Navigating cBioPortal for Cancer Genomics

Anand S. Merchant

CCRIFX Bioinformatics Core

http://ccrifx.cancer.gov

Acknowledgements

This presentation was developed by the CCRIFX core of the Advanced Biomedical Computing Center operated by SAIC-Frederick. Funding for CCRIFX comes from the CCR Bioinformatics Initiative (YT-10-060).

CCRIFX

Natalie Abrams, PhD

Yvonne Edwards, PhD

Parthav Jailwala, MS

Li Jia, MS

Manjula Kasoji, MS

Nathan Starner, BS

Anand S Merchant, MD, PhD

Eric Stahlberg, PhD

ABCC

Brian Luke, PhD

Uma Mudunuri, MS

Jack Collins, PhD

CCR experts

Maggie Cam, PhD

Sean Davis, MD, PhD

Max Lee, PhD

Peter FitzGerald, PhD

David Goldstein, PhD

Sequencing Facility

Yongmei Zhao, MS

Bao Tran, MS

Outline for today

- Introduction to the web application – What is the purpose?
- Website walk-through – Where is the information and how to query it?
- Overview of the Cancer and Data Types
- Advantages and Limitations
- Analysis Features – 'Clickalytics'
- Viewing and Interpretation of results
- Example Case - TCGA Breast Cancer dataset (2012 Nature publication)
- Advanced OncoQueryLanguage (OQL) - Key words and Codes
- References/Tutorials/FAQ/Pre-set queries
- Q&A

What is the cBioPortal for Cancer Genomics?

- Developed and maintained by the Computational Biology Center @ MSKCC – www.cbioportal.org
- The portal stores genomic data from large scale, integrated cancer genomic data sets
- Provides simple download of small data slices
- Common uses
  - Omics data visualization and exploratory analysis
  - Patient portal: match patient to drugs & clinical trials
  - Identification of biomarkers and therapeutic targets
  - Evaluation of MDx and DTC test results

Cancer and Data Types

- Patient samples from 22 different tissues
- 56 cancer datasets totaling 15506 samples
- Cancer Cell Line Encyclopedia (CCLE)
- Multiple data types – Mutations, Copy Number, mRNA/miR expression, methylation, protein/phosphorylation data

Advantages

- Comprehensive coverage of TCGA datasets
- Standardized analytics
- Visualization modes
- Clinical data made available
- Cross-cancer exploration
- Dataset filters
- Programmatic interfaces

Limitations

- No representation of non-TCGA data
- More focus on genomic as compared to transcriptomic data

Querying across all cancers

Memorial Sloan-Kettering Cancer Center

cBioPortal for Cancer Genomics

Visualize, analyze, discover.

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

Modify Query

Select Cancer Study: All Cancer Studies

Select cancer study

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

Select data type

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

General: PI3K-AKT-mTOR signaling (17 genes)

