

Introduction to CBioPortal

Augustin Luna
20 January, 2016

Research Fellow
Department of Biostatistics and Computational Biology
Dana-Farber Cancer Institute

Topics to be Covered

- CBioPortal overview
- Data retrieval from CBioPortal using `cgdsr`
 - List cancer datasets
 - List data types for dataset
 - List case sets
 - Retrieve data for selected genes
 - Retrieve clinical data
- Plotting retrieved data

What is CBioPortal?

- Publication:
 - <http://www.ncbi.nlm.nih.gov/pubmed/23550210>
- Provides molecular profiles and clinical attributes from large-scale cancer genomics projects
- Dataset List
 - http://www.cbioportal.org/data_sets.jsp
- Datasets
 - Total 105 datasets (as of Jan. 2015)
 - TCGA: The Cancer Genome Atlas
 - Cancer Cell Line Encyclopedia
 - NCI-60 (no drug response data)
 - Other collected large-scale datasets from publications

Accessing CBioPortal

- Interactive Website: <http://www.cbiportal.org/>
- Project provides an application programming interface (API)
 - Makes CBioPortal accessible from any programming language
 - http://www.cbiportal.org/web_api.jsp
- `cgdsr` R Package simplifies access to the CBioPortal API

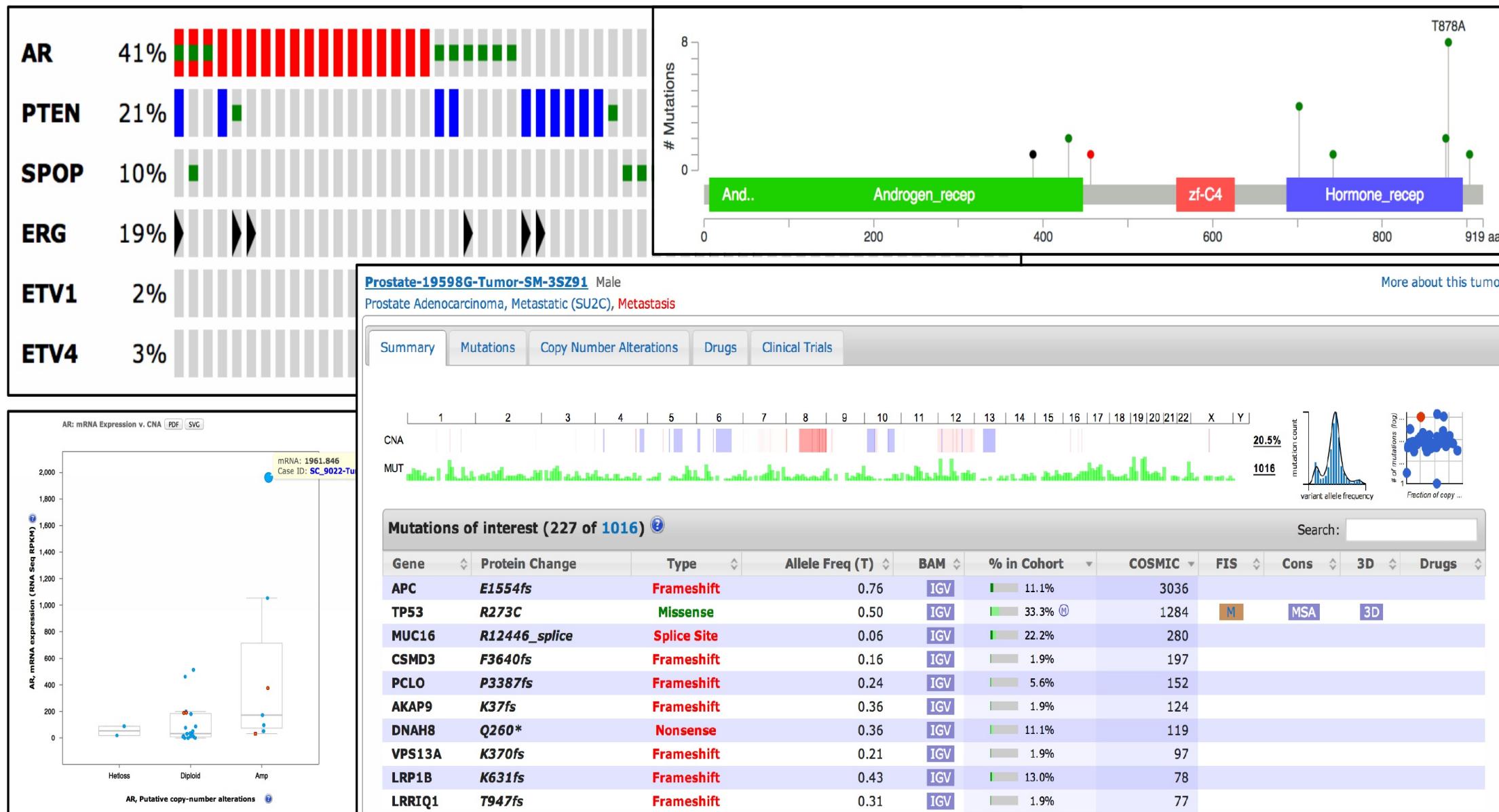
CBioPortal Data Types

- NOTE: Not all samples will have complete data
 - DNA copy number variation
 - mRNA expression
 - microRNA expression
 - Mutation profiles
 - Protein and phospho-protein expression
 - DNA methylation
 - Limited de-identified clinical data

CBioPortal Data Notes

- Notes on the available data
 - <http://www.cbiportal.org/faq.jsp>
- Example notes
 - Clinical data, includes: overall and disease-free survival, gender, age, stage and tumor grade, when available.
 - mRNA and microRNA data is z-scored (i.e. the number of standard deviations away from the mean of expression)

