (Liver) ② P-0015921-T01-IM6, Primary (Lung) ③ P-0015921-T03-IM6, Primary (Lung) ④ P-0015921-T08-IM6, Metastasis (Lymph node) ⑤ P-0015921-T09-IM7 ⑥ P-0015921-T11-XS1, cfDNA

EGFR exon 19 deletion → T790M mutation → ALK fusion → ALK G1202 resistance mutation + several others in cfDNA



MSK CHORD (MSK, 2024)

Groups from *Treatment per Sample (pre/post)*

Groups: (drag to reorder) (A) Post-ERLOTINIB (174) (B) Pre-ERLOTINIB (273) Select all | Deselect all

EGFR mutations: (A) Post-ERLOTINIB vs (B) Pre-ERLOTINIB

Top Y-Axis Max: 41.4

Bottom Y-Axis Max: 41.4

Y-Axis: % #

NM_005228

NM_005228 | ENST00000275493
CCDS5514 | EGFR_HUMAN

4.0% (7 mutated of 174 profiled patients)
AA Change: C797S

(A) Post-ERLOTINIB

	Driver	VUS
	255	10
Missense	144	9
Truncating	0	1
Inframe	111	0
Splice	0	0

(B) Pre-ERLOTINIB

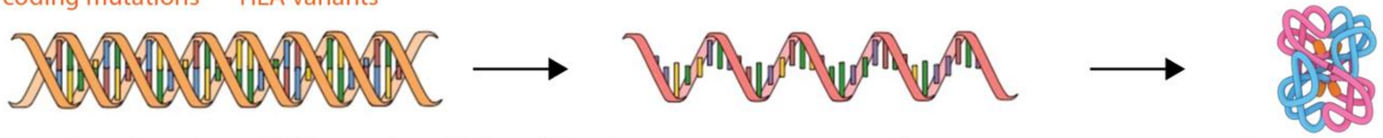
	Driver	VUS
	286	12
Missense	147	12
Truncating	0	0
Inframe	139	0
Splice	0	0

Fisher Exact Two-Sided Test p-value for all mutations - (A) Post-ERLOTINIB vs (B) Pre-ERLOTINIB: 0.827

cBioPortal Roadmap

Add New Data Types & Features

currently supported
enhancements planned
new data types



mutations and genotypes
non-synonymous somatic mutations
non-synonymous germline mutations
mutational signatures
non-coding mutations HLA variants

RNA expression levels
mRNA expression levels in tumor samples
single cell mRNA expression
mRNA expression levels in normal samples

copy-number alterations (CNA)
gene-specific CNA
single cell DNA copy-number
allele-specific CNA

other DNA modifications
DNA methylation
structural rearrangements / fusions
chromatin accessibility

other assays
microbiome
metabolite levels

protein measurements
protein levels
phosphoprotein levels
postranslational modifications

Improve Performance

ClickHouse

Open Source Community & Outreach

Google Summer of Code

ITCR internship

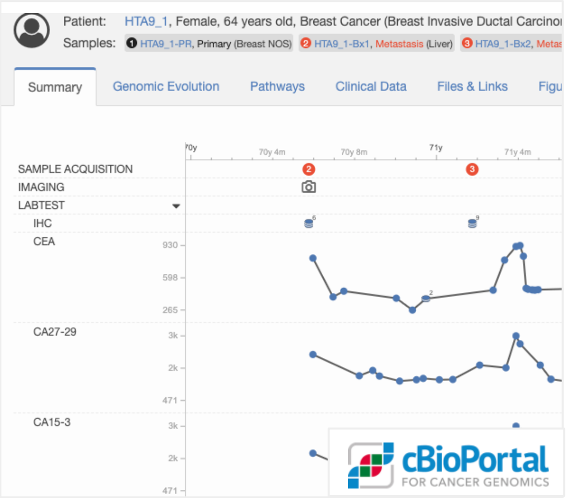
cBioPortal NEWSLETTER

Biannual cBioPortal Hackathons!

