Interactive Mining of TCGA Data

June 19-20, 2017

Anand S. Merchant

About cBioPortal

- The cBioPortal for Cancer Genomics was originally developed at [Memorial Sloan Kettering Cancer Center](#) (MSKCC). The primary architect for the tool is Dr. Chris Sanders, who was Chair of the Computational Biology Program and MSKCC at that time, and is now at the Dana Farber Cancer Institute.

- **The [public cBioPortal site](#) is hosted by the [Center for Molecular Oncology](#) at MSK.**

- The cBioPortal software is now available under an open source license via [GitHub](#).

- The software is now developed and maintained by a multi-institutional team, consisting of MSK, the Dana Farber Cancer Institute, Princess Margaret Cancer Centre in Toronto, Children's Hospital of Philadelphia, The Hyve in the Netherlands, and Bilkent University in Ankara, Turkey.

- **The web resource provides:**

  - Exploration, visualization, and analysis of multidimensional cancer genomics data

  - Molecular profiling data from cancer tissues and cell lines

  - Readily understandable genetic, epigenetic, gene expression, and proteomic events

  - Interactive results for genetic alterations across samples, genes, and pathways

  - When available in the underlying data, link to clinical outcomes

  - Graphical summaries of gene-level data from multiple platforms, network visualization and analysis, survival analysis, patient-centric queries, and software programmatic access
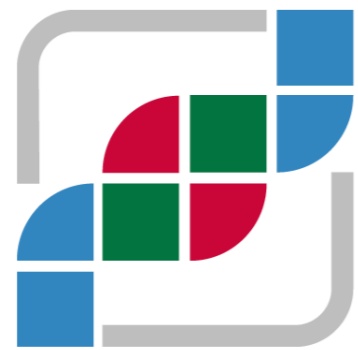
Chris Sanders, AACR 2010 Annual Meeting in Washington D.C.

"Wouldn't it be great if the data that will come from this ongoing TCGA study can be made available in an intuitive manner, to all those who want it? The array data and metadata would be displayed together for interactive analysis so as to maximize biological insights".

The cBio Cancer Genomics Portal is an:

- **Open-access resource for interactive exploration** of multidimensional cancer genomics data sets

- Currently providing access to **data from more than 5,000 tumor samples from 20 cancer studies**.

- Significantly lowers the barriers between complex genomic data and cancer researchers

- **Rapid, intuitive, and high-quality access to molecular profiles and clinical attributes**

- **Empowers researchers** to translate these rich data sets into biologic insights and clinical applications.

Cancer Discov; 2(5); 401–4.©2012 AACR.

cBioPortal for Cancer Genomics
http://cbioportal.org/

Data Sets   Web API   R/MATLAB   Tutorials   FAQ   News   Tools   About   Visualize Your Data

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

Please adhere to the TCGA publication guidelines when using TCGA data in your publications.

Please cite Gao et al. *Sci. Signal.* 2013 & Cerami et al. *Cancer Discov.* 2012 when publishing results based on cBioPortal.
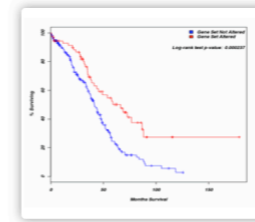
What's New

Sign up for low-volume email news alerts:

[   ]

SUBSCRIBE

Or follow us @cbioportal on Twitter

Query   Download Data

Select Cancer Study:

[Search...   ▼]

No studies selected.