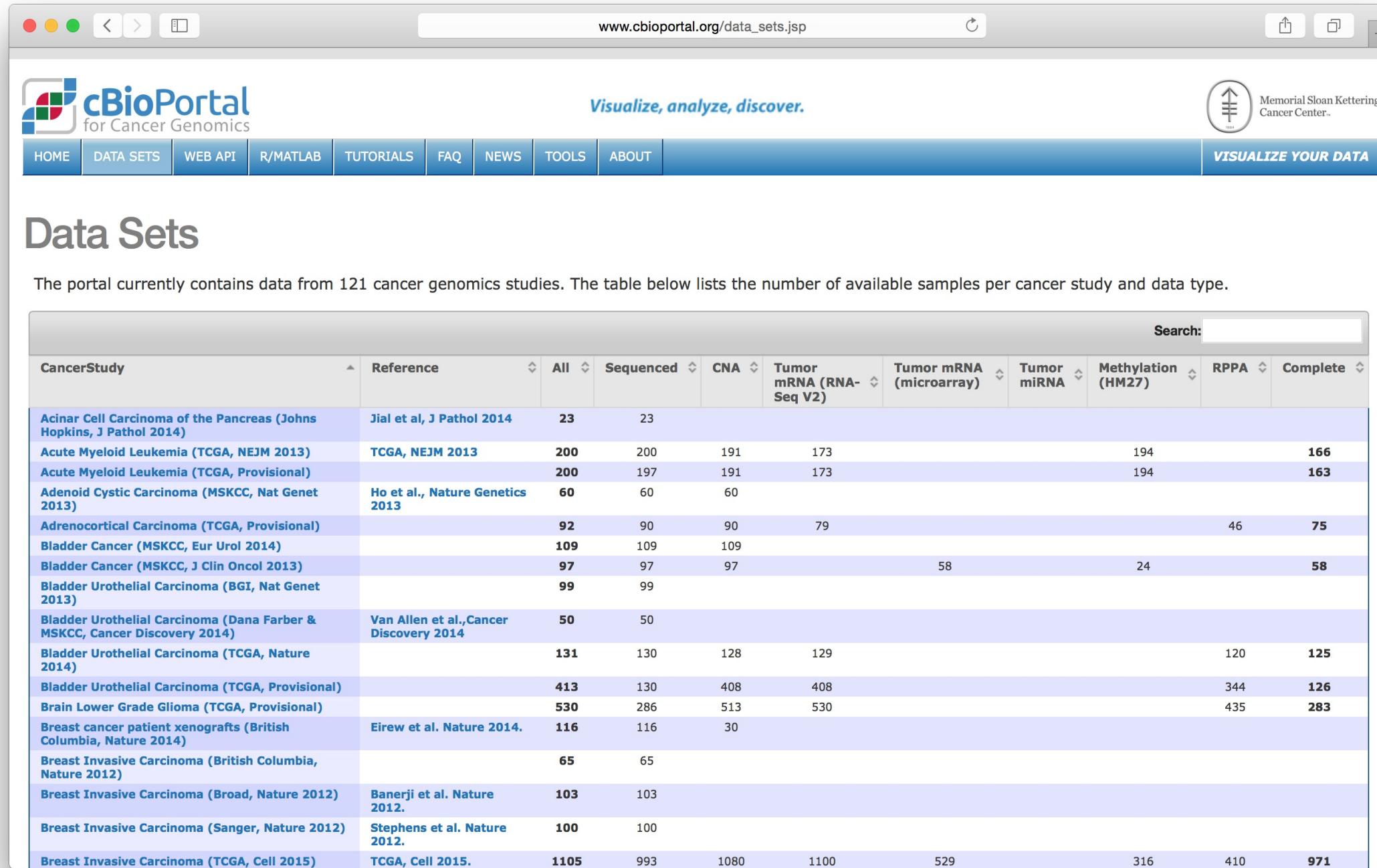
CBioPortal Website Features



CBioPortal Tutorials

The screenshot shows a web browser window displaying the cBioPortal tutorial page at www.cbioportal.org/tutorial.jsp. The page has a header with the cBioPortal logo, a navigation bar with links for HOME, DATA SETS, WEB API, R/MATLAB, TUTORIALS, FAQ, NEWS, TOOLS, and ABOUT. A blue banner across the top right says "Visualize, analyze, discover." and "VISUALIZE YOUR DATA". The main content area features a section titled "Tutorials" with a link to a "Step-by-step Guide to cBioPortal: a Protocol Paper". Below this, there's a summary of a paper by Gao, Aksoy, Dogrusoz, Dresdner, Gross, Sumer, Sun, Jacobsen, Sinha, Larsson, Cerami, Sander, Schultz, titled "Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal". The paper was published in *Sci. Signal.* 6, pl1 (2013). A "Reprint" link is also provided. Another section, "Tutorial #1: Getting Started", has a link to "cBioPortal for Cancer Genomics: Getting started". On the right side, there are three panels: "What's New" (with a note about new jobs at Dana-Farber), "Data Sets" (showing a heatmap of 121 cancer studies), and "Example Queries" (listing RAS/RAF alterations in colorectal cancer, BRCA1 and BRCA2 mutations in ovarian cancer, and POLM hotspot mutations in endometrial cancer).

CBioPortal Datasets



The portal currently contains data from 121 cancer genomics studies. The table below lists the number of available samples per cancer study and data type.

CancerStudy	Reference	All	Sequenced	CNA	Tumor mRNA (RNA-Seq V2)	Tumor 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 (MSKCC, Nat Genet 2013)	Ho et al., Nature Genetics 2013	60	60	60						
Adrenocortical Carcinoma (TCGA, Provisional)		92	90	90	79			46		75
Bladder Cancer (MSKCC, Eur Urol 2014)		109	109	109						
Bladder Cancer (MSKCC, J Clin Oncol 2013)		97	97	97		58		24		58
Bladder Urothelial Carcinoma (BGI, Nat Genet 2013)		99	99							
Bladder Urothelial Carcinoma (Dana Farber & MSKCC, Cancer Discovery 2014)	Van Allen et al., Cancer Discovery 2014	50	50							
Bladder Urothelial Carcinoma (TCGA, Nature 2014)		131	130	128	129			120		125
Bladder Urothelial Carcinoma (TCGA, Provisional)		413	130	408	408			344		126
Brain Lower Grade Glioma (TCGA, Provisional)		530	286	513	530			435		283
Breast cancer patient xenografts (British Columbia, Nature 2014)	Eirew et al. Nature 2014.	116	116	30						
Breast Invasive Carcinoma (British Columbia, Nature 2012)		65	65							
Breast Invasive Carcinoma (Broad, Nature 2012)	Banerji et al. Nature 2012.	103	103							
Breast Invasive Carcinoma (Sanger, Nature 2012)	Stephens et al. Nature 2012.	100	100							
Breast Invasive Carcinoma (TCGA, Cell 2015)	TCGA, Cell 2015.	1105	993	1080	1100	529		316	410	971

CBioPortal Home Data Selection

www.cbioportal.org/index.do

Select Cancer Study:

breast

All Search Results (7)

 Breast (7)

 Invasive Breast Carcinoma (7)

 Breast Invasive Carcinoma (British Columbia, Nature 2012) 65 samples

 Breast cancer patient xenografts (British Columbia, Nature 2014) 116 samples

 Breast Invasive Carcinoma (Broad, Nature 2012) 103 samples

 Breast Invasive Carcinoma (Sanger, Nature 2012) 100 samples

 Breast Invasive Carcinoma (TCGA, Provisional) 1105 samples

 Breast Invasive Carcinoma (TCGA, Nature 2012) 825 samples

 Breast Invasive Carcinoma (TCGA, Cell 2015) 1105 samples

Select Genomic Profiles:

Mutations

Putative copy-number alterations from GISTIC

mRNA Expression data. Select one of the profiles below:

mRNA Expression z-Scores (microarray)

mRNA Expression z-Scores (RNA Seq V2 RSEM)

Enter a z-score threshold ±: 2.0

Select Patient/Case Set:

All Complete Tumors (971)

To build your own case set, try out our enhanced Study View.

