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

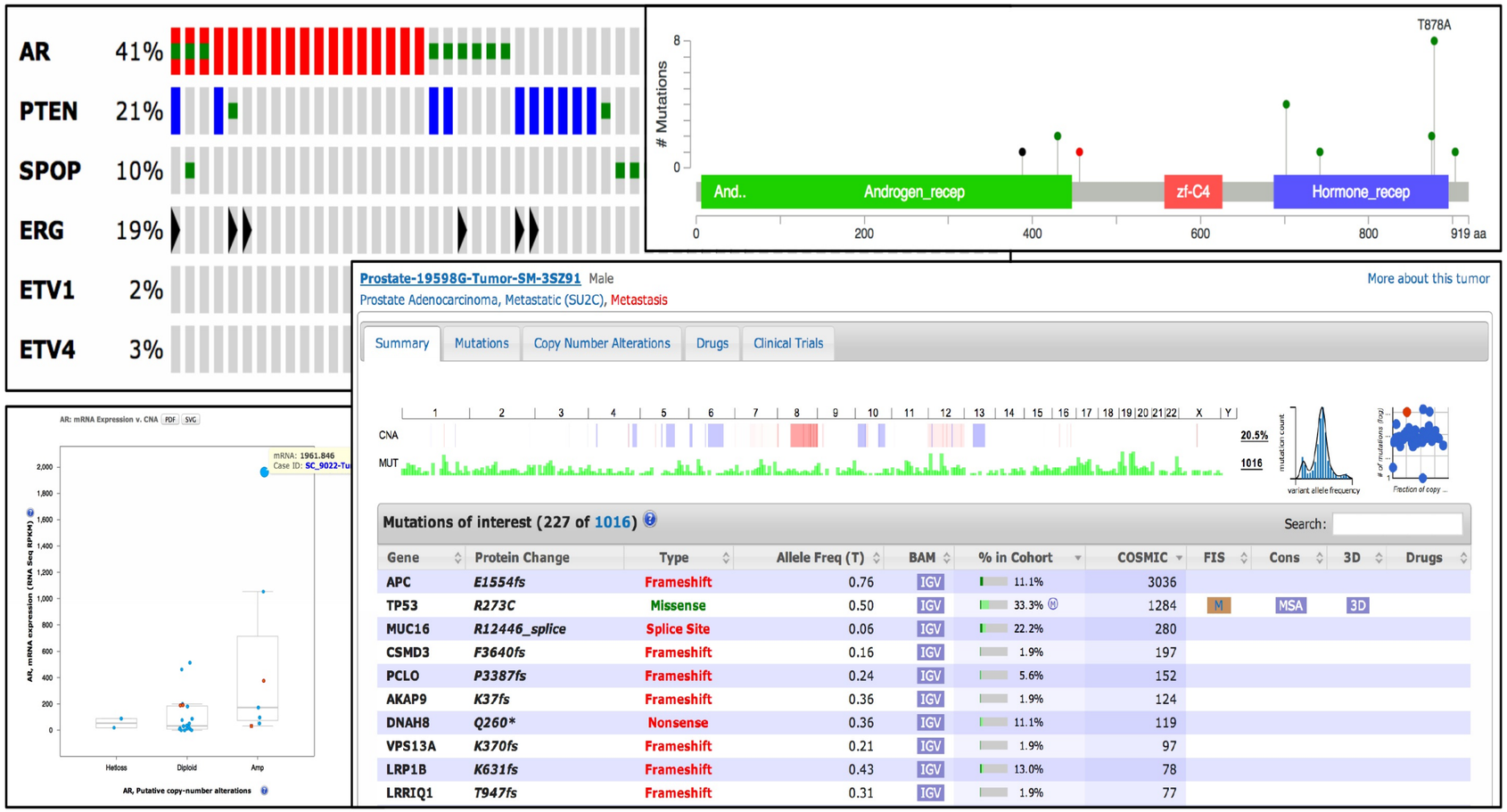
CBioPortal Data Types

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

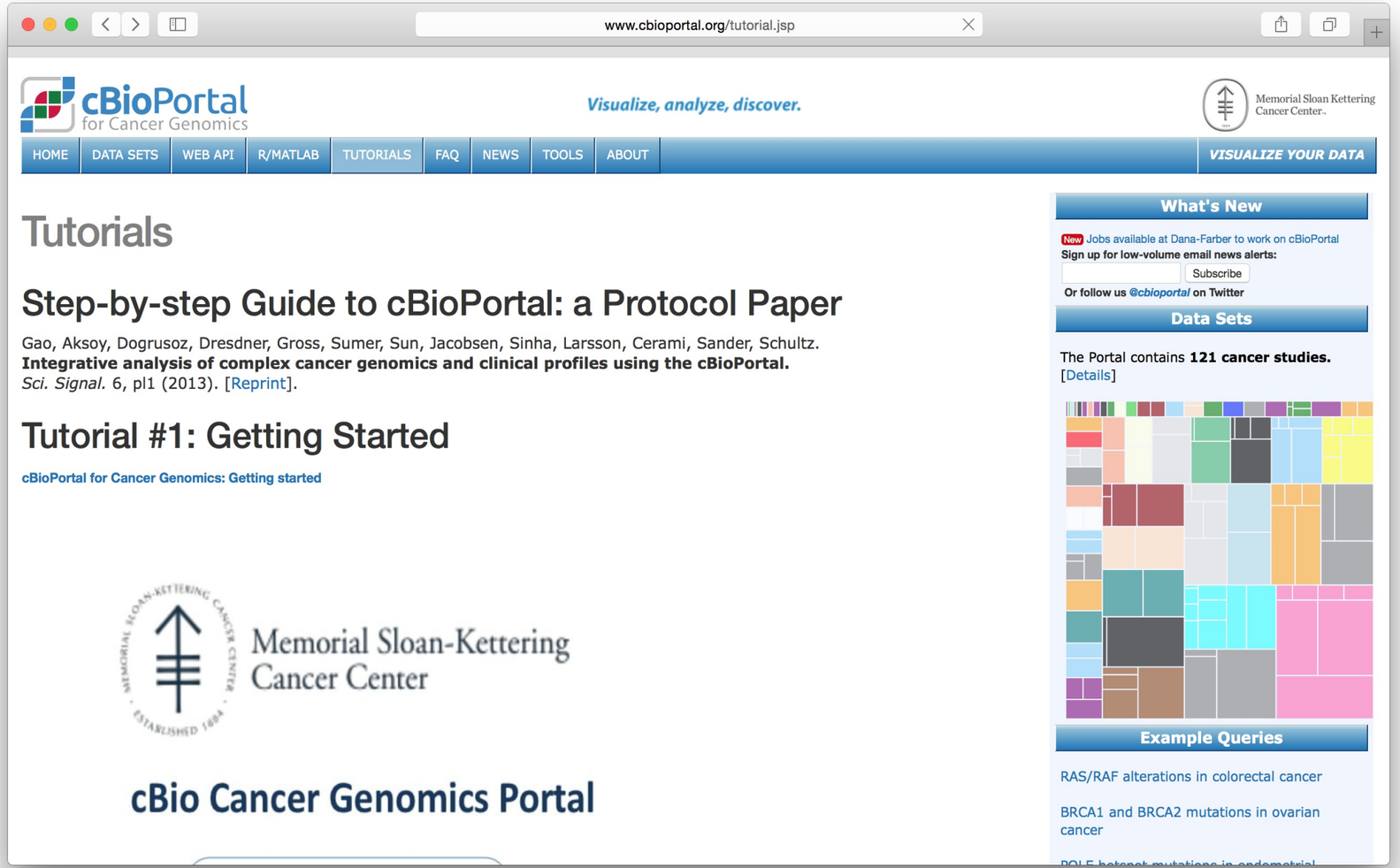
CBioPortal Data Notes

- Notes on the available data
  - http://www.cbioportal.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

