

Introduction to CBioPortal

Augustin Luna
11 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

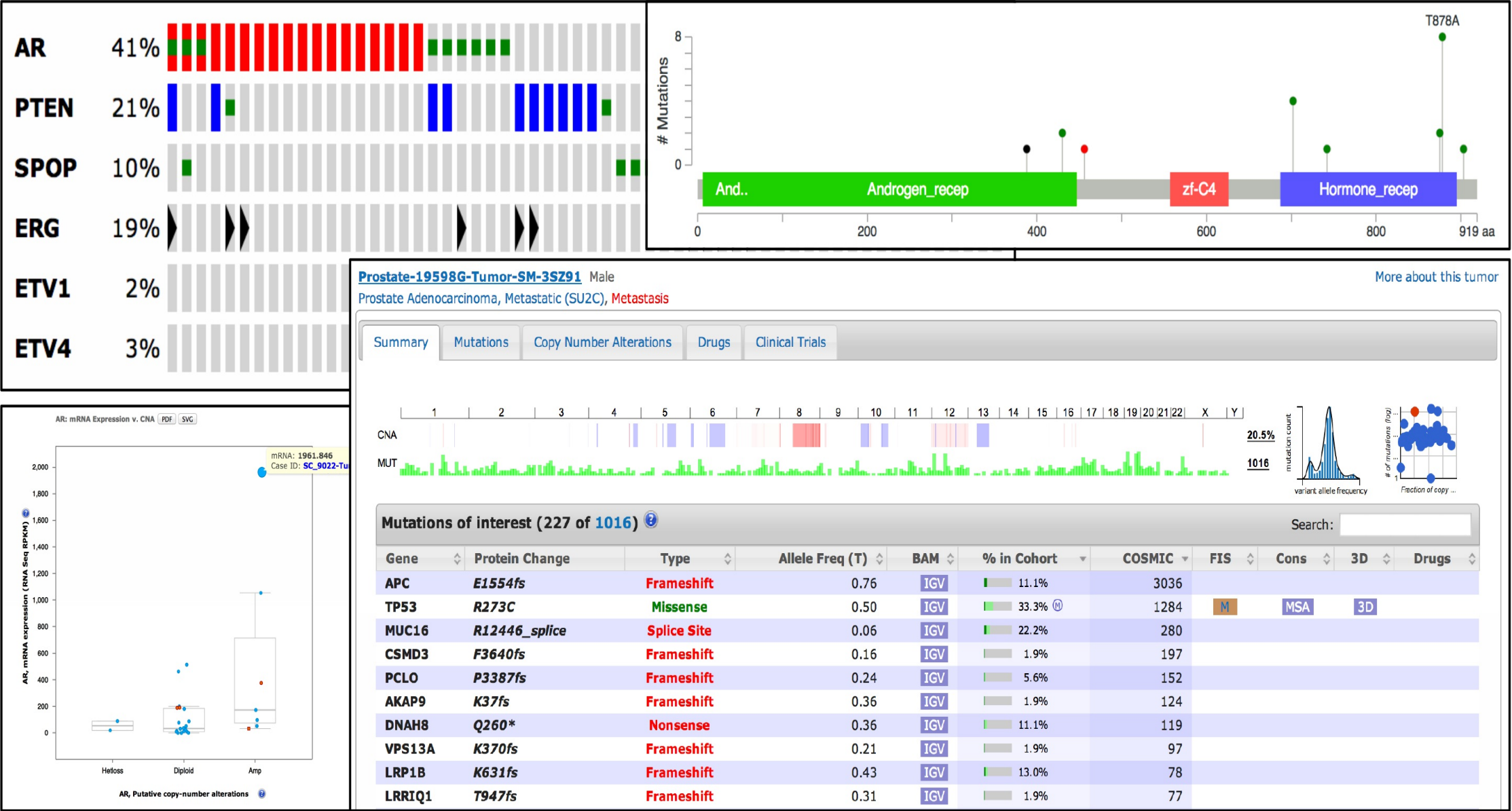
What is CBioPortal?

- 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 Website Features



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.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)

Getting Help

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