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

User-defined List

Select From Recurrently Mutated Genes (MutSig) Select Genes from Recurrent CNAs (Gistic)

TP53 BRAF

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

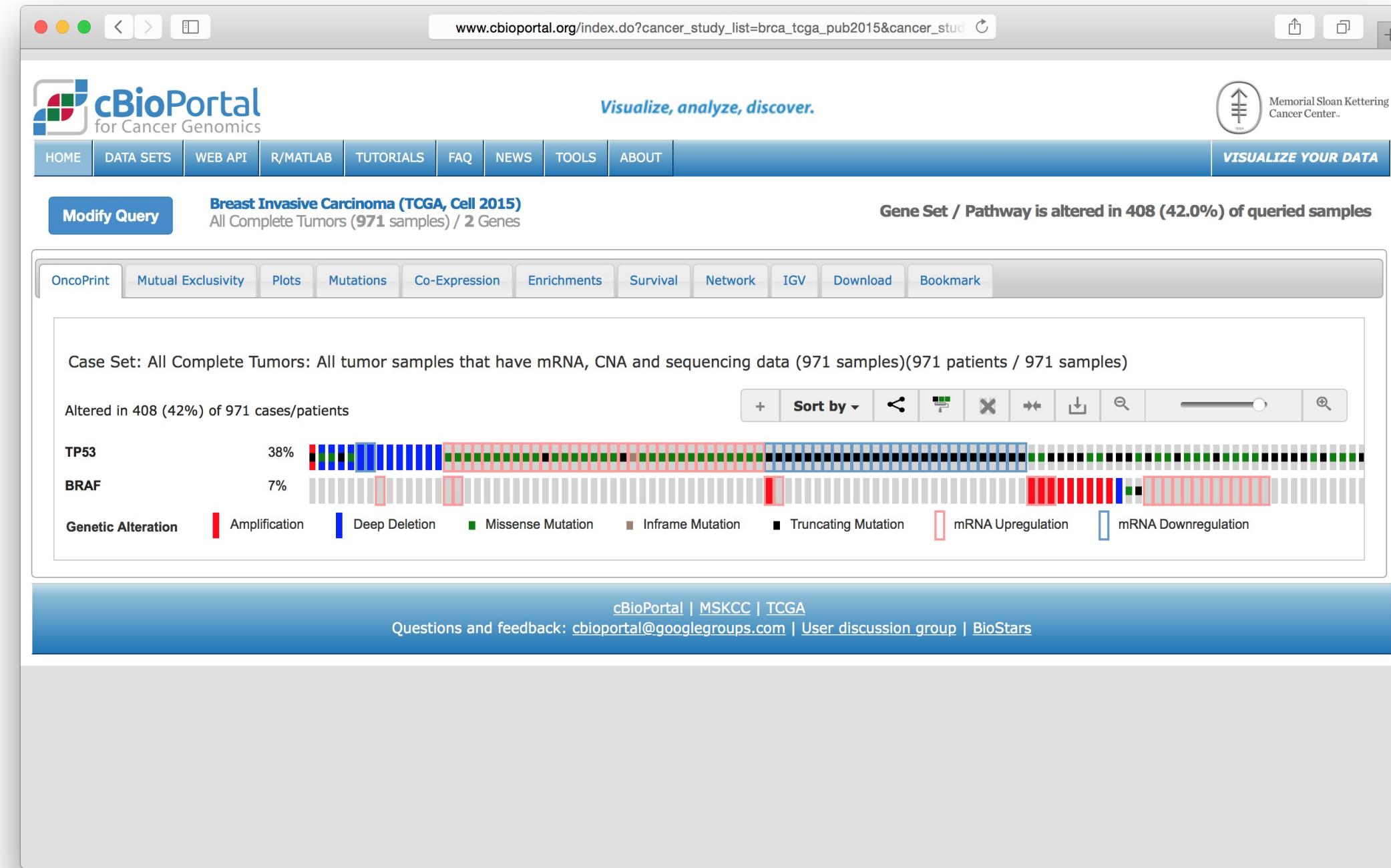
What People are Saying

"This portal is truly the greatest thing since sliced bread. I am making discoveries with it not only in glioblastoma, my primary focus, but in other cancers as well -- it's all so easy with this fantastic tool. And I am enjoying showing it to my colleagues, whose jaws also drop. Thank you a thousand times over for this beautiful public resource. I am looking forward to citing this soon in an upcoming paper..."

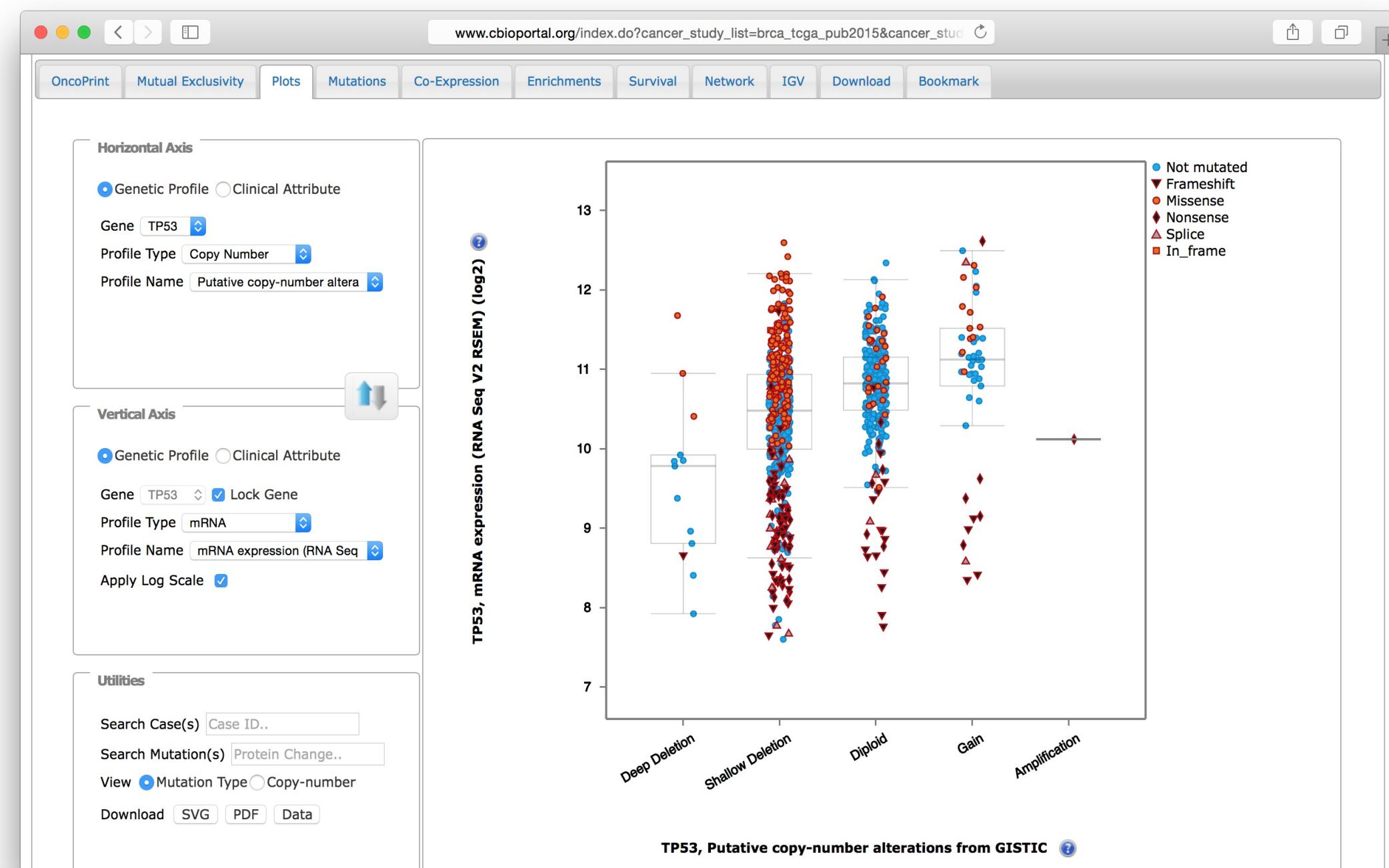
– Associate Professor, University of Virginia