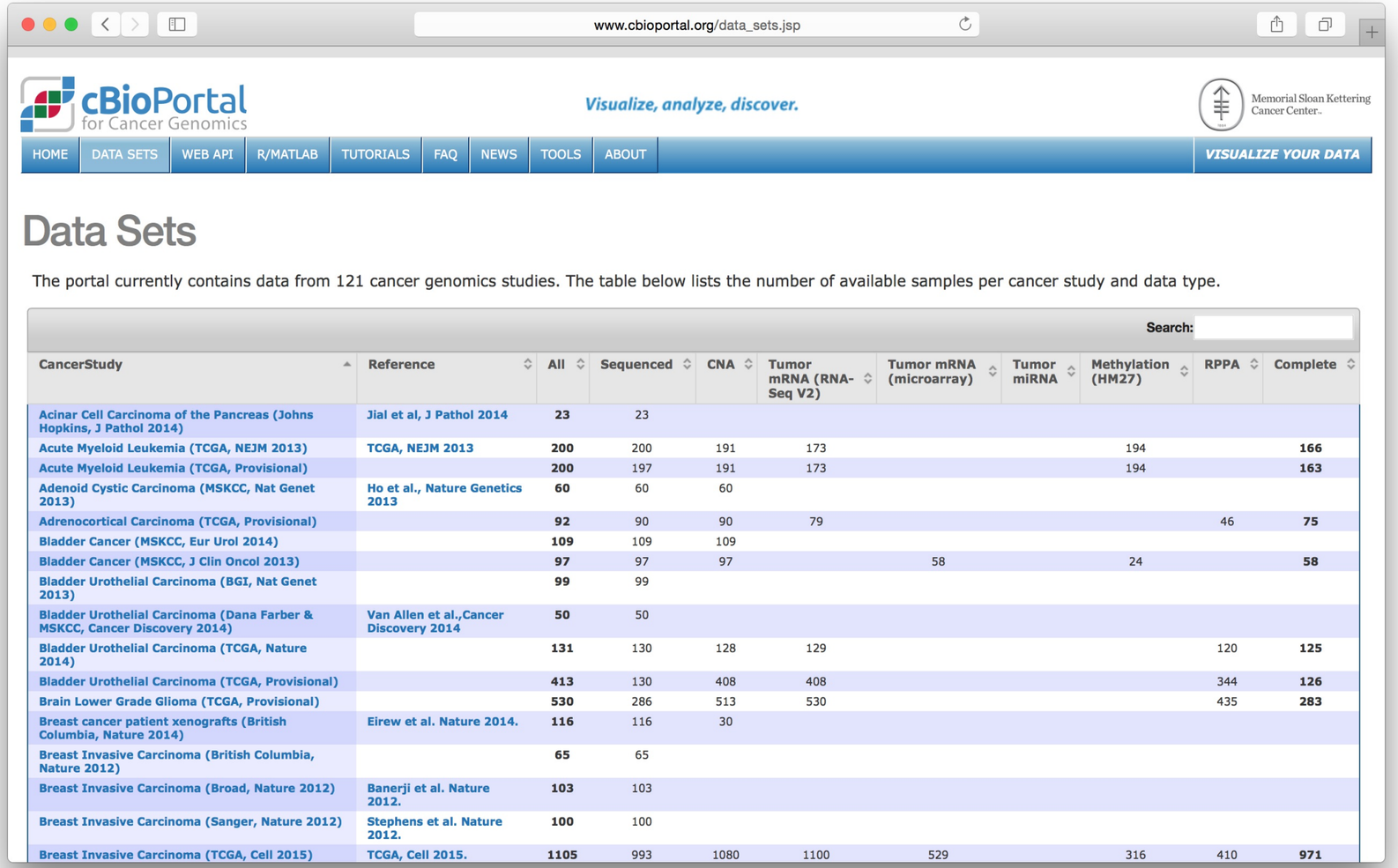
CBioPortal Website Features



CBioPortal Tutorials



CBioPortal Datasets



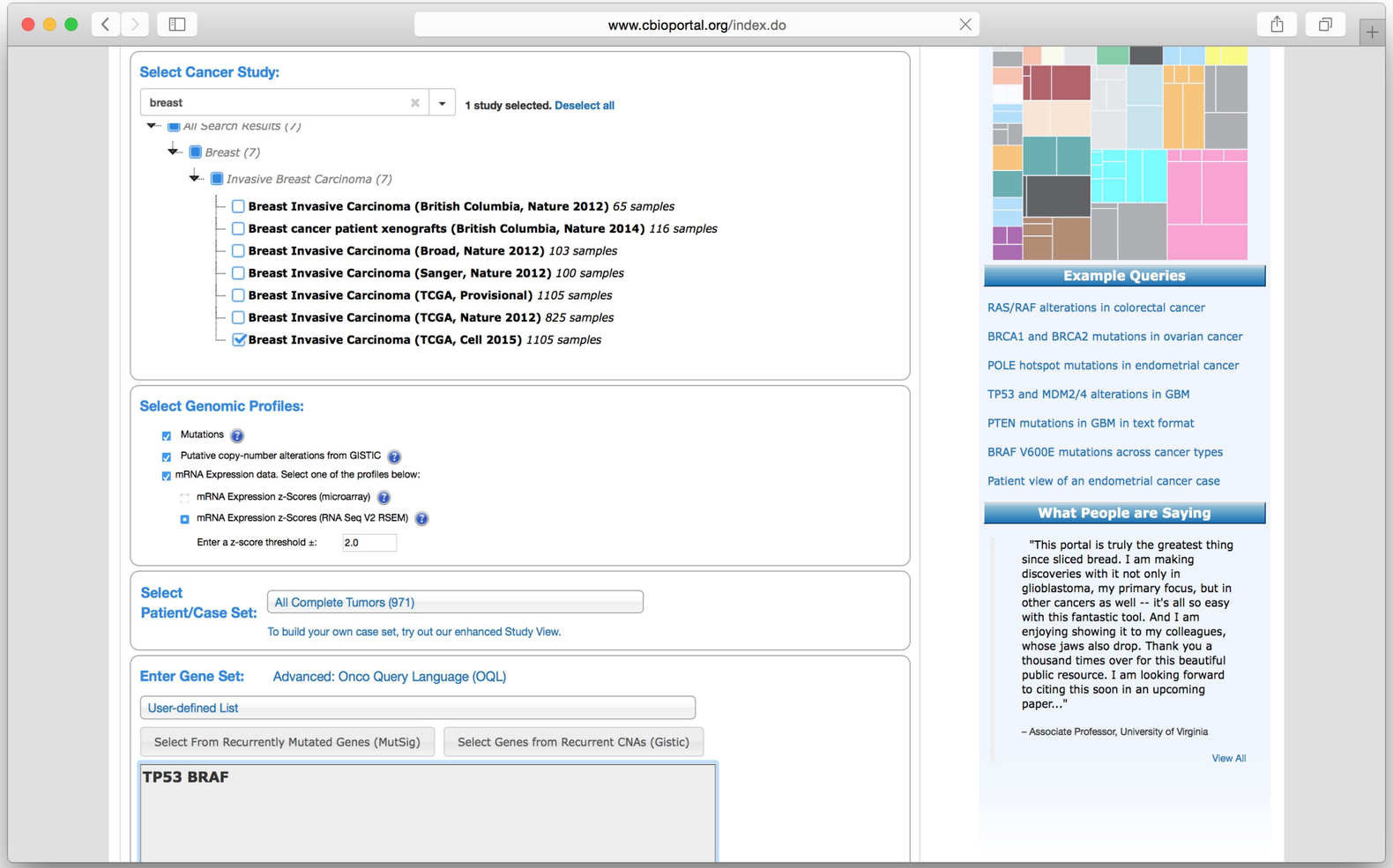
Data Sets

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.

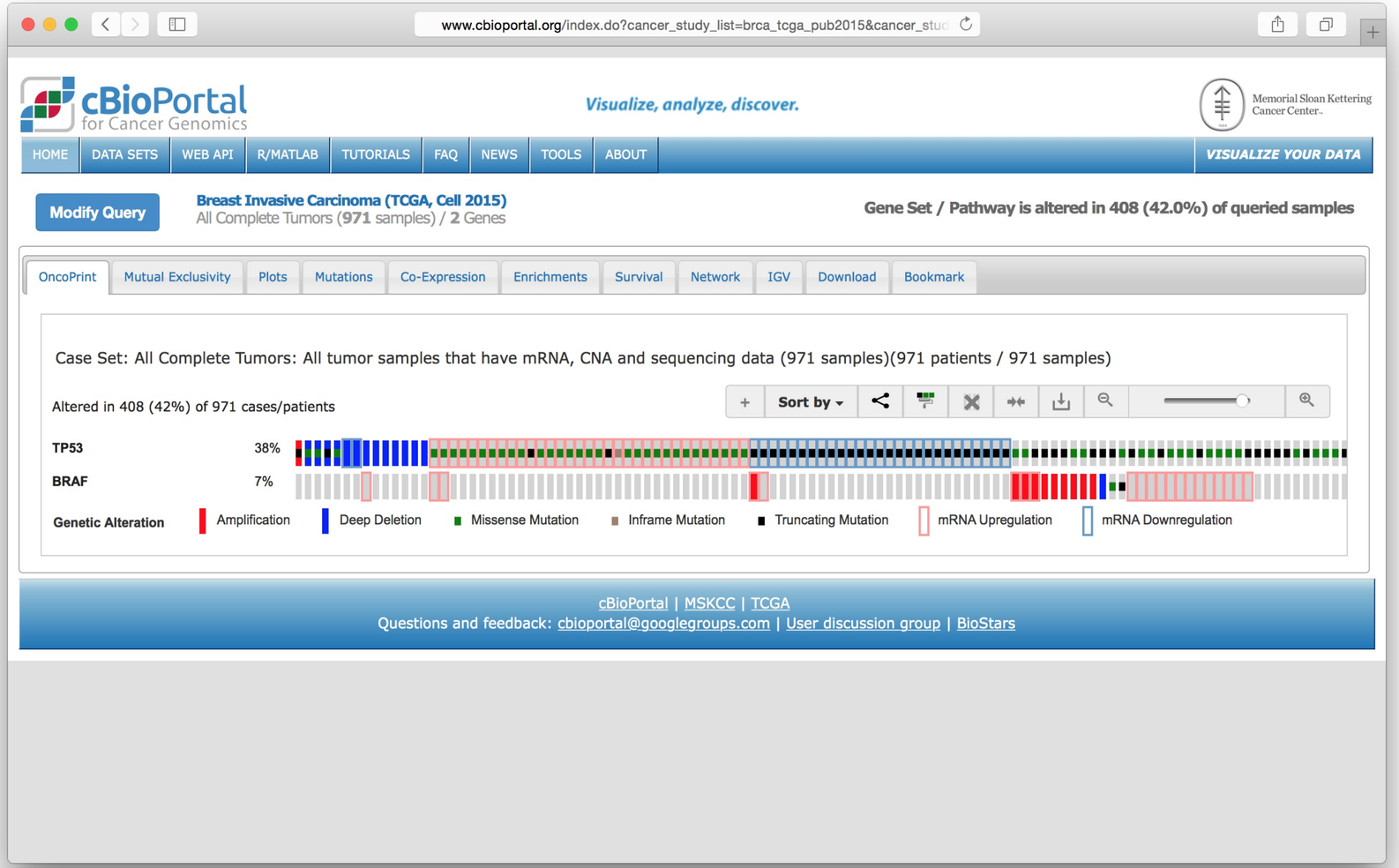
Search:

CancerStudy	Reference	All	Sequenced	CNA	Tumor mRNA (RNA-Seq V2)	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 (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**

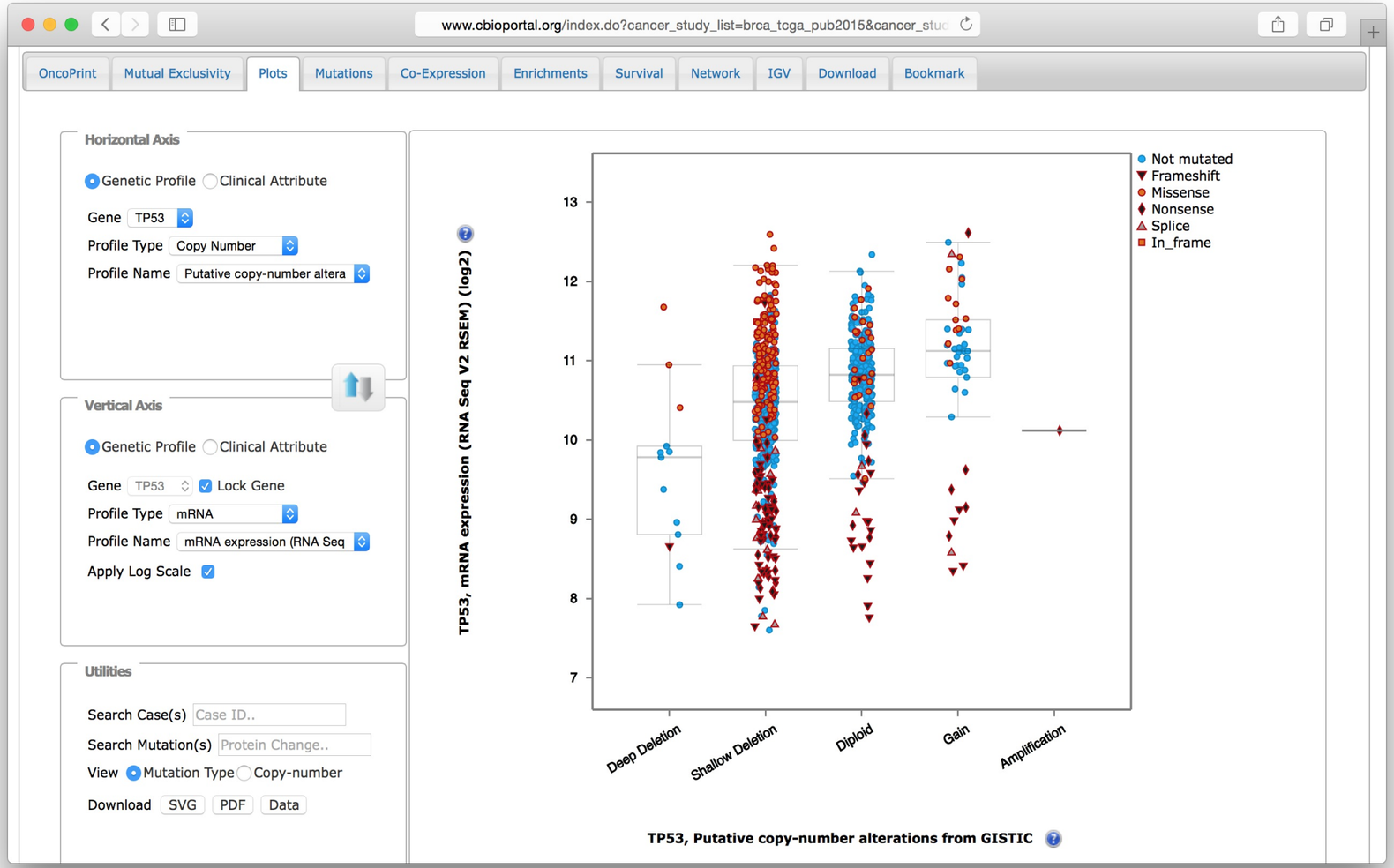
CBioPortal Home Data Selection



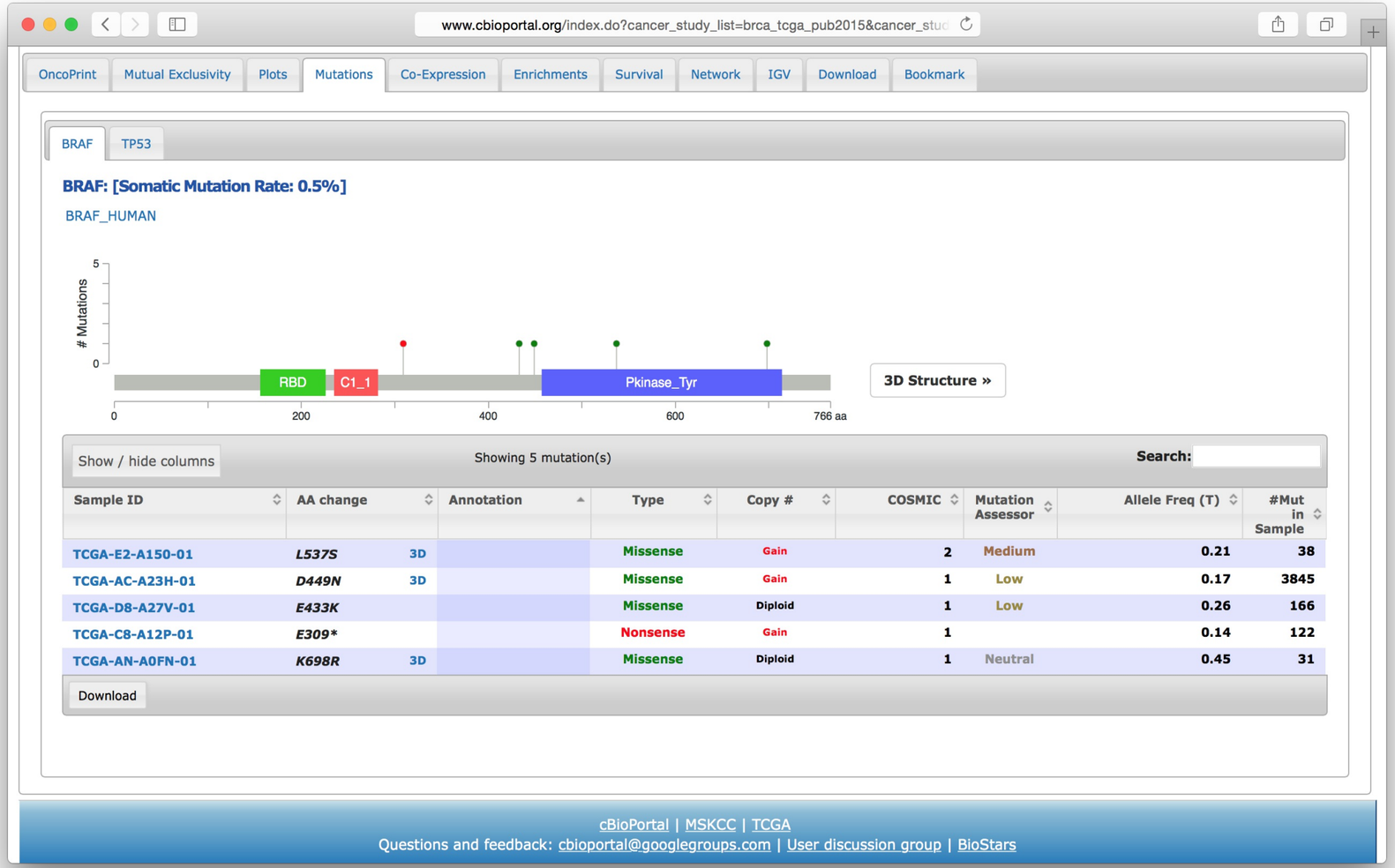