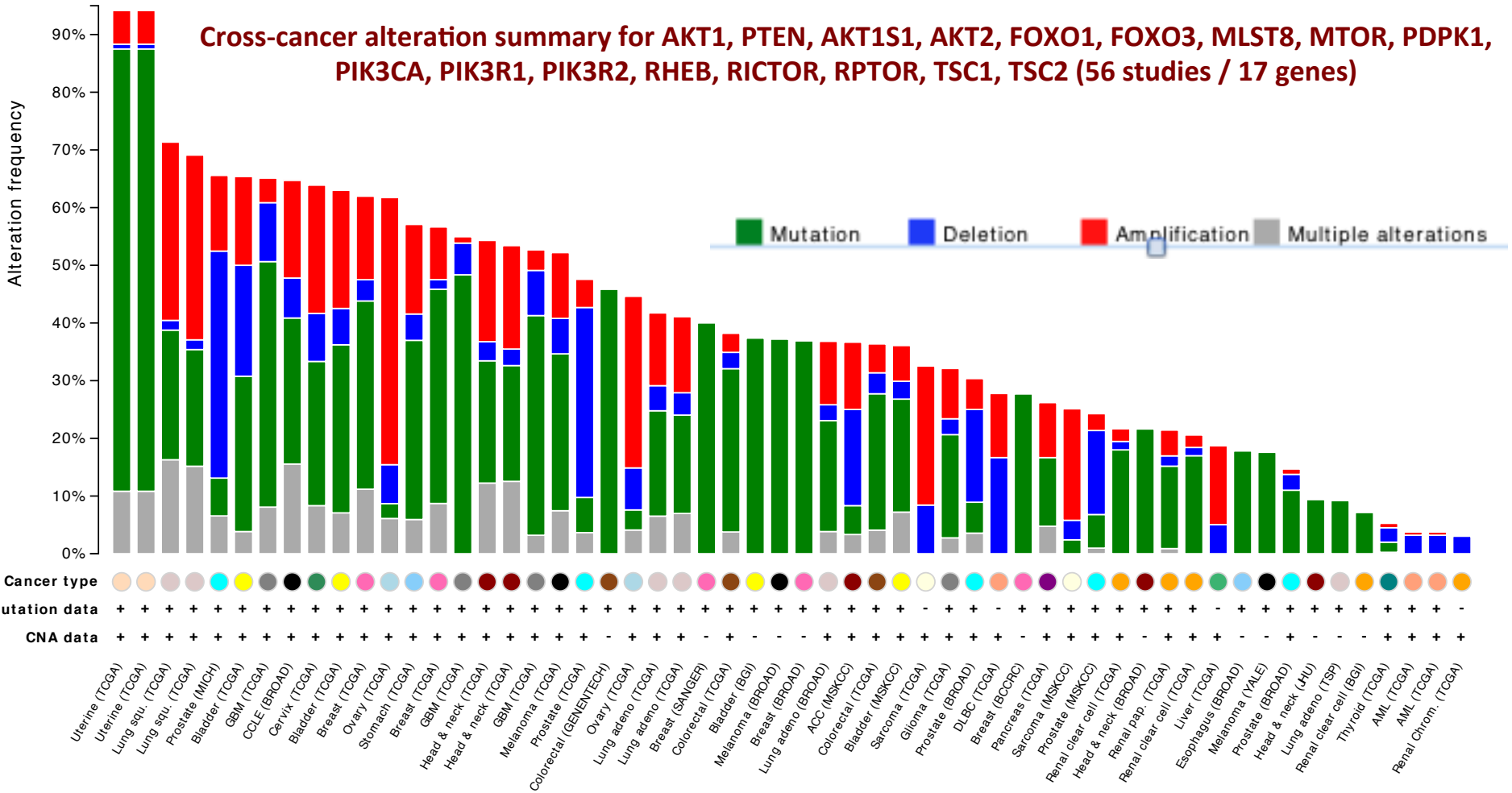
Enter gene symbols

PIK3CA PIK3R1 PIK3R2 PTEN PDPK1 AKT1 AKT2 FOXO1 FOXO3 MTOR RICTOR TSC1 TSC2 RHEB AKT1S1 RPTOR MLST8

All gene symbols are valid.