▼ ☐ *All (151)*
  ▼ ☐ *PanCancer Studies (2)*
      ☐ **MSK-IMPACT Clinical Sequencing Cohort (MSKCC, Nat Med 2017)** *10946 samples*
      ☐ **Pan-Lung Cancer (TCGA, Nat Genet 2016)** *1144 samples*
  ▼ ☐ *Cell lines (2)*
      ☐ **Cancer Cell Line Encyclopedia (Novartis/Broad, Nature 2012)** *1019 samples*
      ☐ **NCI-60 Cell Lines (NCI, Cancer Res. 2012)** *60 samples*
  ▼ ☐ *Adrenal Gland (1)*
    ▼ ☐ *Adrenocortical Carcinoma (1)*

Select Data Type Priority: ⦿ Mutation and CNA   ◯ Only Mutation   ◯ Only CNA

Enter Gene Set: Advanced: Onco Query Language (OQL)

[User-defined List   ]

[**Enter Gene Symbols or Gene Aliases**

                                                                  ]

Submit

Data Sets

The Portal contains **151 cancer studies**. [Details]

Example Queries

• RAS/RAF alterations in colorectal cancer
• BRCA1 and BRCA2 mutations in ovarian cancer
• POLE hotspot mutations in endometrial cancer
• TP53 and MDM2/4 alterations in GBM
• PTEN mutations in GBM in text format
• BRAF V600E mutations across cancer types
• Patient view of an endometrial cancer case

What People are Saying

"Whenever bench scientists ask me how they can look at TCGA data, I've never had a good answer for them. Now I do. The cBio Portal meets a critical need--it is the interface that the cancer research community needs to access the wealth of TCGA. Even as a computational biologist, I use it to follow-up on genes of interest. It makes querying the data much less painful."

*– Postdoctoral Fellow, Oregon Health & Science University*

View All

Complete List of Datasets and Datatypes

cBioPortal FOR CANCER GENOMICS

Data Sets

The portal currently contains data from 151 cancer genomics studies. The table below lists the number of available samples per cancer study and data type.

Search:

CancerStudy	Reference	All	Sequenced	CNA	Tumor mRNA (RNA-Seq)	Tumor mRNA (microarray)	Tumor miRNA	Methylation (HM27)	RPPA	Complete
Acinar Cell Carcinoma of the Pancreas (Johns Hopkins, J Pathol 2014)	Jial et al. J Pathol 2014	23	23							
Acute Myeloid Leukemia (TCGA, NEJM 2013)	TCGA, NEJM 2013	200	200	191	173			194		166
Acute Myeloid Leukemia (TCGA, Provisional)		200	197	191	173			194		163
Adenoid Cystic Carcinoma (FMI, Am J Surg Pathl. 2014)	Ross et al. Am J Surg Pathl. 2014	28	15	28						
Adenoid Cystic Carcinoma (MDA, Clin Cancer Res 2015)	Mitani et al. Clin Cancer Res. 2015	102	102							
Adenoid Cystic Carcinoma (MSKCC, Nat Genet 2013)	Ho et al. Nat Genet 2013	60	60	60						
Adenoid Cystic Carcinoma (Sanger/MDA, JCI 2013)	Stephens et al. JCI 2013	24	24							
Adenoid Cystic Carcinoma of the Breast (MSKCC, J Pathol. 2015)	Martelotto et al. J Pathol. 2015	12	12	12						