[View All](#)

CBioPortal OncoPrint



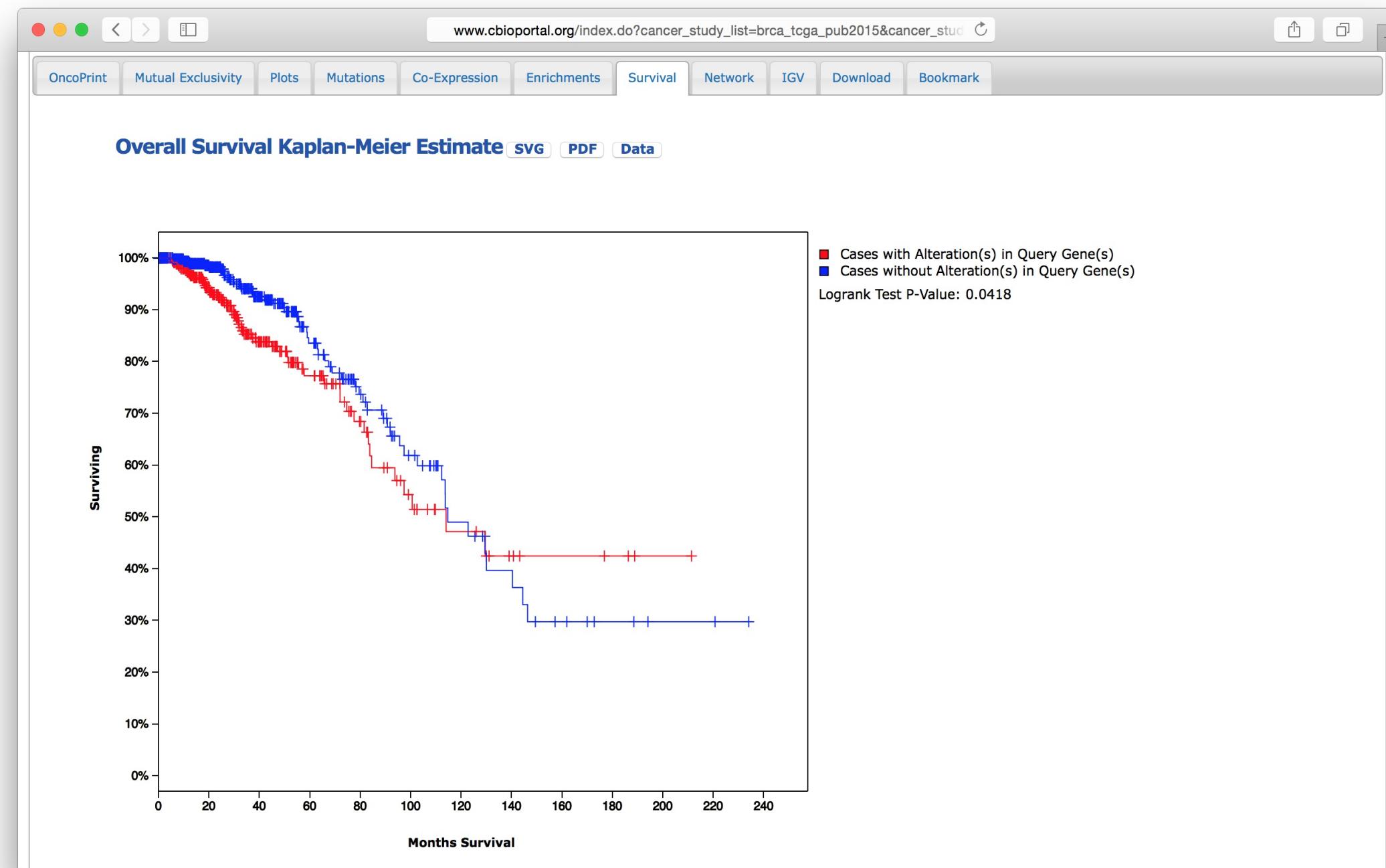
CBioPortal Plots



CBioPortal Mutations



CBioPortal Survival



CBioPortal OncoPrinter

The screenshot shows the OncoPrinter page of the CBioPortal website. At the top, the CBioPortal logo and navigation menu are visible. The main heading "OncoPrinter" is displayed, followed by a sub-instruction: "generates Oncoprints from your own data (v1.0.1)". Below this, a note says "Please input **tab-delimited** genomic alteration events." with a link to "Data format". A "Input Mutation Data" section includes a "Load example data" link and a large text input area. Below it, there's a "Choose File" button with "no file selected". A section for defining gene order is present with a text input field and a "Submit" button. At the bottom, a footer bar contains links to cBioPortal, MSKCC, TCGA, and other resources.

www.cbiportal.org/oncoprinter.jsp

cBioPortal for Cancer Genomics

Visualize, analyze, discover.

Memorial Sloan Kettering Cancer Center

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

OncoPrinter generates Oncoprints from your own data (v1.0.1)

Please input **tab-delimited** genomic alteration events. [Data format](#)

Input Mutation Data [Load example data](#)

Input Mutation Data File
 Choose File no file selected

Please define the order of genes (optional).

Submit

cBioPortal | MSKCC | TCGA
Questions and feedback: cbiportal@googlegroups.com | [User discussion group](#) | [BioStars](#)

CBioPortal Mutation Mapper

The screenshot shows the 'MutationMapper' page of the CBioPortal website. At the top, the CBioPortal logo and navigation links (HOME, DATA SETS, WEB API, R/MATLAB, TUTORIALS, FAQ, NEWS, TOOLS, ABOUT) are visible. To the right, there's a 'Visualize, analyze, discover.' tagline and the Memorial Sloan Kettering Cancer Center logo. A large blue button labeled 'VISUALIZE YOUR DATA' is prominent. Below this, the 'MutationMapper' section title is displayed, along with a note that it 'interprets mutations with protein annotations (v1.0.1)'. It instructs users to input tab-delimited mutation data, providing links for 'Data Format' and 'Load example data'. A large empty rectangular area is provided for file upload. Below this area, there's a 'Upload your own Mutation File' section with a 'Choose File' button (showing 'no file selected') and a 'Visualize' button. At the bottom, a footer bar contains links to cBioPortal, MSKCC, and TCGA, and a 'Questions and feedback' link.

www.cbioportal.org/mutation_mapper.jsp

cBioPortal for Cancer Genomics

Visualize, analyze, discover.