Visualization tools for different data modalities

cBioPortal

Multimodal Data

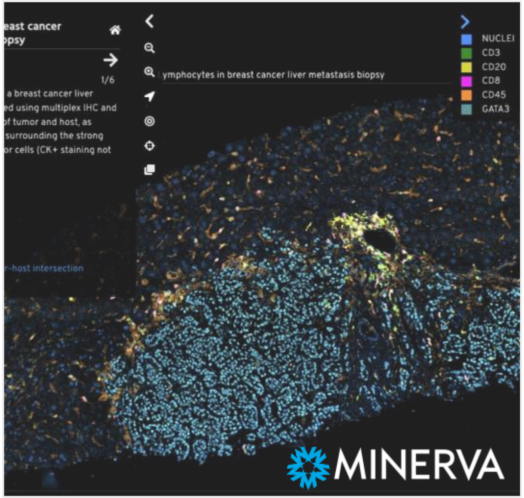


Patient: HTA9_1, Female, 64 years old, Breast Cancer (Breast Invasive Ductal Carcino

Samples: ❶ HTA9_1-PR, Primary (Breast NOS) ② HTA9_1-Bx1, Metastasis (Liver) ③ HTA9_1-Bx2, Metas

MINERVA

Multiplex Imaging Data



east cancer
opsy

1/6

a breast cancer liver
ed using multiplex IHC and
of tumor and host, as
surrounding the strong
or cells (CK+ staining not