cBioPortal for Cancer Genomics: Data to knowledge

Tumor DNA

DNA sequencer, microarrays …

Tumor and normal sequences
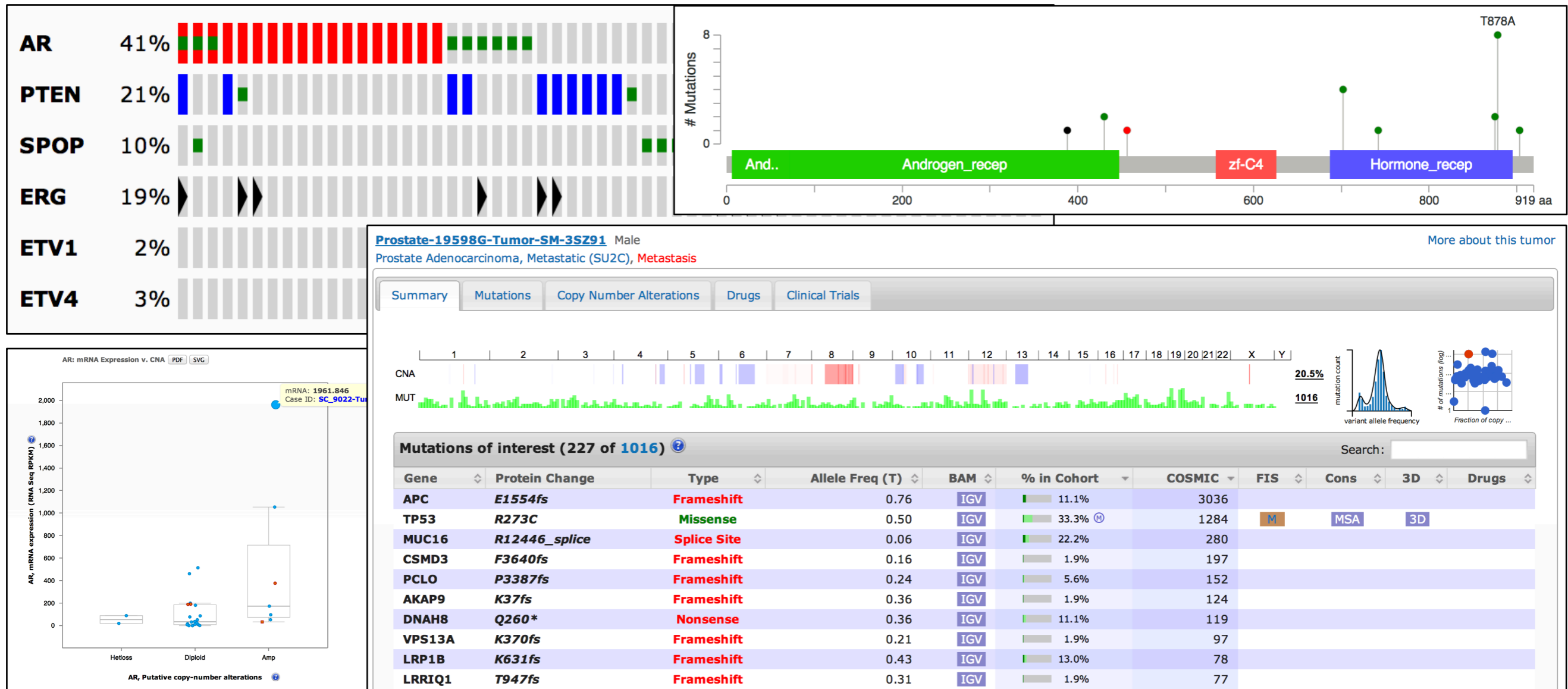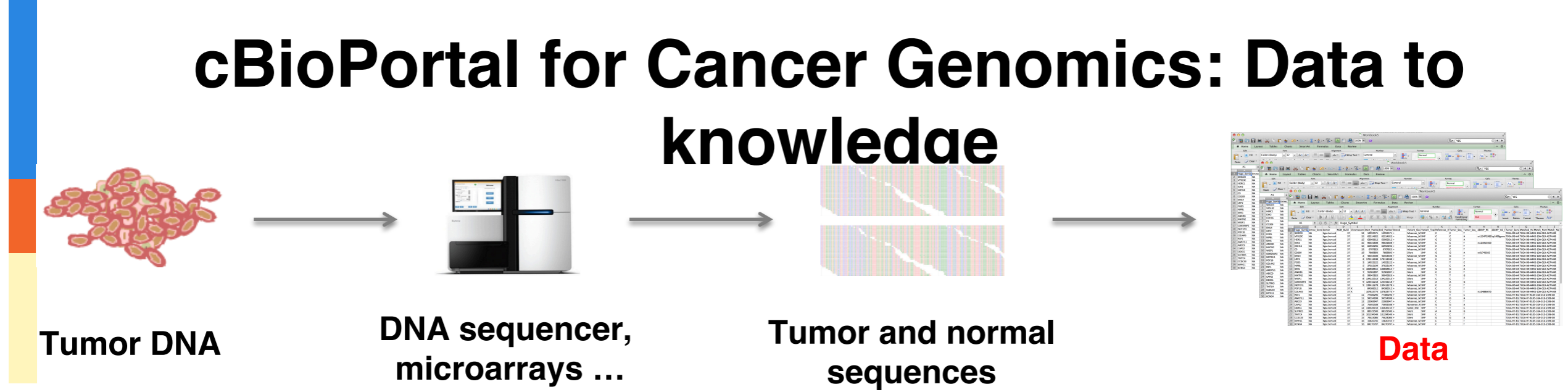
Data