Memorial Sloan Kettering Cancer Center

VISUALIZE YOUR DATA

MutationMapper interprets mutations with protein annotations (v1.0.1)

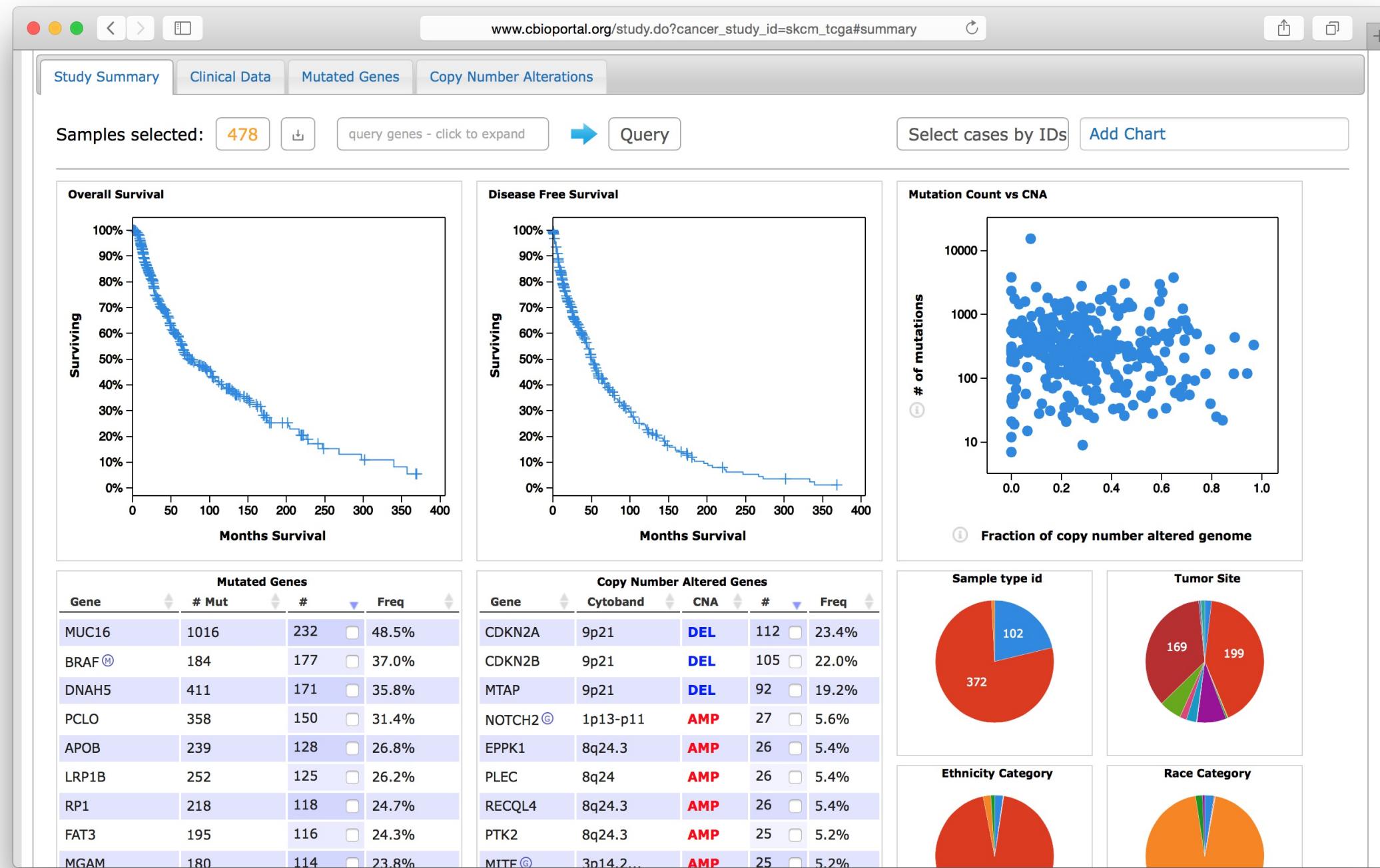
Please input **tab-delimited** mutation data. [Data Format](#) [Load example data](#)

Upload your own Mutation File

no file selected

cBioPortal | MSKCC | TCGA
Questions and feedback: cbioportal@googlegroups.com | [User discussion group](#) | [BioStars](#)

CBioPortal Study: Summary



CBioPortal Study: Clinical Data

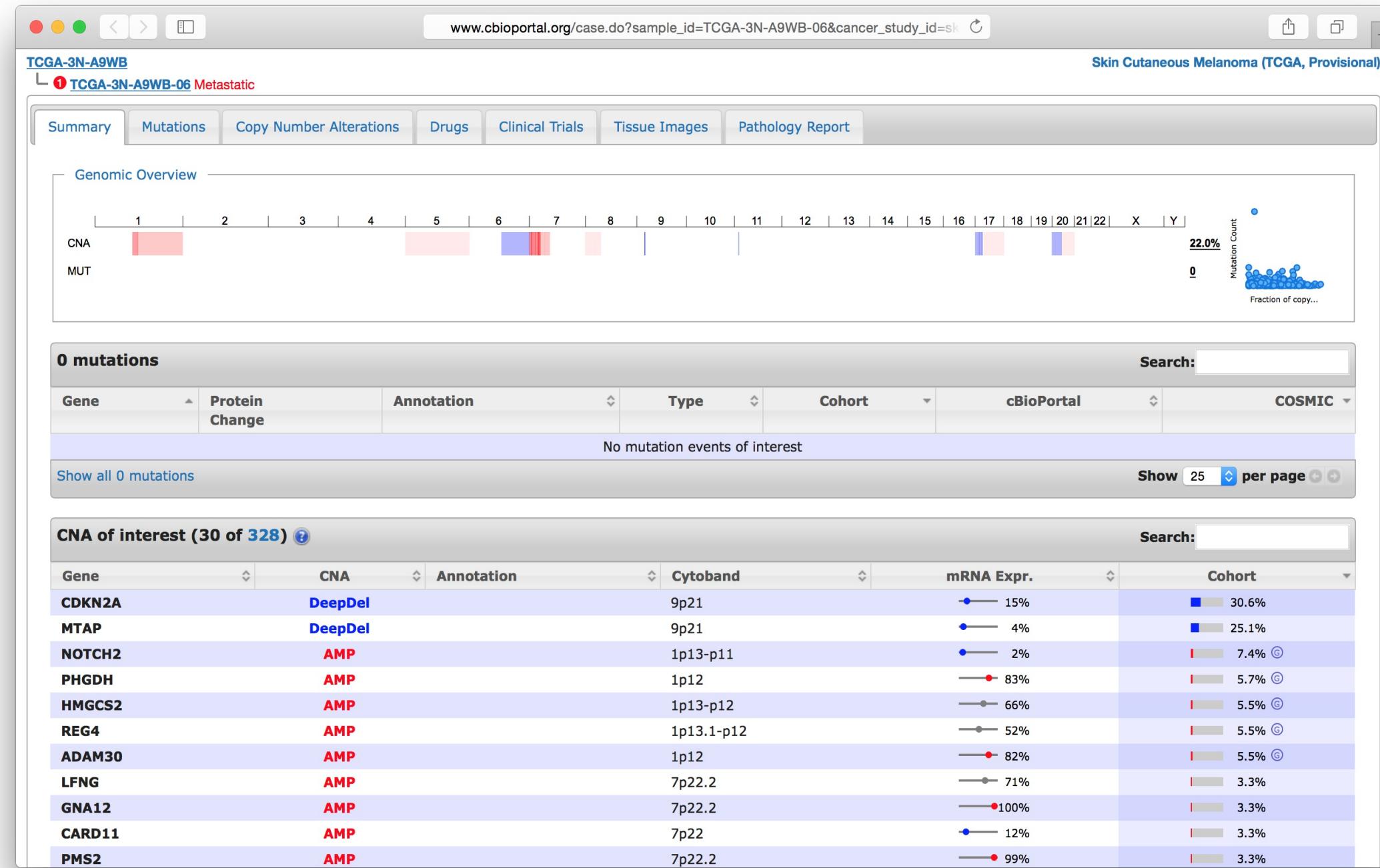
www.cbioportal.org/study.do?cancer_study_id=skcm_tcga#clinical