r-host intersection

ymphocytes in breast cancer liver metastasis biopsy

CELLxGENE

Single Cell Data

Developed By

Memorial Sloan Kettering Cancer Center

UHN Toronto General Toronto Western Princess Margaret Toronto Rehab

CARIS LIFE SCIENCES

Dana-Farber Cancer Institute

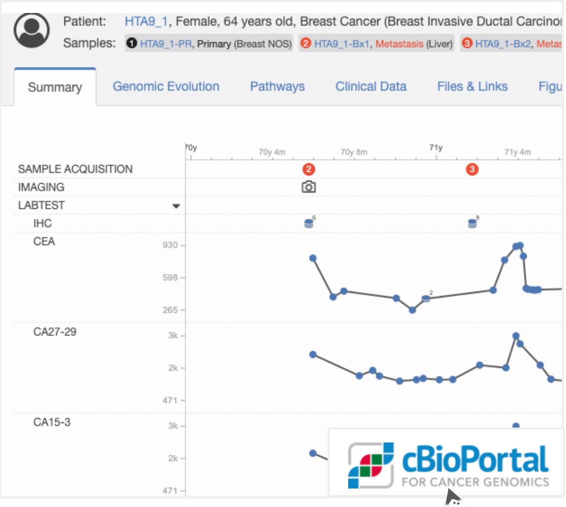
The Children's Hospital of Philadelphia

thehyve

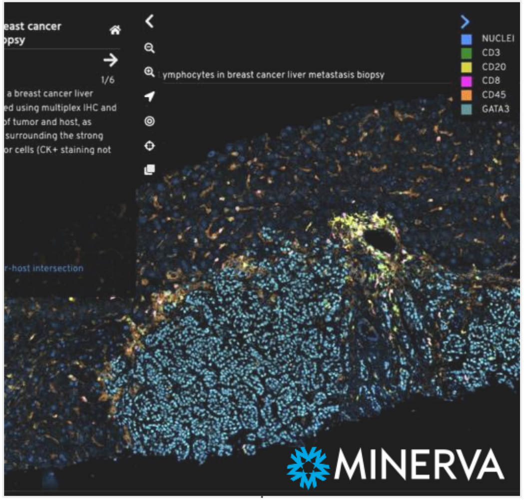
Laboratory of Systems Pharmacology

Visualization tools for different data modalities

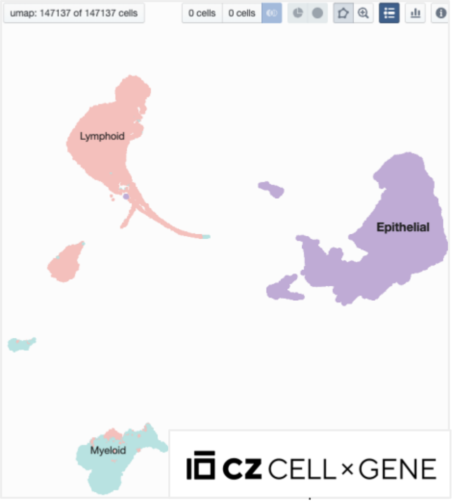
cBioPortal
Multimodal Data



MINERVA
Multiplex Imaging Data



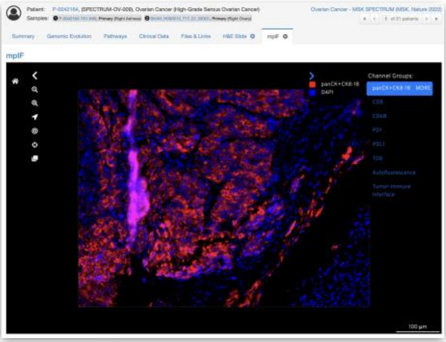
CELLxGENE
Single Cell Data

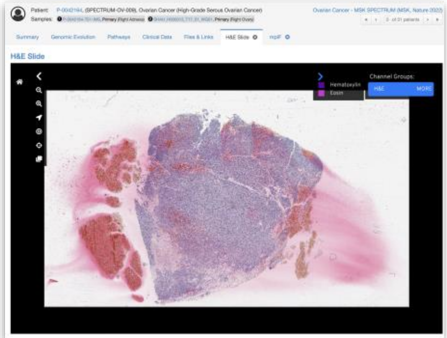


Tools Integration & Joint analysis on derived data elements

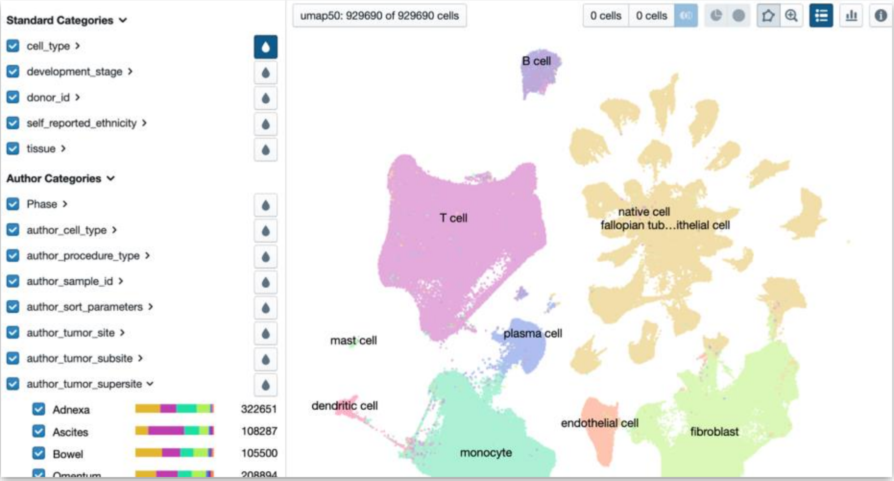
MSK SPECTRUM

Multiplex imaging and H&E (Minerva)


Patient: P-0042164, (SPECTRUM-OV-009), Ovarian Cancer (High-Grade Serous Ovarian Cancer)