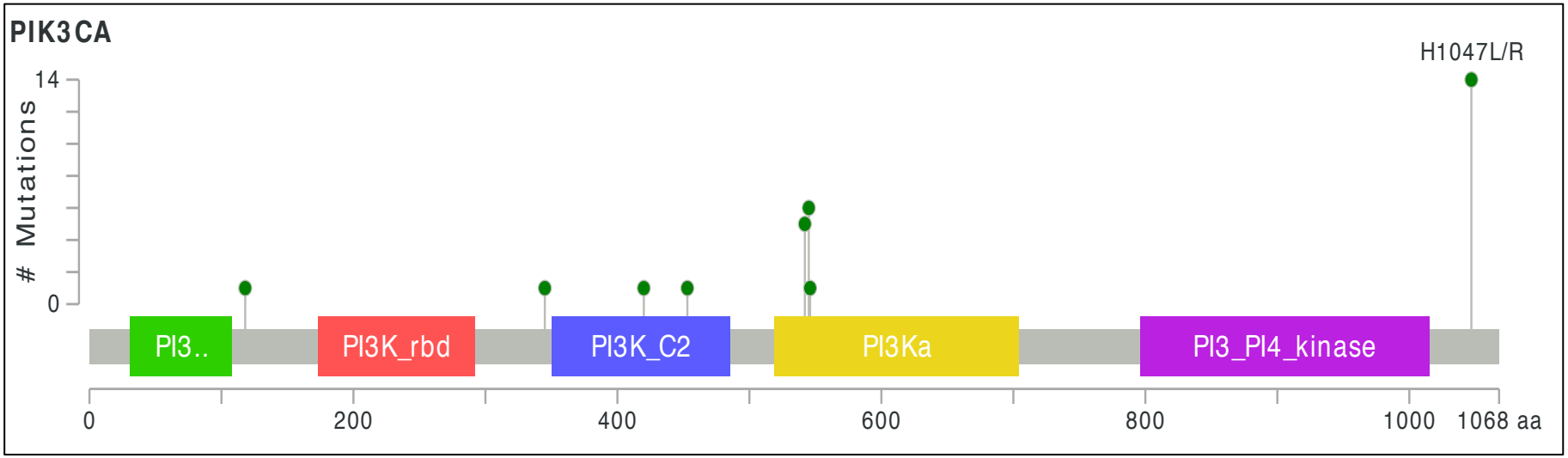
Submit

Generating results across cancers



CANCER TYPES

Mutation Chart for Query Gene



Case ID	AA Change	Type	COSMIC	FIS	Cons	3D	VS	Allele Freq (T)	#Mut in Sample
TCGA-AN-A0XO	E542K	Missense	693	L	msa	3D	U	NA	52
TCGA-E2-A153	E545K	Missense	1245	L	msa	3D	U	NA	20
TCGA-E2-A10C	E545K	Missense	1245	L	msa	3D	U	NA	286
TCGA-BH-A0W7	E726K	Missense	13	L	msa	3D	U	NA	125
TCGA-BH-A0HU	GNR106del	IF del	1				U	NA	34
TCGA-A8-A0A1	E545K	Missense	1245	L	msa	3D	U	NA	29
TCGA-A8-A08T	E545K	Missense	1245	L	msa	3D	U	NA	37
TCGA-A8-A08B	E545K	Missense	1245	L	msa	3D	U	NA	21
TCGA-A8-A075	E545K	Missense	1245	L	msa	3D	U	NA	51
TCGA-A8-A06P	E545K	Missense	1245	L	msa	3D	U	NA	20

TCGA Breast Nature 2012

- **Comprehensive molecular portraits of human breast tumours**

- **Tissues -** Primary breast cancers
- **Assays –**
  - genomic DNA copy number arrays,
  - DNA methylation,
  - exome sequencing,
  - messenger RNA arrays,
  - microRNA sequencing and
  - reverse-phase protein arrays

- **Findings**
  - Somatic mutations in only three genes (TP53, PIK3CA and GATA3) occurred at >10% incidence across all breast cancers;
  - numerous subtype-associated and novel gene mutations including the enrichment of specific mutations in GATA3, PIK3CA and MAP3K1 with the luminal A subtype.

- **Nature. 2012 Oct 4;490(7418):61-70. doi: 10.1038/nature11412. Epub 2012 Sep 23.**

Querying Single Cancer Datasets

Select Cancer Study: Breast Invasive Carcinoma (TCGA, Nature 2012)

The Cancer Genome Atlas (TCGA) Breast Invasive Carcinoma project. 825 cases.