CBioPortal OncoPrint



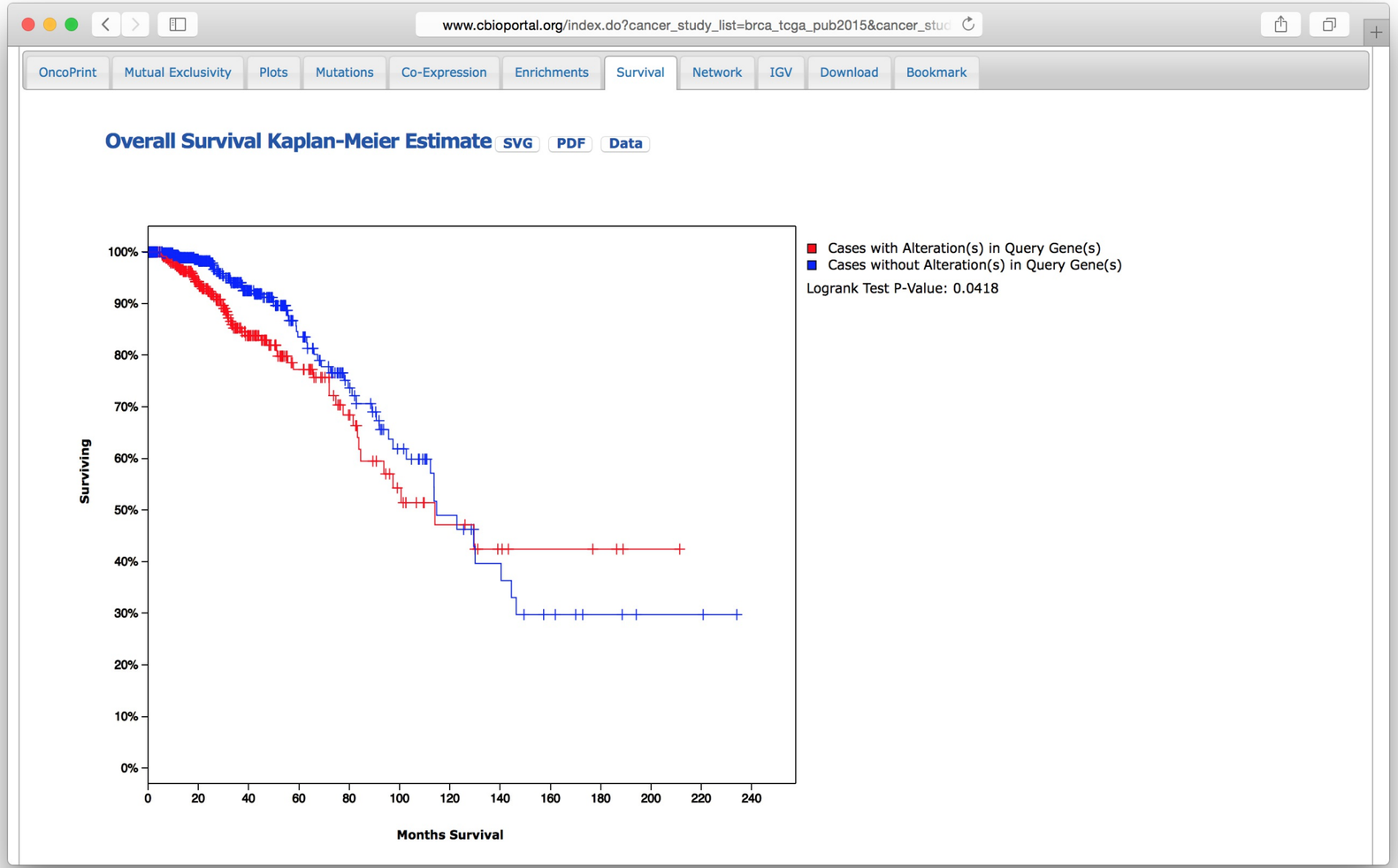
CBioPortal Plots



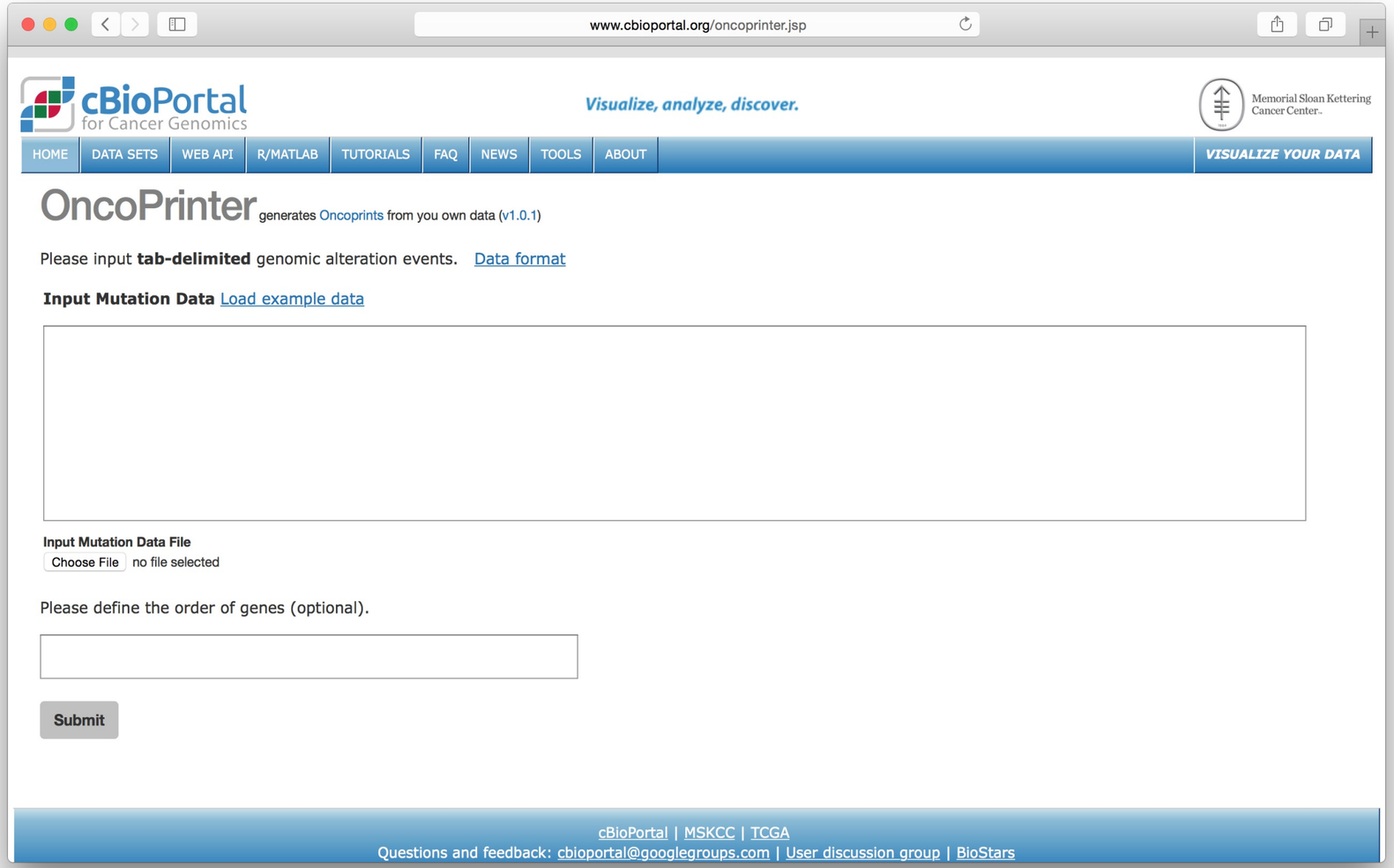
CBioPortal Mutations



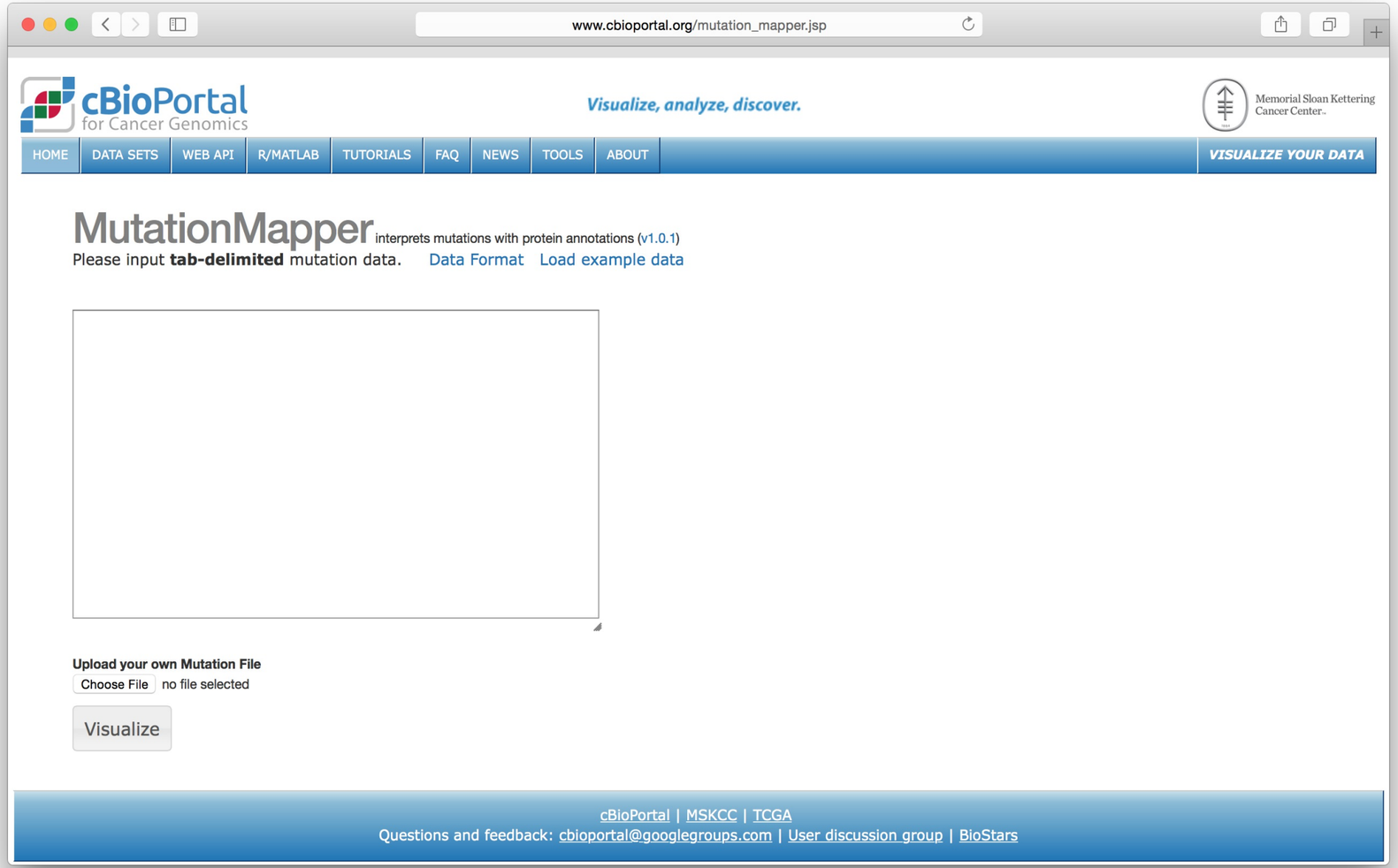
CBioPortal Survival



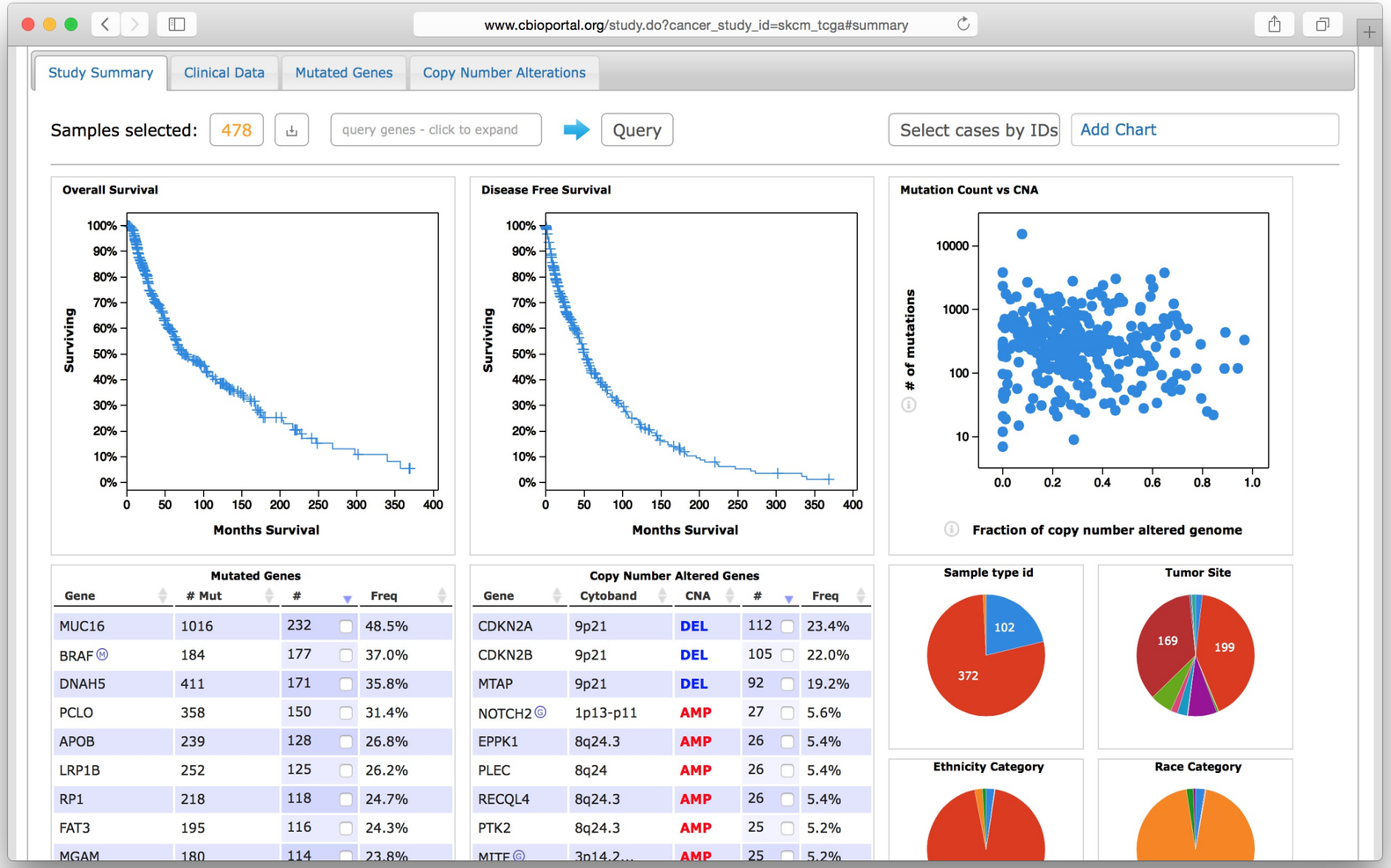
CBioPortal OncoPrinter