Study Summary Clinical Data Mutated Genes Copy Number Alterations

Copy CSV Show / hide columns Showing 478 samples Search: Clark level at diagnosis

PATIENT_ID	Sample ID	Adjuvant Postop...	American Joint ...	American Joint ...	American Joint ...	Birth from Init...	Clark level at ...	CNA	Days to Last Fo...	Days t Sampl
TCGA-3N-A9WB	TCGA-3N-A9WB-06							0.220		552
TCGA-3N-A9WC	TCGA-3N-A9WC-06	NO	6th	M0	T2b	-30286	IV	0.350	1856	1709
TCGA-3N-A9WD	TCGA-3N-A9WD-06	NO	7th	M0	T2a	-30163	III	0.252		248
TCGA-BF-A1PU	TCGA-BF-A1PU-01	NO	7th	M0	T4b	-17025	III	0.683	387	338
TCGA-BF-A1PV	TCGA-BF-A1PV-01	[Unknown]	7th	M0	T4b	-27124	IV	0.187	14	106
TCGA-BF-A1PX	TCGA-BF-A1PX-01	NO	7th	M0	T4b	-20626	III	7.620e-4		122
TCGA-BF-A1PZ	TCGA-BF-A1PZ-01	NO	7th	M0	T4a	-26240	III	0.225	12	97
TCGA-BF-A1Q0	TCGA-BF-A1Q0-01	NO	7th	M0	T4b	-29380	III	0.645	17	75
TCGA-BF-A3DJ	TCGA-BF-A3DJ-01	NO	7th	M0	T4b	-13332	III	0.332	464	183
TCGA-BF-A3DL	TCGA-BF-A3DL-01	NO	7th	M0	T3b	-30805	IV	0.377	28	229
TCGA-BF-A3DM	TCGA-BF-A3DM-01	NO	7th	M0	T2b	-23076	V	0.156	14	61
TCGA-BF-A3DN	TCGA-BF-A3DN-01	NO	7th	M0	T3b	-29801	IV	0.415	32	177
TCGA-BF-A5EO	TCGA-BF-A5EO-01	NO	7th	M0	T4b	-23940	III	0.214	338	134
TCGA-BF-A5EP	TCGA-BF-A5EP-01	NO	7th	M0	T4b	-27622	V	0.830	11	131
TCGA-BF-A5EQ	TCGA-BF-A5EQ-01	NO	7th	M0	T4b	-23353	III	0.215	12	119
TCGA-BF-A5ER	TCGA-BF-A5ER-01	NO	7th	M0	T4b	-23199	IV	0.299	12	91
TCGA-BF-A5ES	TCGA-BF-A5ES-01	NO	7th	M0	T4b	-28092	IV	0.123	19	89
TCGA-BF-A9VF	TCGA-BF-A9VF-01	NO	7th	M0	T4b	-28312	III	0.505	12	154
TCGA-BF-AAOU	TCGA-BF-AAOU-01	NO	7th	M0	T4b	-26750	IV	0.326	13	199
TCGA-BF-AAOX	TCGA-BF-AAOX-01	NO	7th	M0	T4b	-30318	IV	0.180	11	167
TCGA-BF-AAP0	TCGA-BF-AAP0-06	NO	7th	M1	T4	-14773		0.0251	14	177
TCGA-BF-AAP1	TCGA-BF-AAP1-01	NO	7th	M0	T4b	-31666	IV	0.454	14	132
TCGA-BF-AAP2	TCGA-BF-AAP2-01	NO	7th	M0	T3b	-22734	III	0.135	13	128
TCGA-BF-AAP4	TCGA-BF-AAP4-01	NO	7th	M0	T4b	-22303	V	0.410	15	58