Nature 2012. Raw data via the TCGA Data Portal. TCGA, Nature 2012 View details

Select cancer study

Select Genomic Profiles:

- [x] Mutations
- [x] Putative copy-number alterations from GISTIC
- [x] mRNA Expression data. Select one of the profiles below:
  - ( ) mRNA Expression z-Scores (microarray)
  - (•) mRNA/miRNA expression Z-scores (all genes)
    - Enter a z-score threshold ±: 2.0
- [x] protein/phosphoprotein level (RPPA)
  - Enter a RPPA z-score threshold ±: 2.0

Select omic profile

Select Patient/Case Set: Complete samples (mutations, copy-number, expressio... Build Case Set

Select patient set

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

User-defined List

Select From Recurrently Mutated Genes (MutSig)

PIK3CA TP53 MAP3K1 KMT2C GATA3 CDH1 MAP2K4

All gene symbols are valid.

Enter gene symbols

Submit

Screenshot of Results

OncoPrint | Mutual Exclusivity | Plots | Mutations | Protein Changes | Survival | Network | IGV | Download | Bookmark

OncoPrint (What are OncoPrints?) PDF SVG

Q: Are there any differences between PIK3CA and TP53 mutations in this sample set?

Customize

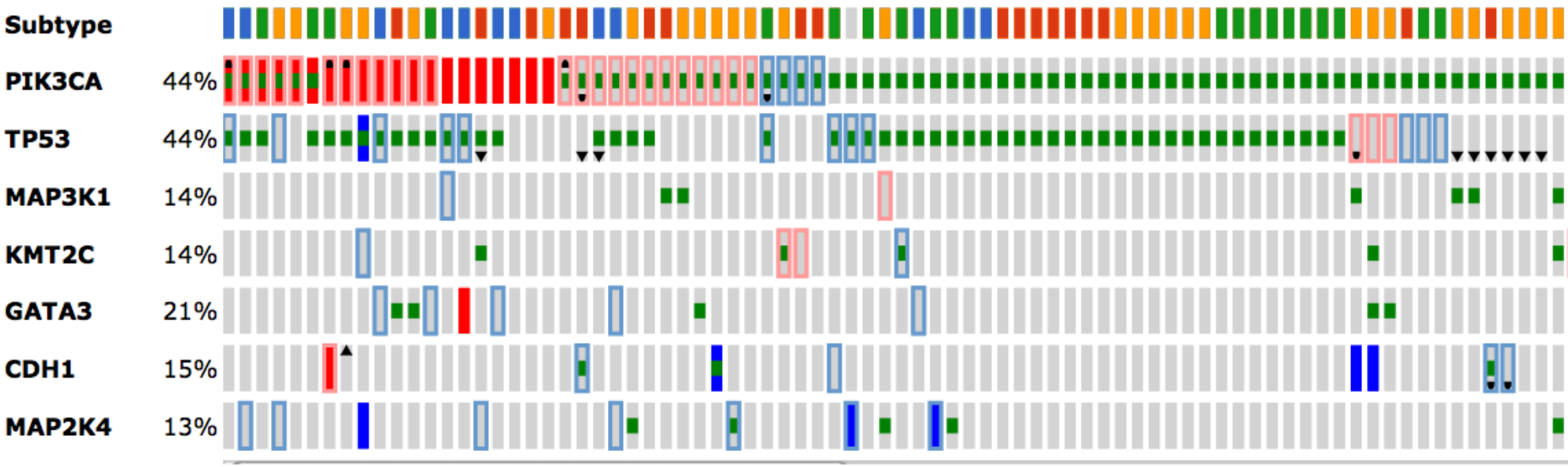
Zoom

☑ Remove Unaltered Cases ☐ Remove Whitespace

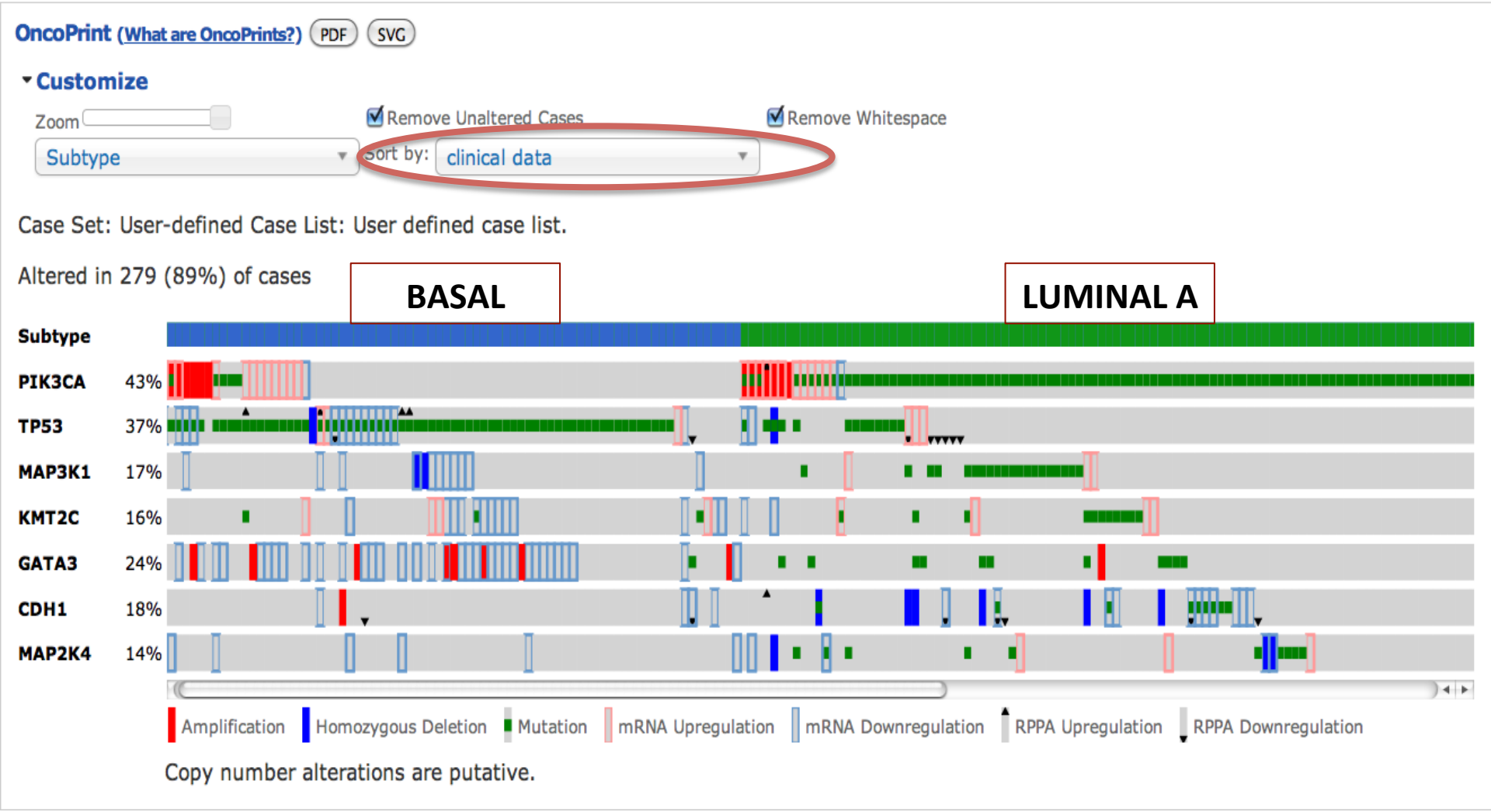
Subtype Sort by: gene data

Case Set: Complete samples (mutations, copy-number, expression): Samples with complete data (463 samples)

Altered in 411 (89%) of cases



Basal samples have more TP53 mutations and Luminal A samples have more PIK3CA mutations



Mutual Exclusivity Statistics

OncoPrint | Mutual Exclusivity | Plots | Mutations | Protein Changes | Survival | Network

Gene	PIK3CA	TP53	GATA3
PIK3CA	---	0.048988	0.025178
TP53		---	0
GATA3			---

Indicates if alterations in the specified genes for the selected cases is significant for mutual exclusivity and/or co-occurrence.