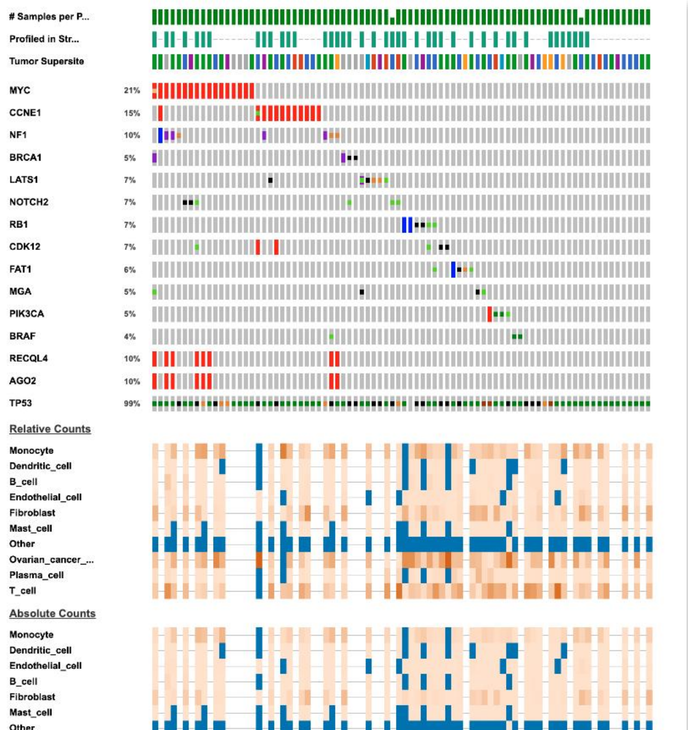

Patient: P-0042164, (SPECTRUM-OV-009), Ovarian Cancer (High-Grade Serous Ovarian Cancer)

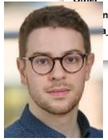
Single Cell Data (CellXGene)



cbioportal.org/study/summary?id=msk_spectrum_tme_2022

Oncoprints combining WGS, MSK-IMPACT, scRNASeq (cBioPortal)

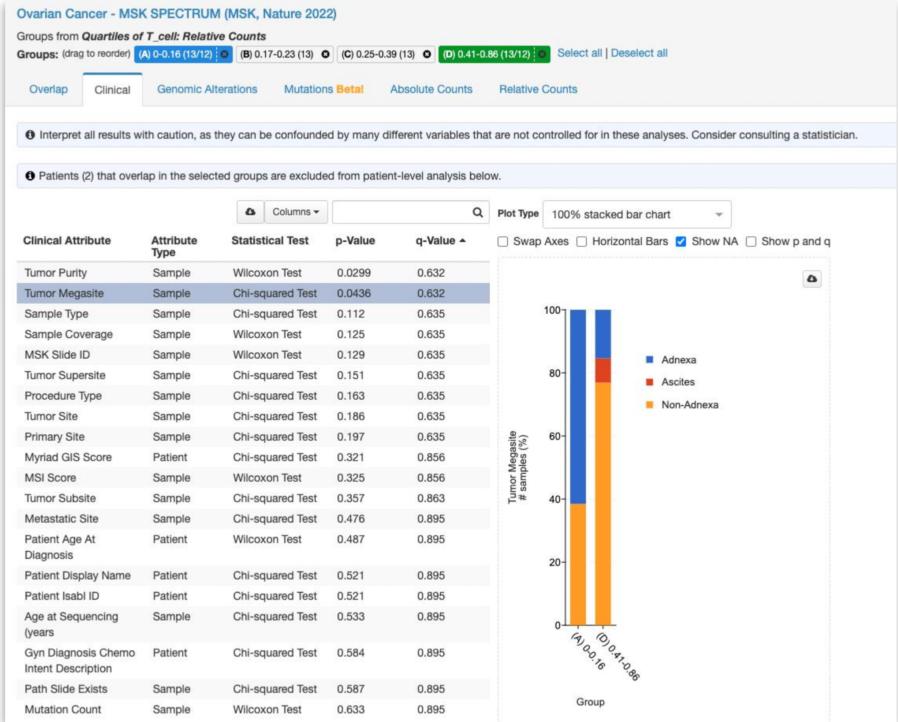



Ignacio
Vázquez-García


Sohrab
Shah

Compare Cell Types across samples

T-cell high vs low samples



Ovarian Cancer - MSK SPECTRUM (MSK, Nature 2022)