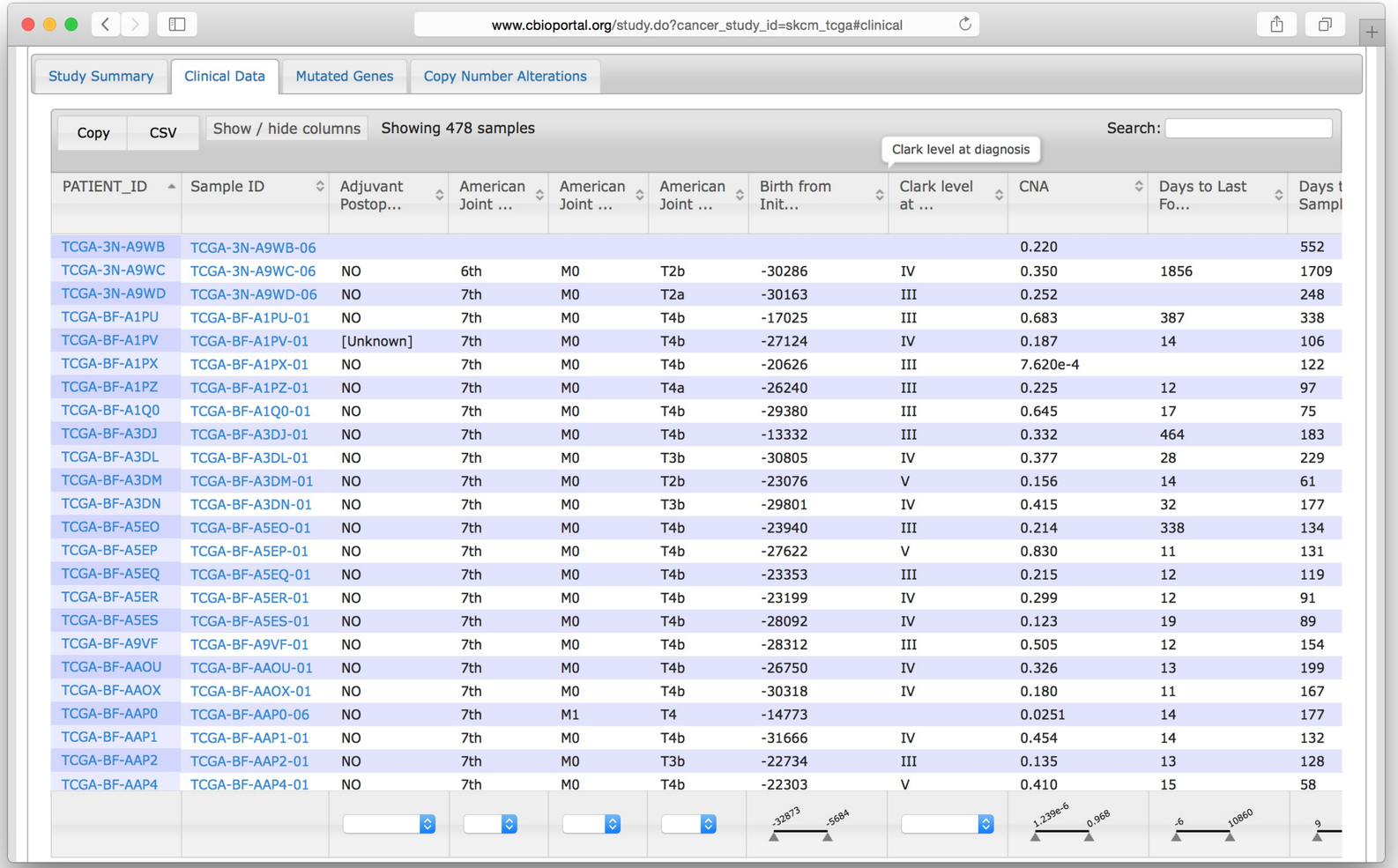
CBioPortal Mutation Mapper



CBioPortal Study: Summary

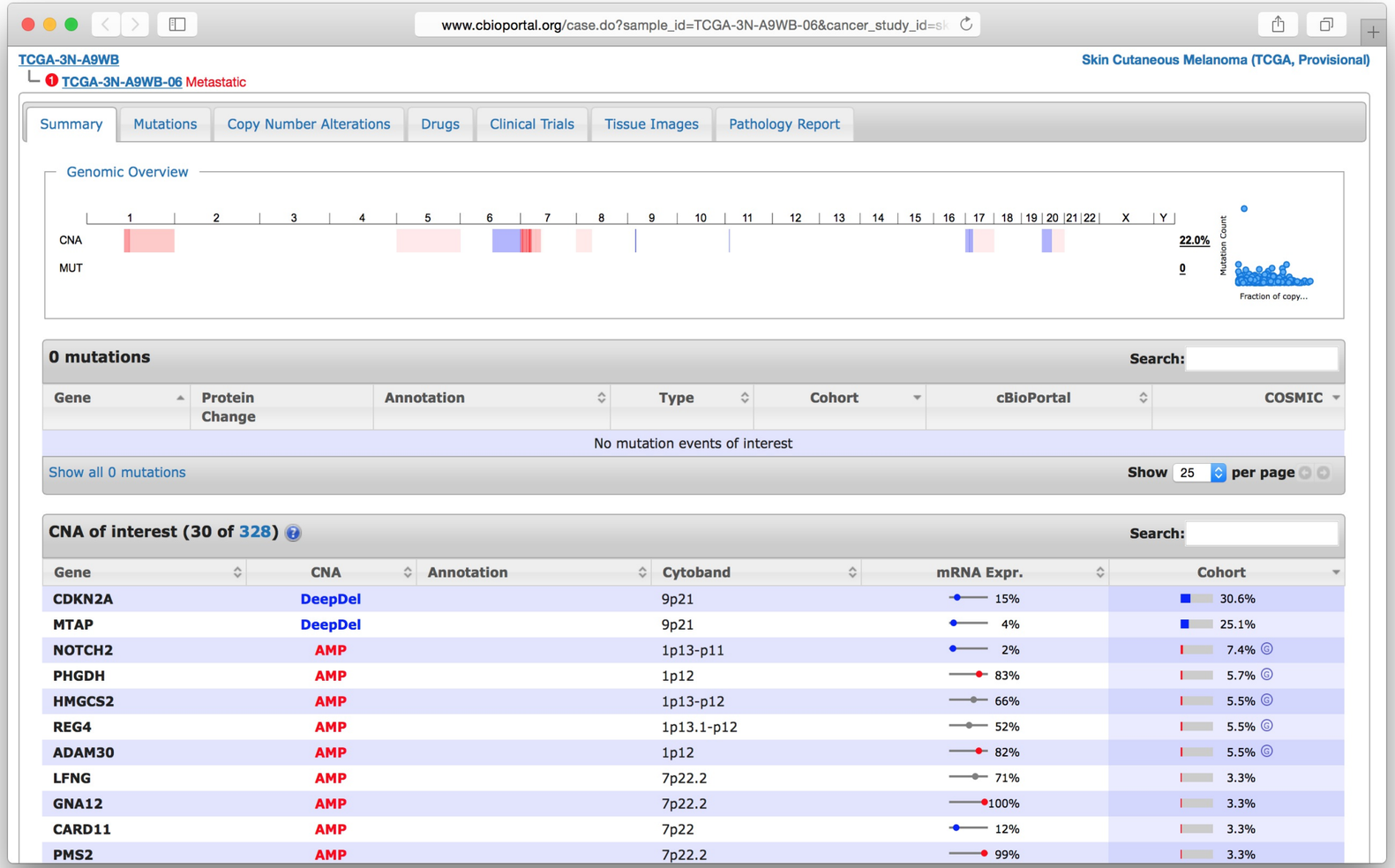


CBioPortal Study: Clinical Data

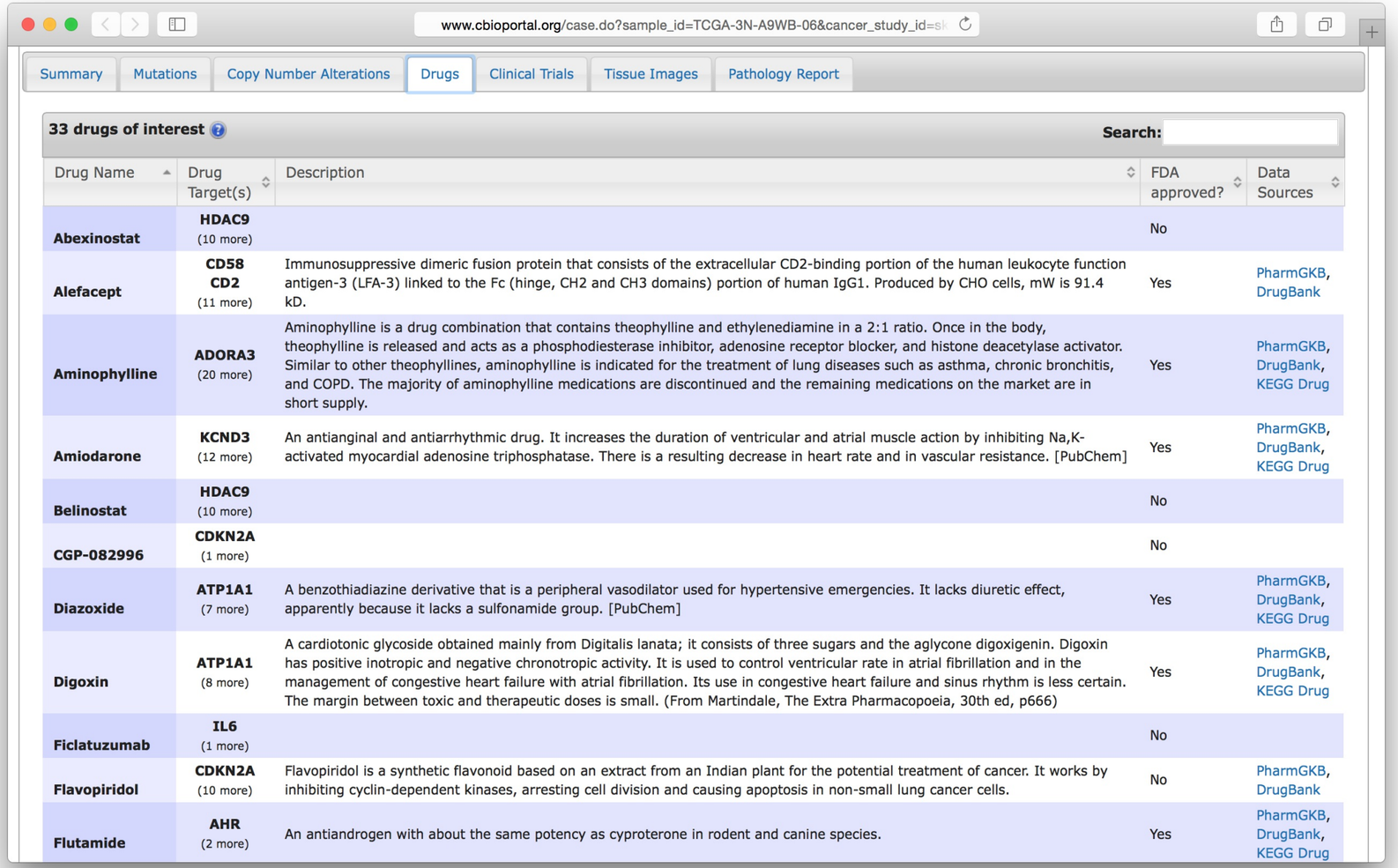


PATIENT_ID	Sample ID	Adjuvant Postop...	American Joint ...	American Joint ...	American Joint ...	Birth from Init...	Clark level at ...	CNA	Days to Last Fo...	Days t Sampl
TCGA-3N-A9WB	TCGA-3N-A9WB-06							0.220		552
TCGA-3N-A9WC	TCGA-3N-A9WC-06	NO	6th	M0	T2b	-30286	IV	0.350	1856	1709
TCGA-3N-A9WD	TCGA-3N-A9WD-06	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