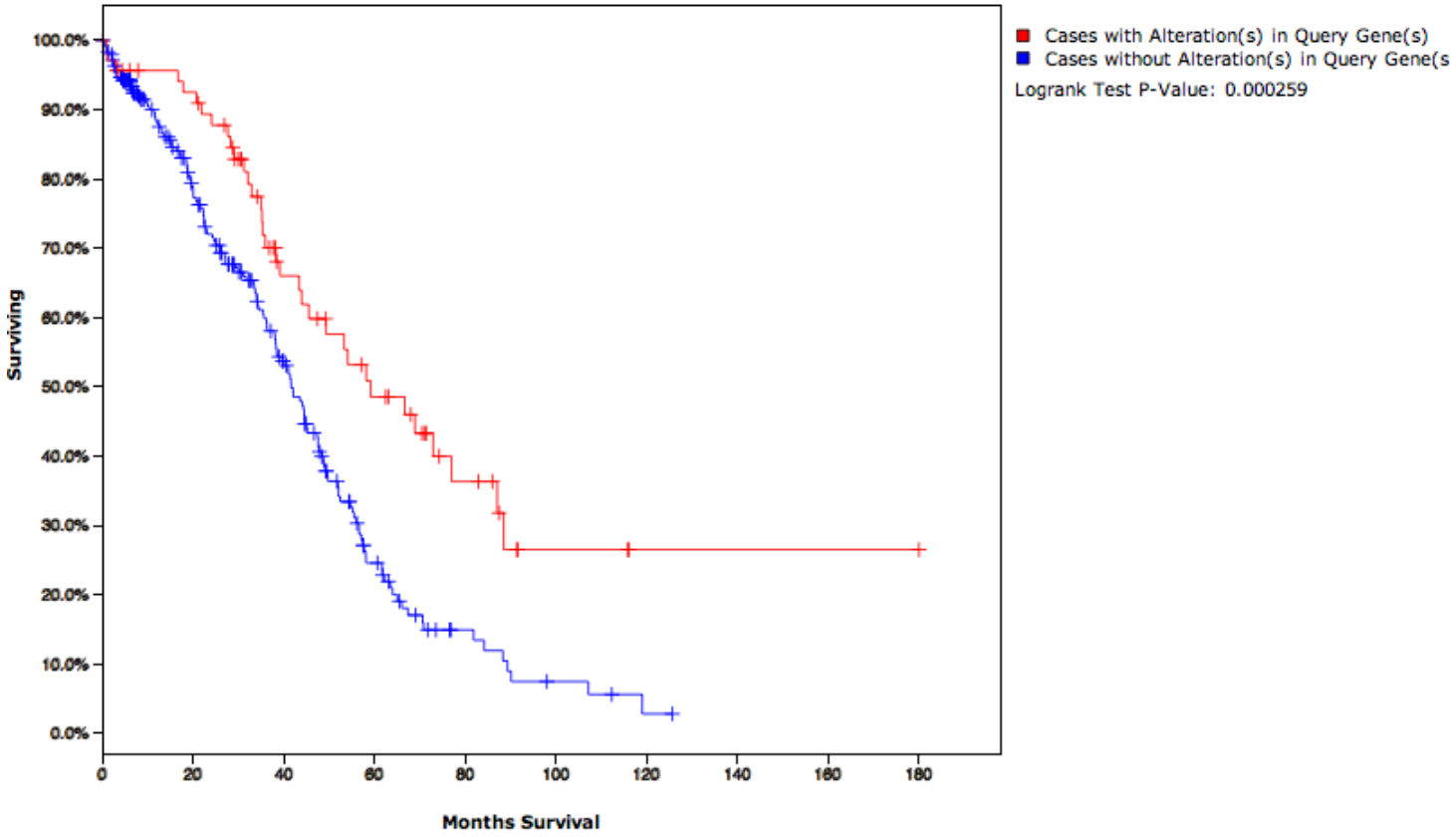
p-values <0.05, as derived via Fisher's Exact test are outlined in red.
p-values are *not* adjusted for FDR.