Groups from ***Quartiles of T_cell: Relative Counts***

Groups: (drag to reorder) (A) 0-0.16 (13/12) (B) 0.17-0.23 (13) (C) 0.25-0.39 (13) (D) 0.41-0.86 (13/12) Select all | Deselect all

Interpret all results with caution, as they can be confounded by many different variables that are not controlled for in these analyses. Consider consulting a statistician.

Patients (2) that overlap in the selected groups are excluded from patient-level analysis below.

Clinical Attribute	Attribute Type	Statistical Test	p-Value	q-Value ▲
Tumor Purity	Sample	Wilcoxon Test	0.0299	0.632
Tumor Megasite	Sample	Chi-squared Test	0.0436	0.632
Sample Type	Sample	Chi-squared Test	0.112	0.635
Sample Coverage	Sample	Wilcoxon Test	0.125	0.635
MSK Slide ID	Sample	Wilcoxon Test	0.129	0.635
Tumor Supersite	Sample	Chi-squared Test	0.151	0.635
Procedure Type	Sample	Chi-squared Test	0.163	0.635
Tumor Site	Sample	Chi-squared Test	0.186	0.635
Primary Site	Sample	Chi-squared Test	0.197	0.635
Myriad GIS Score	Patient	Chi-squared Test	0.321	0.856
MSI Score	Sample	Wilcoxon Test	0.325	0.856
Tumor Subsite	Sample	Chi-squared Test	0.357	0.863
Metastatic Site	Sample	Chi-squared Test	0.476	0.895
Patient Age At Diagnosis	Patient	Wilcoxon Test	0.487	0.895
Patient Display Name	Patient	Chi-squared Test	0.521	0.895
Patient Isabl ID	Patient	Chi-squared Test	0.521	0.895
Age at Sequencing (years	Sample	Chi-squared Test	0.533	0.895
Gyn Diagnosis Chemo Intent Description	Patient	Chi-squared Test	0.584	0.895
Path Slide Exists	Sample	Chi-squared Test	0.587	0.895
Mutation Count	Sample	Wilcoxon Test	0.633	0.895

Plot Type: 100% stacked bar chart

☐ Swap Axes ☐ Horizontal Bars ☑ Show NA ☐ Show p and q

Compare cell types across sites

Ovarian Cancer - MSK SPECTRUM (MSK, Nature 2022)

Groups from ***Tumor Supersite***

Groups: (drag to reorder) (A) Adnexa (28/22) (B) Ascites (2) (C) Bowel (7) (D) Omentum (22/18) (E) Other (11/8) (F) Peritoneum (7) (G) Upper Quadrant (5/4)