Intuitive interface, quick response time, reduction of complexity
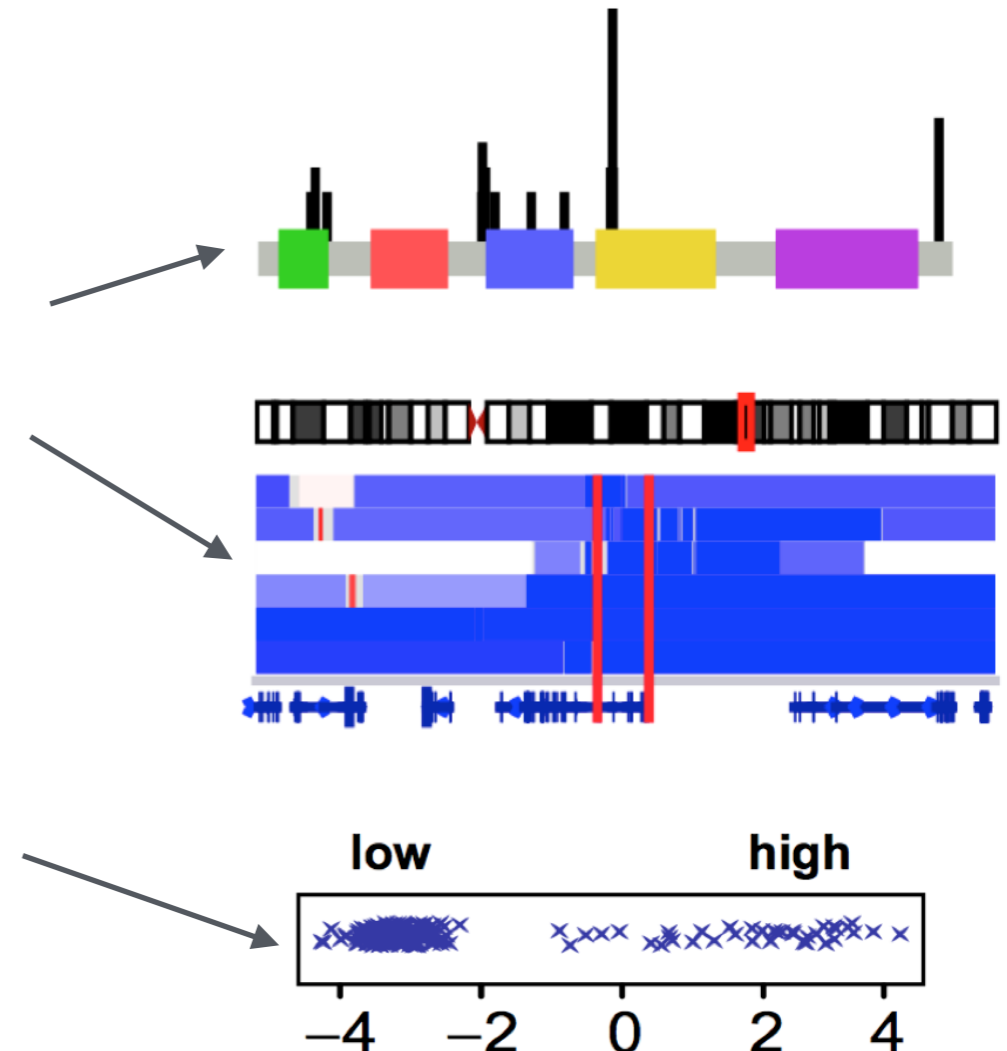
Reduction of complexity: Event calls
Which genes are altered in which samples?