Legend
Strong tendency towards mutual exclusivity (0 < Odds Ratio < 0.1)
Some tendency towards mutual exclusivity (0.1 < Odds Ratio < 0.5)
No association (0.5 < Odds Ratio < 2)
Tendency toward co-occurrence (2 < Odds Ratio < 10)
Strong tendendency towards co-occurrence (Odds Ratio > 10)
No events recorded for one or both genes

Kaplan-Meier Survival Curve

Ovarian Serous Cystadenocarcinoma; Sequenced Tumors: (316); BRCA1 and BRCA2

Overall Survival Kaplan-Meier Estimate PDF SVG



	#total cases	#cases deceased	median months survival
Cases with Alteration(s) in Query Gene(s)	69	34	59.10
Cases without Alteration(s) in Query Gene(s)	246	147	41.56

Advanced cancer genomic data visualization: The Onco Query Language (OQL)

You can use the Onco Query Language (OQL) to select and define genetic alterations for all output on the cBioPortal for Cancer Genomics, including the OncoPrint, heat map, and data downloads.

Genetic Alterations

Users can define genetic alterations for three data types:

Data Type	Keyword	Categories and Levels	Default
Copy Number Alterations	CNA	AMP Amplified HOMDEL Homozygously Deleted GAIN Gained HETLOSS Heterozygously Deleted	AMP and HOMDEL
Mutations	MUT	MUT Show mutated cases MUT = x Specific mutations or mutation types.	All somatic, non-synonymous mutations
mRNA Expression	EXP	EXP < -x Under-expression is less than x SDs below the mean. EXP > x Over-expression is greater than x SDs above the mean. The comparison operators <= and >= also work.	At least 2 standard deviations (SD) from the mean.
Protein/phosphoprotein level (RPPA)	PROT	PROT < -x Protein-level under-expression is less than x SDs below the mean. PROT > x Protein-level over-expression is greater than x SDs above the mean. The comparison operators <= and >= also work.	At least 2 standard deviations (SD) from the mean.

Modifying earlier query…

Select Patient/Case Set: User-defined Case List Build Case Set

Enter case IDs below:

TCGA-AO-A0J6 TCGA-AO-A0JL TCGA-AO-A124 TCGA-AO-A129 TCGA-AO-A12F TCGA-AQ-A04J TCGA-AR-A0TP TCGA-AR-A0TS TCGA-AR-A0U1 TCGA-AR-A0U4 TCGA-AR-A1AH TCGA-AR-A1AI TCGA-AR-A1AJ TCGA-AR-A1AQ TCGA-AR-A1AR TCGA-AR-A1AY TCGA-B6-A0I2 TCGA-B6-A0IJ TCGA-B6-A0IQ TCGA-B6-A0RE TCGA-B6-A0RT TCGA-B6-A0RU TCGA-B6-A0WX TCGA-BH-A0AV TCGA-BH-A0B3 TCGA-BH-A0B9 TCGA-BH-A0BG TCGA-BH-A0BL TCGA-BH-A0BW TCGA-BH-A0DL TCGA-BH-A0E0 TCGA-BH-A0E6 TCGA-BH-A0RX TCGA-BH-A0WA TCGA-BH-A18K TCGA-BH-A18Q TCGA-BH-A18V TCGA-BH-A1F0 TCGA-C8-A12K TCGA-C8-A12V TCGA-C8-A131 TCGA-C8-A134 TCGA-D8-A142 TCGA-D8-A147 TCGA-E2-A14N TCGA-E2-A14R TCGA-E2-A14X TCGA-E2-A14Y TCGA-E2-A150 TCGA-E2-A158 TCGA-E2-A159 TCGA-E2-A1B5