Data Set: Relative Counts

p-Value: 0.144

q-Value: 0.479

High in ... ☐ Significant only

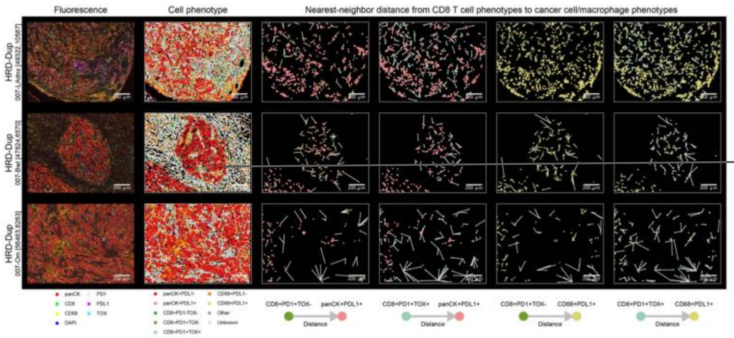
Relative Counts	μ in (A) Adnexa	μ in (B) Ascites	μ in (C) Bowel	μ in (D) Omentum	μ in (E) Other	μ in (F) Peritoneum	μ in (G) Upper Quadrant	σ i
Dendritic_cell	0.00	0.02	0.00	0.00	0.00	0.01	0.00	0.0
B_cell	0.01	0.03	0.02	0.03	0.08	0.01	0.00	0.0
T_cell	0.21	0.38	0.29	0.30	0.31	0.49	0.19	0.1
Monocyte	0.26	0.22	0.28	0.13	0.15	0.18	0.25	0.1
Mast_cell	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
Ovarian_cancer_cell	0.34	0.33	0.21	0.29	0.26	0.10	0.44	0.2
Plasma_cell	0.02	0.00	0.01	0.04	0.03	0.05	0.00	0.0
Fibroblast	0.14	0.02	0.18	0.17	0.14	0.15	0.10	0.1

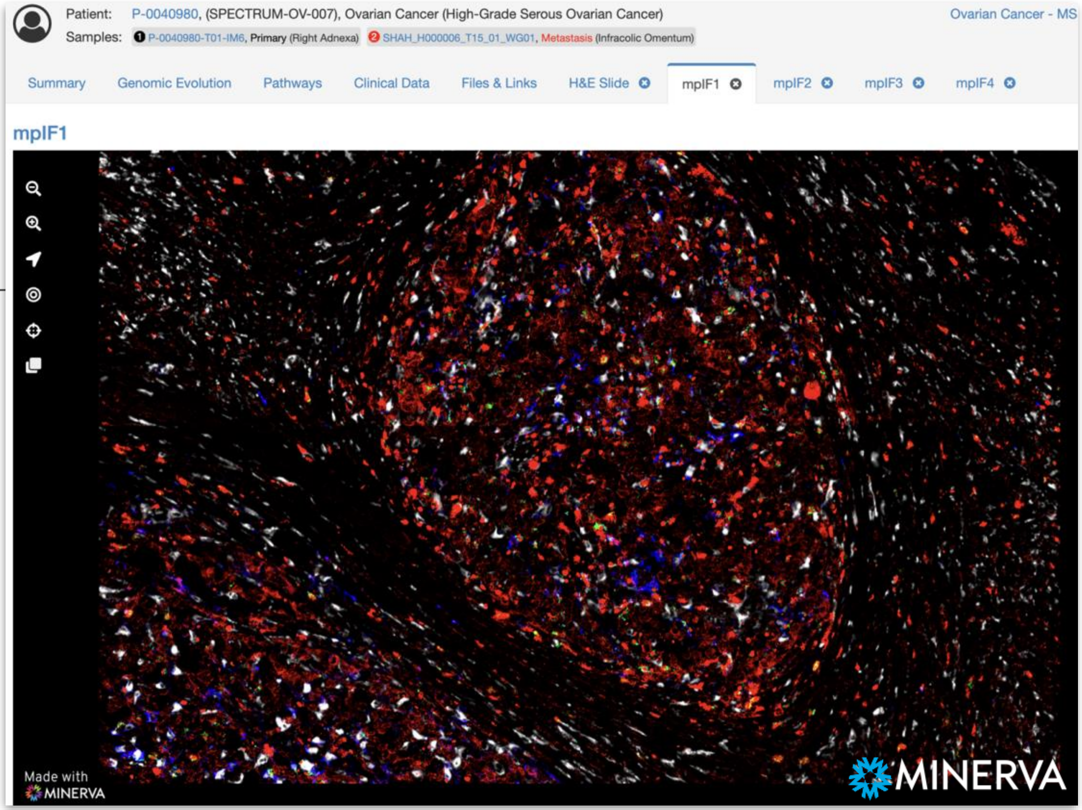
Showing 1-8 of 8

Explore Article Figures in more detail

Extended Data Fig. 13: Site-specific determinants of spatial interactions between cancer cells, T cells and macrophages.