CBioPortal Patient: Summary

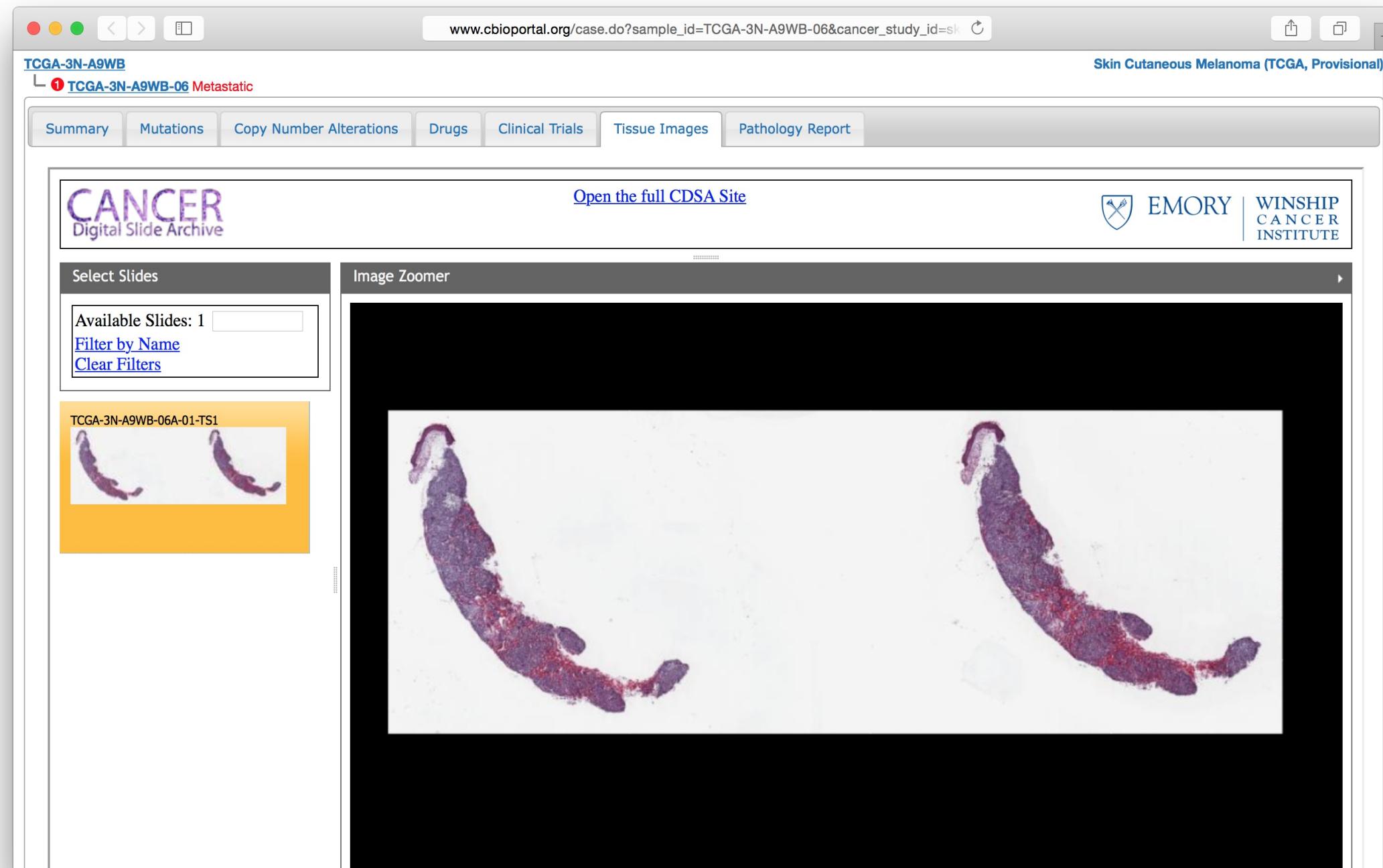


CBioPortal Patient: Drug Targets

The screenshot shows a web browser window for the CBioPortal website (www.cbioperl.org/case.do?sample_id=TCGA-3N-A9WB-06&cancer_study_id=sk). The page is titled '33 drugs of interest'. The 'Drugs' tab is active in the navigation bar. A search bar is present at the top right. The main content is a table with the following columns: Drug Name, Drug Target(s), Description, FDA approved?, and Data Sources.

Drug Name	Drug Target(s)	Description	FDA approved?	Data Sources
Abexinostat (10 more)	HDAC9		No	
Alefacet (11 more)	CD58 CD2	Immunosuppressive dimeric fusion protein that consists of the extracellular CD2-binding portion of the human leukocyte function antigen-3 (LFA-3) linked to the Fc (hinge, CH2 and CH3 domains) portion of human IgG1. Produced by CHO cells, mW is 91.4 kD.	Yes	PharmGKB, DrugBank
Aminophylline (20 more)	ADORA3	Aminophylline is a drug combination that contains theophylline and ethylenediamine in a 2:1 ratio. Once in the body, theophylline is released and acts as a phosphodiesterase inhibitor, adenosine receptor blocker, and histone deacetylase activator. Similar to other theophyllines, aminophylline is indicated for the treatment of lung diseases such as asthma, chronic bronchitis, and COPD. The majority of aminophylline medications are discontinued and the remaining medications on the market are in short supply.	Yes	PharmGKB, DrugBank, KEGG Drug
Amiodarone (12 more)	KCND3	An antianginal and antiarrhythmic drug. It increases the duration of ventricular and atrial muscle action by inhibiting Na,K-activated myocardial adenosine triphosphatase. There is a resulting decrease in heart rate and in vascular resistance. [PubChem]	Yes	PharmGKB, DrugBank, KEGG Drug
Belinostat (10 more)	HDAC9		No	
CGP-082996 (1 more)	CDKN2A		No	
Diazoxide (7 more)	ATP1A1	A benzothiadiazine derivative that is a peripheral vasodilator used for hypertensive emergencies. It lacks diuretic effect, apparently because it lacks a sulfonamide group. [PubChem]	Yes	PharmGKB, DrugBank, KEGG Drug
Digoxin (8 more)	ATP1A1	A cardiotonic glycoside obtained mainly from Digitalis lanata; it consists of three sugars and the aglycone digoxigenin. Digoxin has positive inotropic and negative chronotropic activity. It is used to control ventricular rate in atrial fibrillation and in the management of congestive heart failure with atrial fibrillation. Its use in congestive heart failure and sinus rhythm is less certain. The margin between toxic and therapeutic doses is small. (From Martindale, The Extra Pharmacopoeia, 30th ed, p666)	Yes	PharmGKB, DrugBank, KEGG Drug
Ficlatuzumab (1 more)	IL6		No	
Flavopiridol (10 more)	CDKN2A	Flavopiridol is a synthetic flavonoid based on an extract from an Indian plant for the potential treatment of cancer. It works by inhibiting cyclin-dependent kinases, arresting cell division and causing apoptosis in non-small lung cancer cells.	No	PharmGKB, DrugBank
Flutamide (2 more)	AHR	An antiandrogen with about the same potency as cyproterone in rodent and canine species.	Yes	PharmGKB, DrugBank, KEGG Drug

CBioPortal Patient: Histology



CBioPortal Patient: Pathology

www.cbiportal.org/case.do?sample_id=TCGA-3N-A9WB-06&cancer_study_id=sk

TCGA-3N-A9WB
① TCGA-3N-A9WB-06 Metastatic

Skin Cutaneous Melanoma (TCGA, Provisional)

Summary Mutations Copy Number Alterations Drugs Clinical Trials Tissue Images Pathology Report

PATIENT EXAMINATION

PATIENT NAME: [REDACTED] COLLECTED: [REDACTED]
MED. REC. #: [REDACTED] CLIENT: [REDACTED]
SOC. SEC. #: [REDACTED] PHYSICIAN: [REDACTED]
GENDER: [REDACTED] ADDITIONAL: [REDACTED]
ACCOUNT #: [REDACTED] COPY TO/NOTES: [REDACTED]

SPECIMEN(S) RECEIVED: 1C00-3
A: Left flank melanoma excision
B: Left axillary nodule