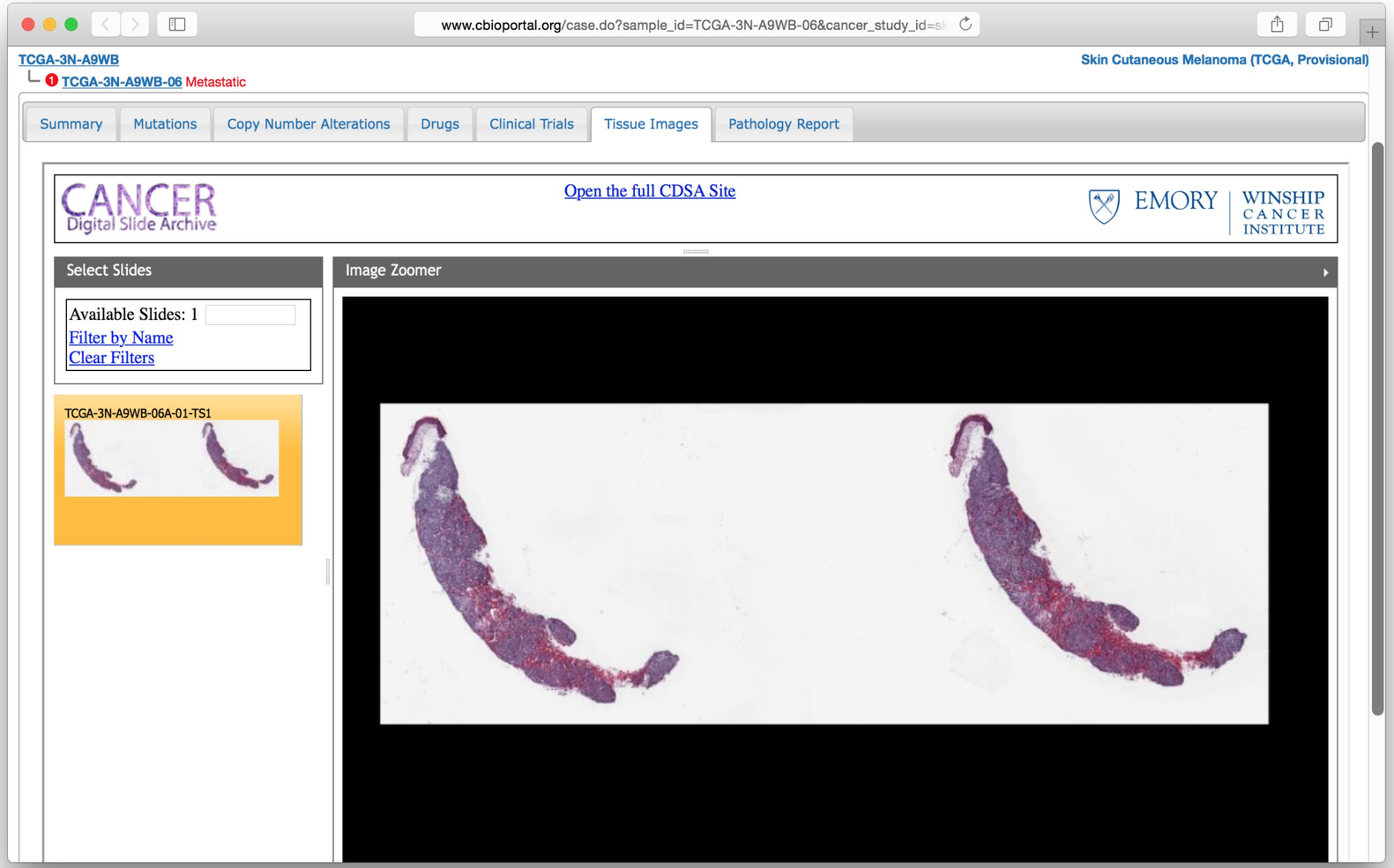


CBioPortal Patient: Drug Targets

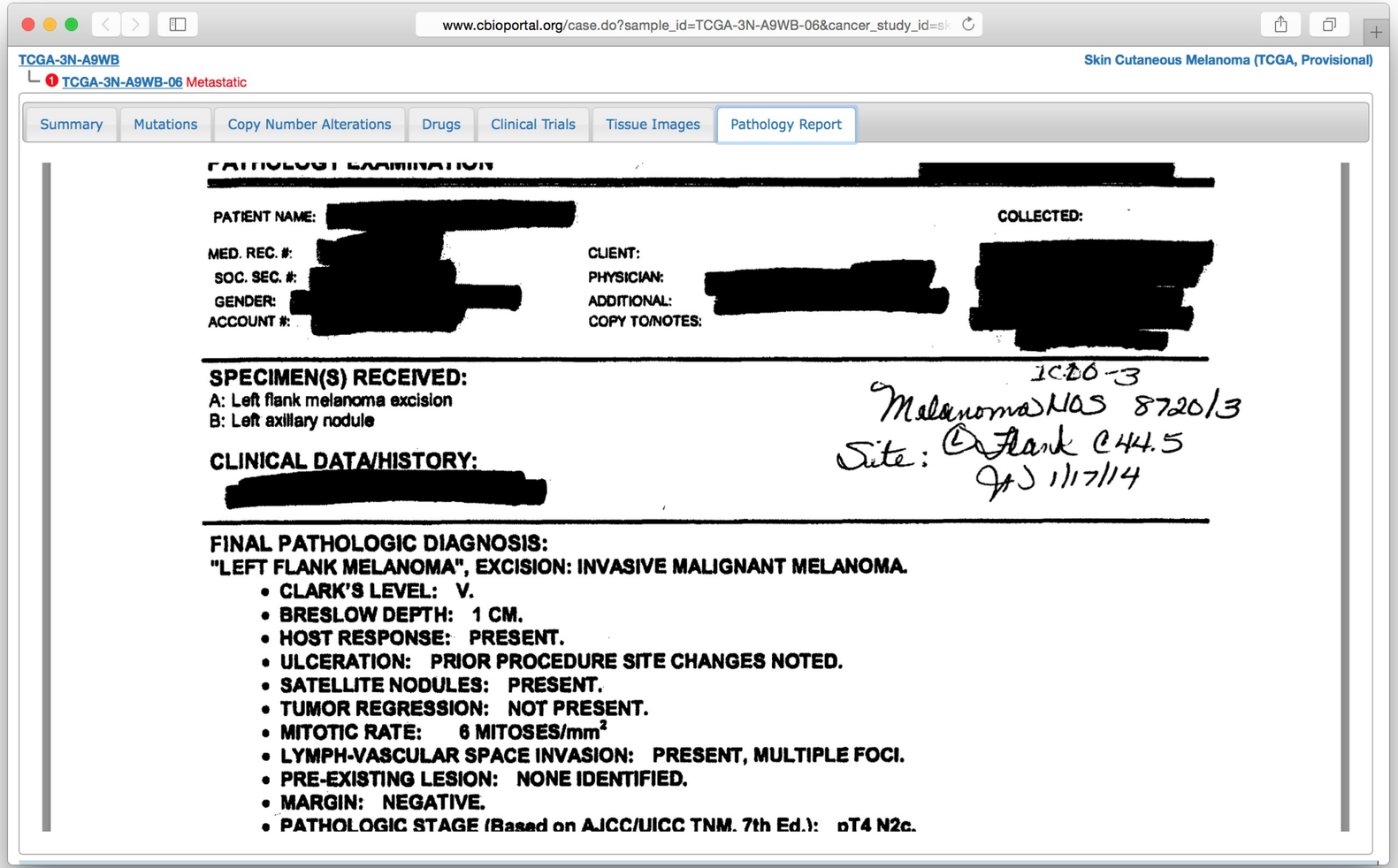


Drug Name	Drug Target(s)	Description	FDA approved?	Data Sources
Abexinostat	**HDAC9** (10 more)		No	
Alefacept	**CD58** **CD2** (11 more)	Immunosuppressive dimeric fusion protein that consists of the extracellular CD2-binding portion of the human leukocyte function antigen-3 (LFA-3) linked to the Fc (hinge, CH2 and CH3 domains) portion of human IgG1. Produced by CHO cells, mW is 91.4 kD.	Yes	PharmGKB, DrugBank
Aminophylline	**ADORA3** (20 more)	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	**KCND3** (12 more)	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	**HDAC9** (10 more)		No	
CGP-082996	**CDKN2A** (1 more)		No	
Diazoxide	**ATP1A1** (7 more)	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	**ATP1A1** (8 more)	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	**IL6** (1 more)		No	
Flavopiridol	**CDKN2A** (10 more)	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	**AHR** (2 more)	An antiandrogen with about the same potency as cyproterone in rodent and canine species.	Yes	PharmGKB, DrugBank, KEGG Drug