From: Ovarian cancer mutational processes drive site-specific immune evasion





Patient: P-0040980, (SPECTRUM-OV-007), Ovarian Cancer (High-Grade Serous Ovarian Cancer)

Samples: ❶ P-0040980-T01-IM6, Primary (Right Adnexa) ❷ SHAH_H000006_T15_01_WG01, Metastasis (Infracolic Omentum)

mpIF1

cBioPortal and CRDC

- HTAN is submitting data to CDS

cBioPortal and CRDC

- HTAN is submitting data to CDS
- GDC data now available in cBioPortal for TCGA, CPTAC

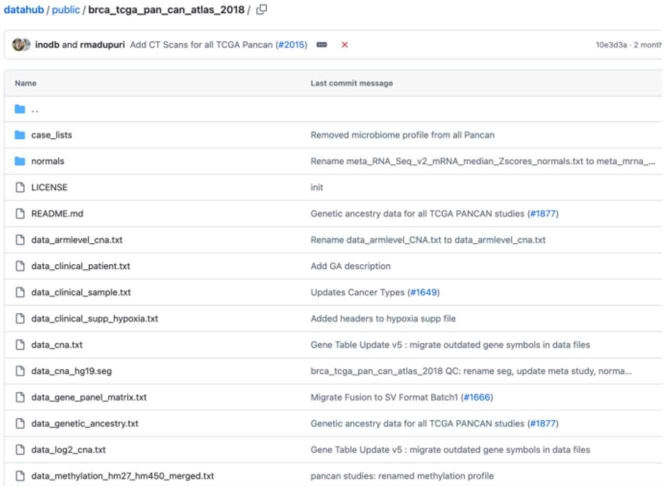
Breast

Invasive Breast Carcinoma

☐	Breast Invasive Carcinoma (TCGA, Cell 2015)	818 samples
☐	Breast Invasive Carcinoma (TCGA, Firehose Legacy)	1108 samples
☐	Breast Invasive Carcinoma (TCGA, Nature 2012)	825 samples
☐	Breast Invasive Carcinoma (TCGA, PanCancer Atlas)	1084 samples
☐	Invasive Breast Carcinoma (TCGA, GDC)	1103 samples

cBioPortal and CRDC

- HTAN is submitting data to CDS
- GDC data now available in cBioPortal for TCGA, CPTAC
- Future: Use CDS as “Datahub” for cBioPortal?
  - Currently using GitHub
  - Could develop an interactive interface for:
    - Browsing publicly available studies
    - Study upload
    - Discussion forum to discuss specific studies
    - Nominating publications for submission