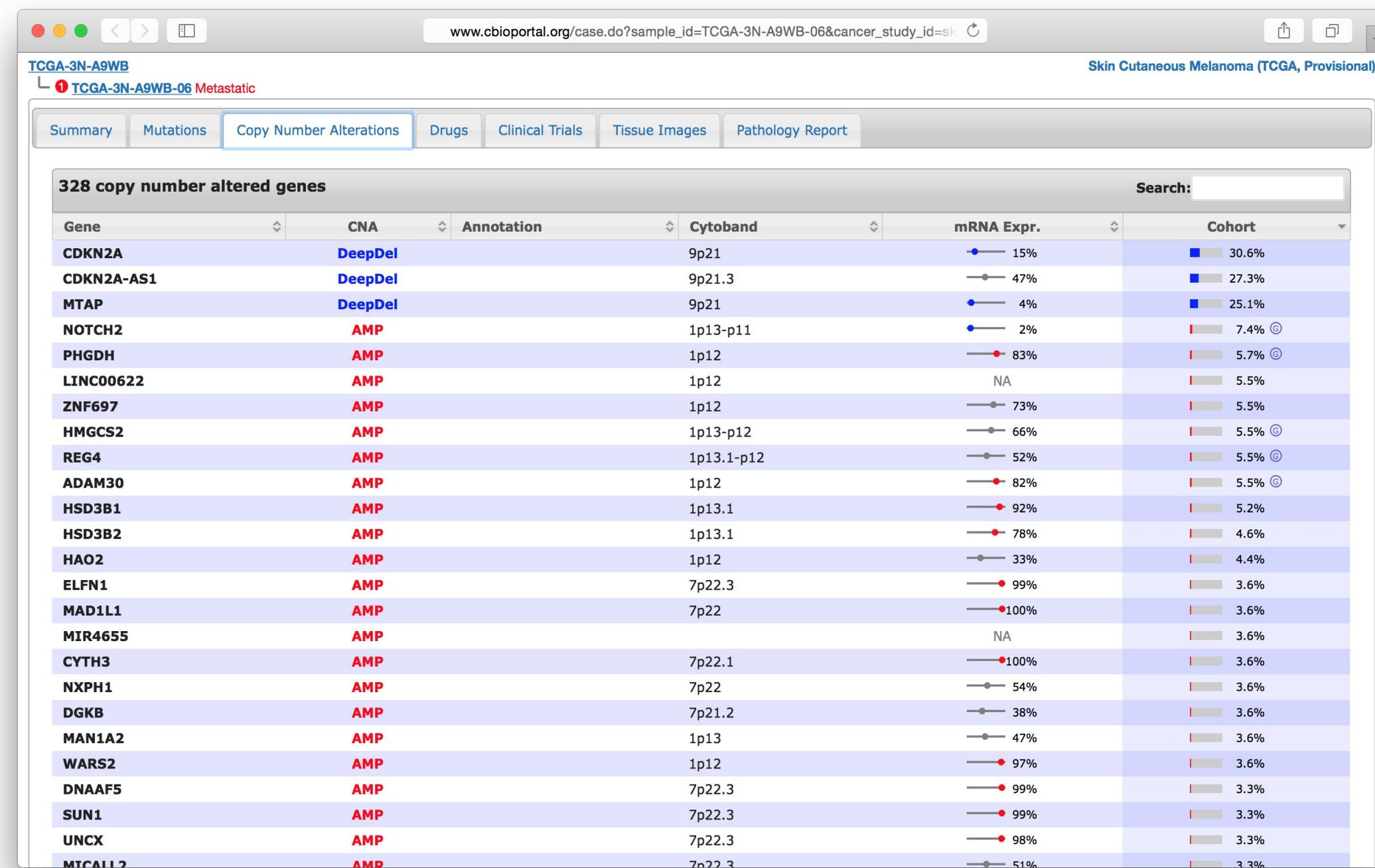
CLINICAL DATA/HISTORY: [REDACTED]

Melanoma NOS 8720/3
Site: C44.5
JG 11/7/14

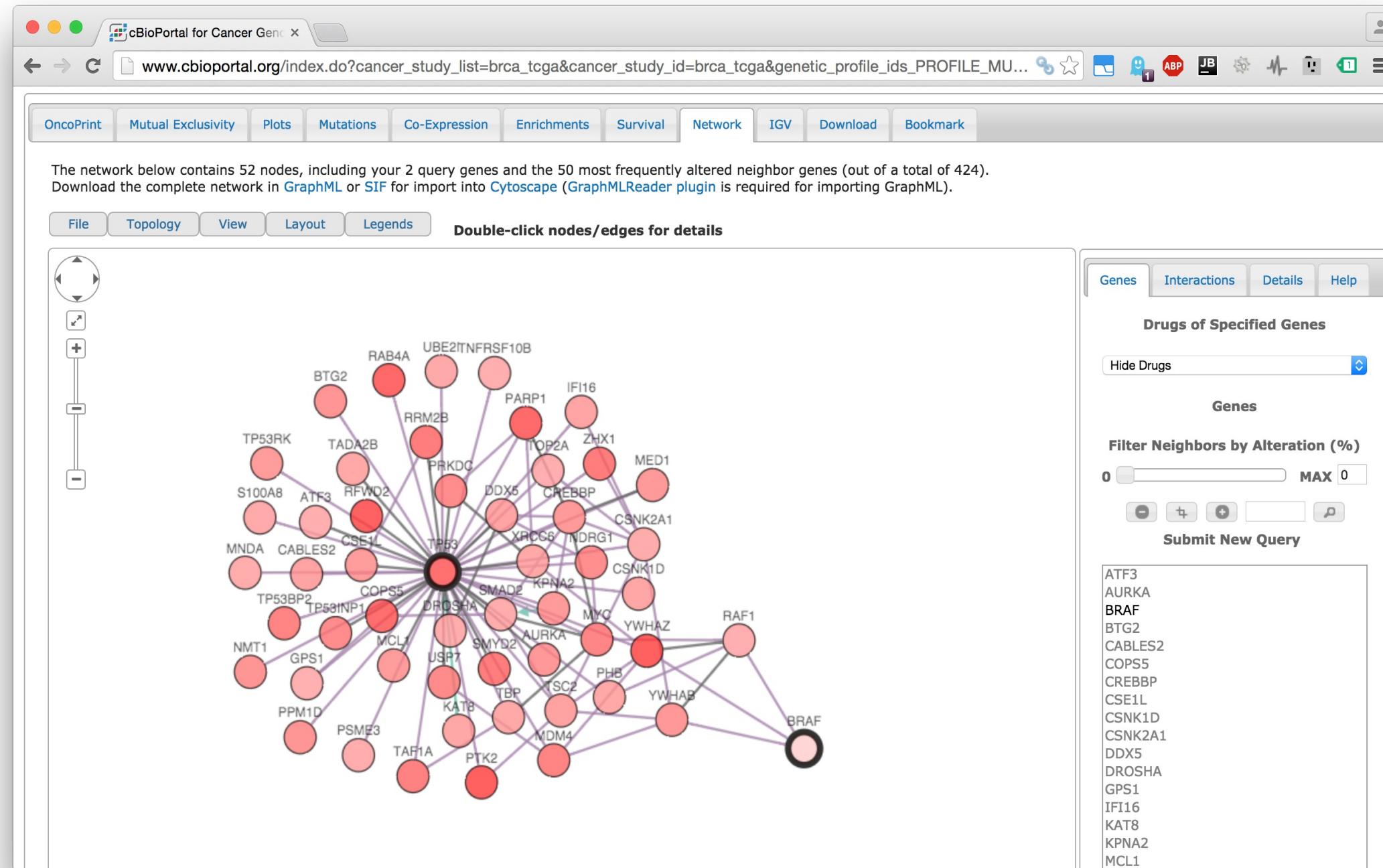
FINAL PATHOLOGIC DIAGNOSIS:
"LEFT FLANK MELANOMA", EXCISION: INVASIVE MALIGNANT MELANOMA.

- CLARK'S LEVEL: V.
- BRESLOW DEPTH: 1 CM.
- HOST RESPONSE: PRESENT.
- ULCERATION: PRIOR PROCEDURE SITE CHANGES NOTED.
- SATELLITE NODULES: PRESENT.
- TUMOR REGRESSION: NOT PRESENT.
- MITOTIC RATE: 6 MITOSES/mm²
- LYMPH-VASCULAR SPACE INVASION: PRESENT, MULTIPLE FOCI.
- PRE-EXISTING LESION: NONE IDENTIFIED.
- MARGIN: NEGATIVE.
- PATHOLOGIC STAGE (Based on AJCC/UICC TNM, 7th Ed.): pT4 N2c.

CBioPortal Patient: Copy Number



CBioPortal Patient: Network



What is cgdsr?

- Website and Tutorial (Vignette):
 - <https://cran.r-project.org/web/packages/cgdsr/index.html>
- Provides programmatic access to CBioPortal data

Getting Help

- CBioPortal Google Group
 - <http://groups.google.com/group/cbioportal>
- Biostars
 - <https://www.biostars.org/t/cbioportal/>