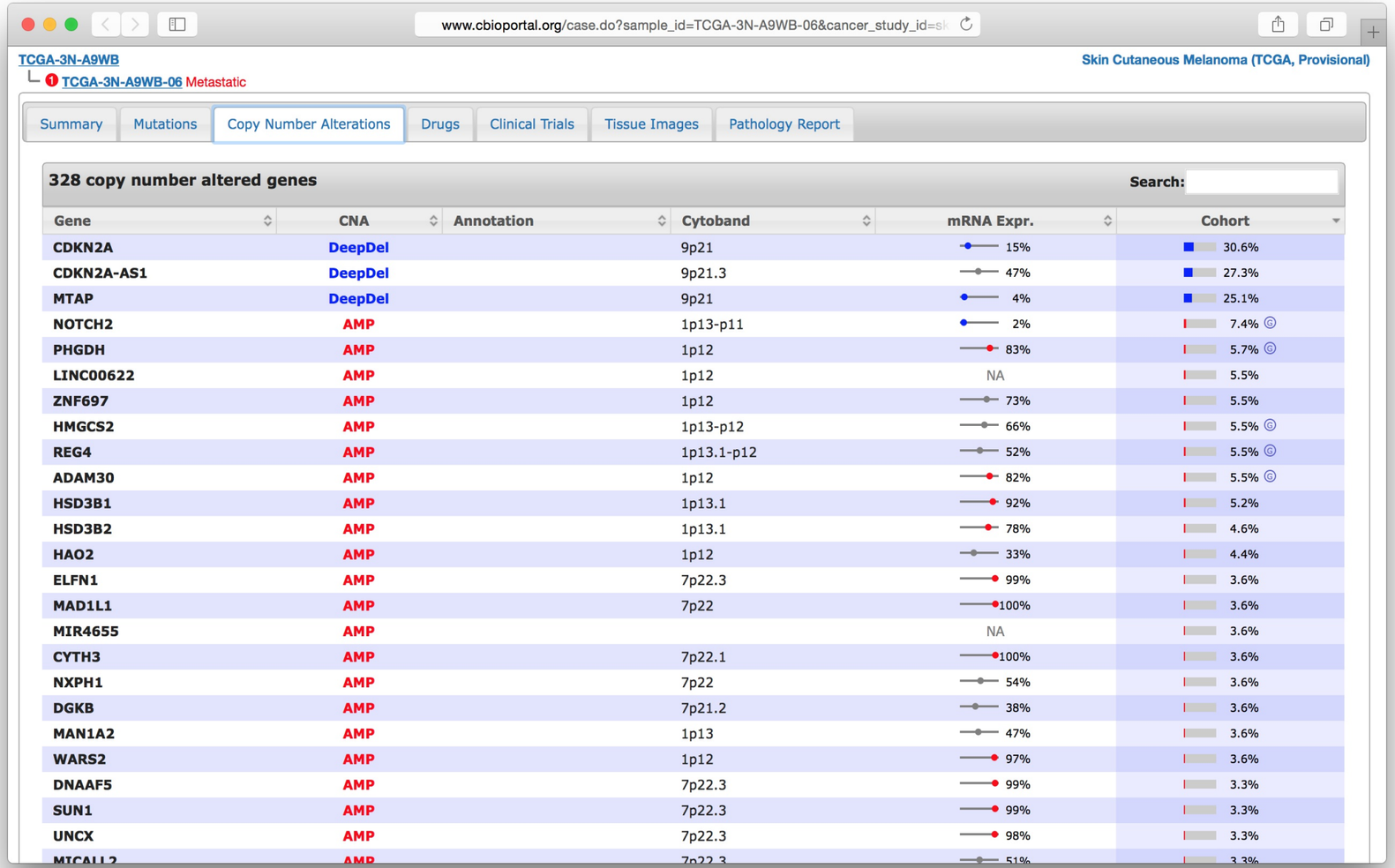
CBioPortal Patient: Histology



CBioPortal Patient: Pathology

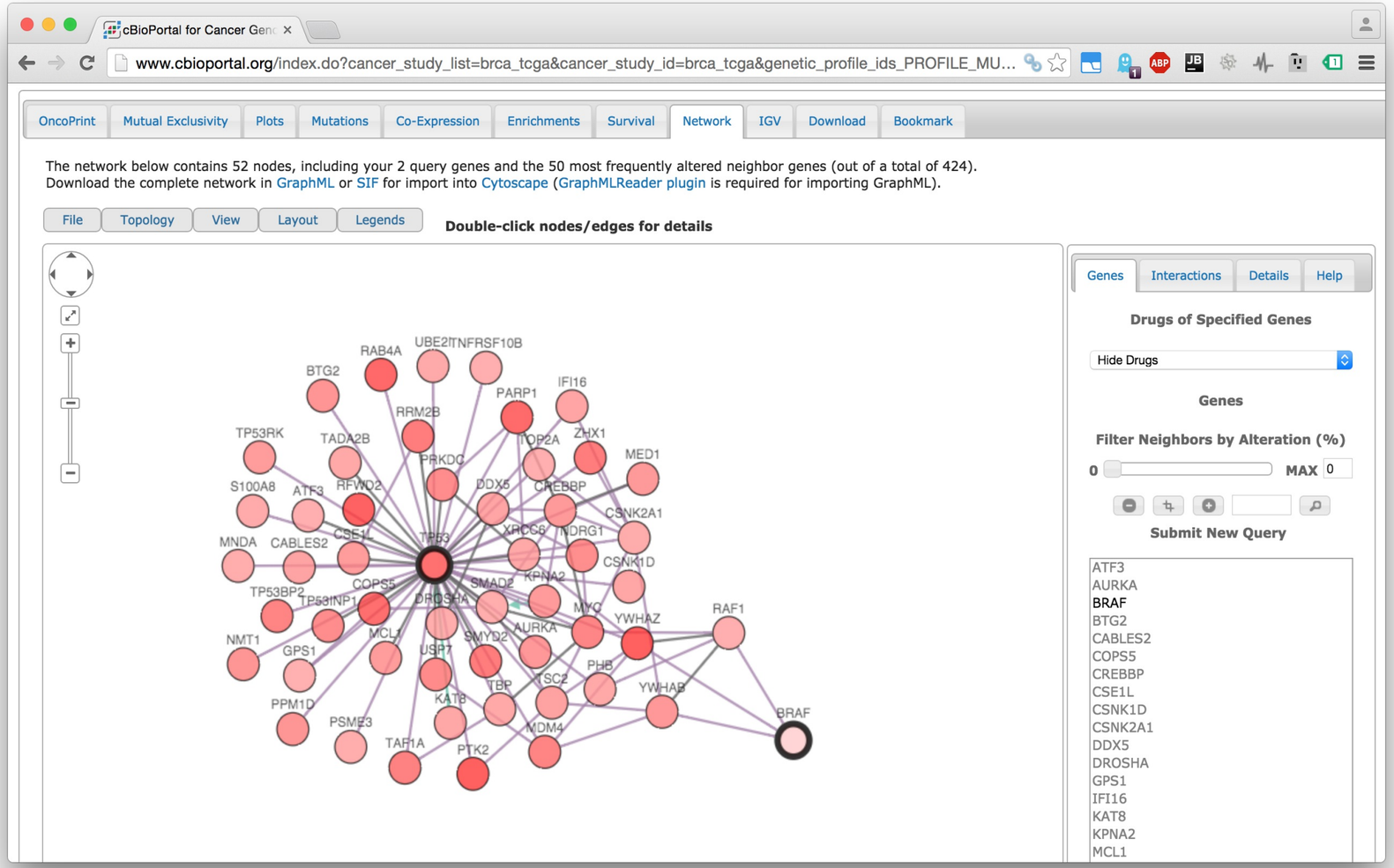


CBioPortal Patient: Copy Number



Gene	CNA	Annotation	Cytoband	mRNA Expr.	Cohort
CDKN2A	DeepDel		9p21	15%	30.6%
CDKN2A-AS1	DeepDel		9p21.3	47%	27.3%
MTAP	DeepDel		9p21	4%	25.1%
NOTCH2	AMP		1p13-p11	2%	7.4%
PHGDH	AMP		1p12	83%	5.7%
LINC00622	AMP		1p12	NA	5.5%
ZNF697	AMP		1p12	73%	5.5%
HMGCS2	AMP		1p13-p12	66%	5.5%
REG4	AMP		1p13.1-p12	52%	5.5%
ADAM30	AMP		1p12	82%	5.5%
HSD3B1	AMP		1p13.1	92%	5.2%
HSD3B2	AMP		1p13.1	78%	4.6%
HAO2	AMP		1p12	33%	4.4%
ELFN1	AMP		7p22.3	99%	3.6%
MAD1L1	AMP		7p22	100%	3.6%
MIR4655	AMP			NA	3.6%
CYTH3	AMP		7p22.1	100%	3.6%
NXPH1	AMP		7p22	54%	3.6%
DGKB	AMP		7p21.2	38%	3.6%
MAN1A2	AMP		1p13	47%	3.6%
WARS2	AMP		1p12	97%	3.6%
DNAAF5	AMP		7p22.3	99%	3.3%
SUN1	AMP		7p22.3	99%	3.3%
UNCX	AMP		7p22.3	98%	3.3%
MICALL2	AMP		7p22.3	51%	3.3%

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/