Event calls

Data type	Alteration event calls
Mutations	Non-synonymous somatic mutations
Copy number changes	Homozygous deletion or amplification
Methylation	Epigenetic silencing
mRNA and/or DNA	Gene fusions
mRNA expression changes	Over- or under-expression

low   high

−4 −2 0 2 4

Alteration types and thresholds can be customized for each gene.

Live Demo
http://cbioportal.org

cBioPortal resources / instructions

Science Signaling

Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal

Jianjiong Gao[1], Bülent Arman Aksoy[1], Ugur Dogrusoz[2], Gideon Dresdner[1], Benjamin Gross[1], S. Onur Sumer[1], Yichao Sun[1], Anders Jacobsen[1], Rileen Sinha[1], Erik Larsson[3], Ethan Cerami[1,4], Chris Sander[1], and Nikolaus Schultz[1]

Slide courtesy: Nikolaus Schultz

cBioPortal resources / instructions

Science Signaling

PROTOCOLS

Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal

Jianjiong Gao[1], Bülent Arman Aksoy[1], Ugur Dogrusoz[2], Gideon Dresdner[1], Benjamin Gross[1], S. Onur Sumer[1], Yichao Sun[1], Anders Jacobsen[1], Rileen Sinha[1], Erik Larsson[3], Ethan Cerami[1,4], Chris Sander[1], and Nikolaus Schultz[1]

Google

Search for topics   🔍

Groups

NEW TOPIC   C   Mark all as read   Actions ▾   Filters ▾

My groups

Home

◄ **cBioPortal for Cancer Genomics Discussion Group**   Shared publicly
60 of 568 topics (99+ unread) ☆   g+1

Slide courtesy: Nikolaus Schultz

cBioPortal resources / instructions

Science Signaling

PROTOCOLS

Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal

Jianjiong Gao[1], Bülent Arman Aksoy[1], Ugur Dogrusoz[2], Gideon Dresdner[1], Benjamin Gross[1], S. Onur Sumer[1], Yichao Sun[1], Anders Jacobsen[1], Rileen Sinha[1], Erik Larsson[3], Ethan Cerami[1,4], Chris Sander[1], and Nikolaus Schultz[1]

Google   Search for topics   🔍

Groups   NEW TOPIC   ↻   Mark all as read   Actions ▼   Filters ▼

My groups   ◄ **cBioPortal for Cancer Genomics Discussion Group**   Shared publicly
Home   60 of 568 topics (99+ unread) ☆   g+1

cBioPortal for Cancer Genomics   *Visualize, analyze, discover.*

HOME   DATA SETS   WEB API   R/MATLAB   TUTORIALS   FAQ   NEWS   ABOUT

Portal FAQs

Slide courtesy: Nikolaus Schultz

Link to more exercises:

http://cbio.mskcc.org/cancergenomics/portalworkshop/

Video clip on sources and potential usage of Data Mining by Dr. Atul Butte, UCSF

https://vimeo.com/156680823

National Cancer Institute

The Cancer Genome Atlas — Understanding genomics to improve cancer care

ENCODE

HMP

NIH HUMAN MICROBIOME PROJECT

CCLE Cancer Cell Line Encyclopedia

fitbit

HOME   BROWSE   ANALYSIS TOOLS

Broad-N

Food and Drug Administration
MedWatch

BROAD INSTITUTE

What is Broad   News and Publications

Home › For the Scientific Community Science › Projects › Connectivity Map › Connectivity Map

Connectivity Map

HEP
Human Epigenome Project

NIH LINCS PROGRAM

National Cancer Institute
Surveillance Epidemiology and End Results

Home   About SEER   Cancer Statistics   Datasets & Software   Publications

Cerner

HOME   ABOUT   CENTERS   DATA   ASSAYS

Epic

UK 10K

LINC
of biology by
and other cell
exposed to a v

dbGaP
GENOTYPE and PHENOTYPE

PG KB   PharmGKB

Go   Advanced search