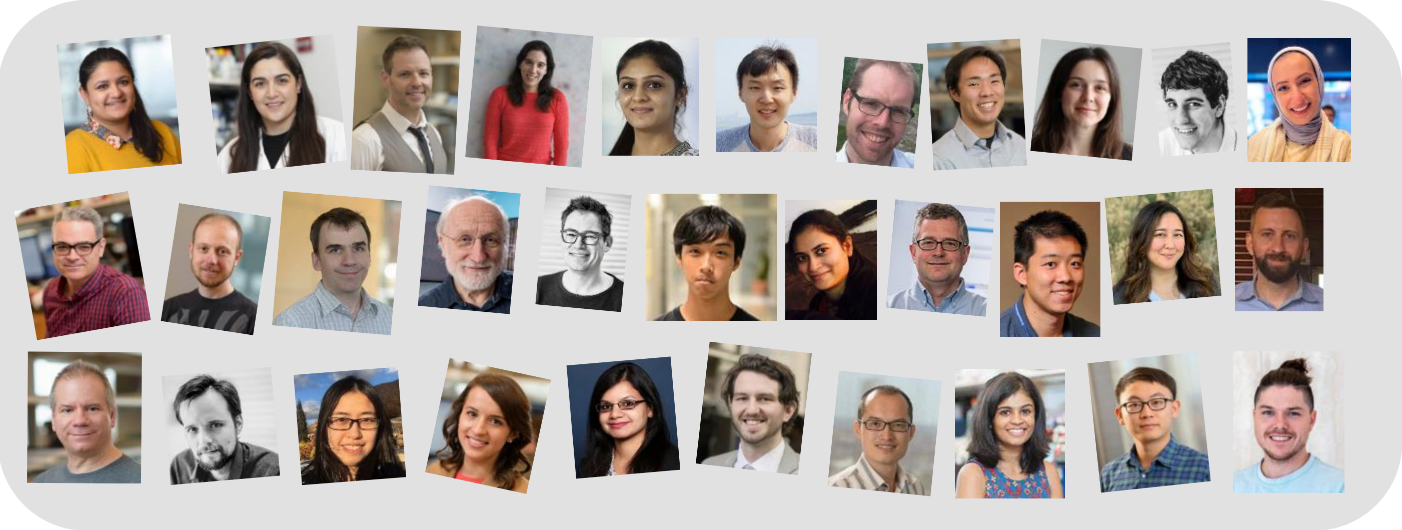
datahub / public / brca_tcga_pan_can_atlas_2018 /

inodb and **rmadupuri** Add CT Scans for all TCGA Pancan (#2015) ✕ 10e3d3a · 2 month

Name	Last commit message
..	
case_lists	Removed microbiome profile from all Pancan
normals	Rename meta_RNA_Seq_v2_mRNA_median_Zscores_normals.txt to meta_mrna_...
LICENSE	init
README.md	Genetic ancestry data for all TCGA PANCAN studies (#1877)
data_armlevel_cna.txt	Rename data_armlevel_CNA.txt to data_armlevel_cna.txt
data_clinical_patient.txt	Add GA description
data_clinical_sample.txt	Updates Cancer Types (#1649)
data_clinical_supp_hypoxia.txt	Added headers to hypoxia supp file
data_cna.txt	Gene Table Update v5 : migrate outdated gene symbols in data files
data_cna_hg19.seg	brca_tcga_pan_can_atlas_2018 QC: rename seg, update meta study, norma...
data_gene_panel_matrix.txt	Migrate Fusion to SV Format Batch1 (#1666)
data_genetic_ancestry.txt	Genetic ancestry data for all TCGA PANCAN studies (#1877)
data_log2_cna.txt	Gene Table Update v5 : migrate outdated gene symbols in data files
data_methylation_hm27_hm450_merged.txt	pancan studies: renamed methylation profile

cBioPortal Contributors and Funders

cBioPortal
FOR CANCER GENOMICS



ITCR
Informatics Technology for Cancer Research

CIViC
Clinical Interpretation of Variants in Cancer

MINERVA

Galaxy

igv

DSA

NDEx

NG-CHM

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

MusicaTK

Memorial Sloan Kettering Cancer Center

Dana-Farber Cancer Institute

UHN
Toronto General
Toronto Western
Princess Margaret
Toronto Rehab

Children's Hospital of Philadelphia

CARIS LIFE SCIENCES

The Hyve

HARVARD MEDICAL SCHOOL

SE4BIO

Bilkent University

NIH NATIONAL CANCER INSTITUTE

NCI CANCER MOONSHOT
Human Tumor Atlas Network
A program of the National Cancer Institute of the National Institutes of Health

BREAK THROUGH CANCER

Kravis Center for Molecular Oncology

FOUNDATION gray.

AACR
American Association for Cancer Research

Dana-Farber Cancer Institute

Prostate Cancer Foundation
Curing Together.

MOVEMBER FOUNDATION