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

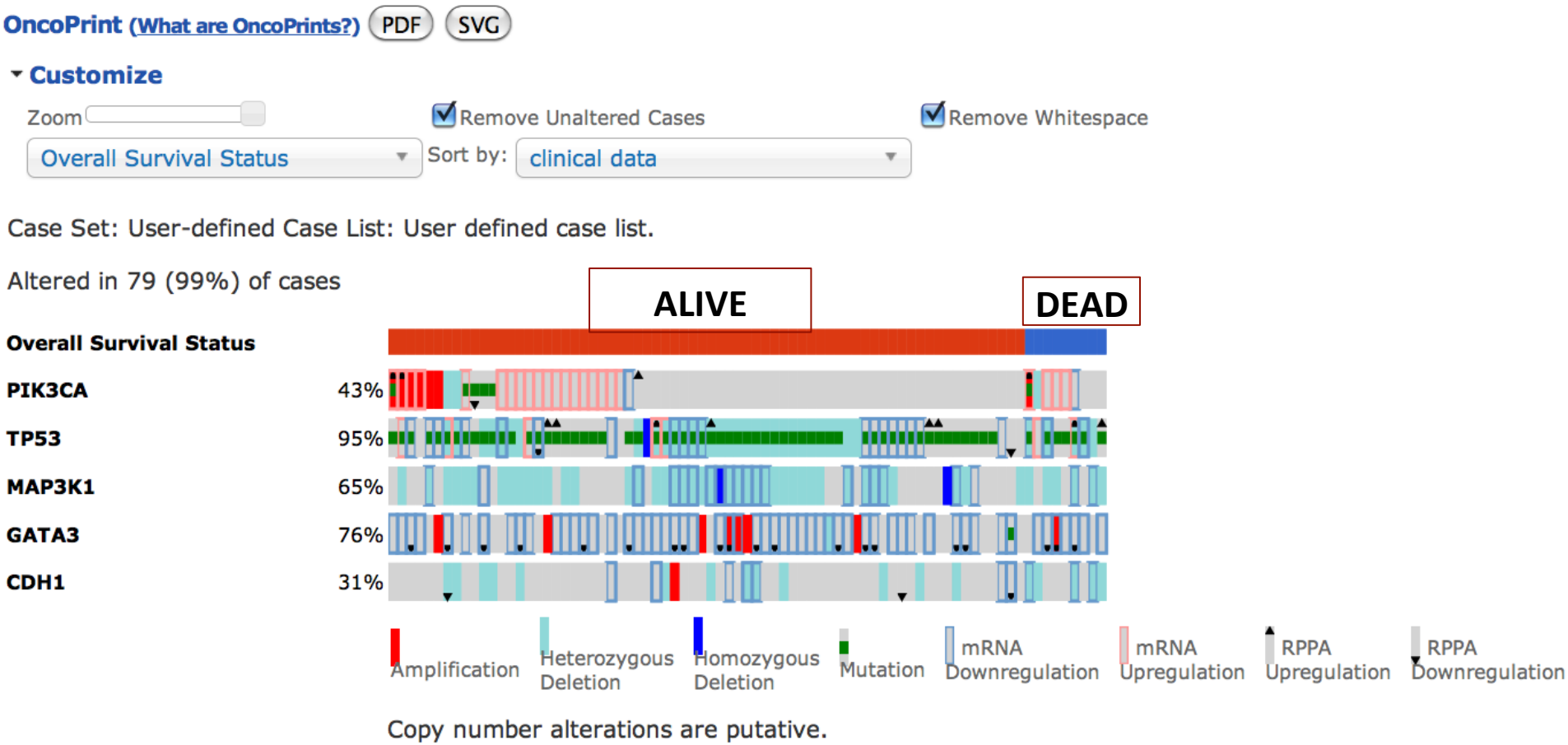
User-defined List

Select From Recurrently Mutated Genes (MutSig)

DATATYPES: AMP HOMDEL HETLOSS MUT EXP > 1.5 EXP<=-1.5 PROT >1.5 PROT <=-1.5; PIK3CA TP53 MAP3K1 GATA3 CDH1

All gene symbols are valid.

…displays specific information



No clear correlation between PIK3CA and TP53 alterations to Overall Survival status in this subset of 80 Basal patients (display Survival Curve on Live Demo)

References/Citations

- **Cerami et al.** The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery*. May 2012 2; 401.
- **Gao et al.** Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6, pl1 (2013).

Q&A

Thanks!