

Visualizing The Complexities Of Genomic Datasets

Vibha Raghavan
Hoffmann-La Roche



Disclosure information

No conflicts of interest to disclose

Overview

1. Background
 1. What is next-generation sequencing (NGS)?
 2. Why is visualization of genomic data important?
2. Maftools (R package)
3. ComplexHeatmaps (R package)
4. cBioPortal for Cancer Genomics (web browser)

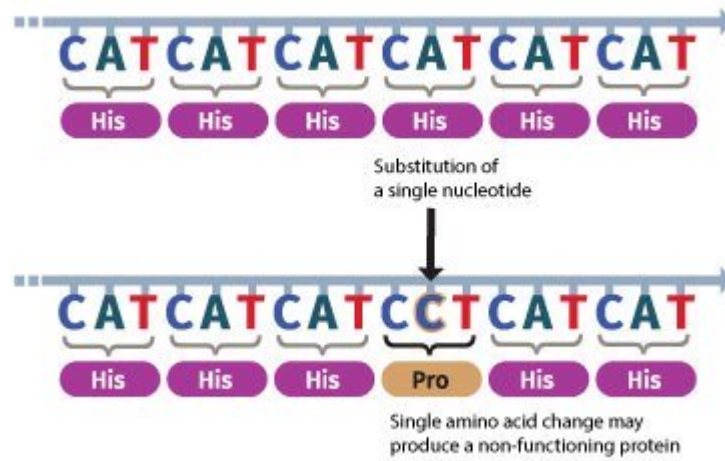
BACKGROUND

What is Next-Generation Sequencing (NGS)?

- Modern sequencing technology that allows sequencing of DNA and RNA
- Cheaper and quicker than previously used Sanger sequencing

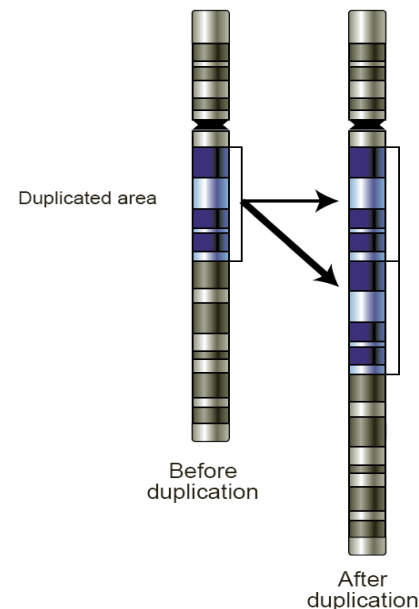
Single nucleotide variant

change in a nucleotide in the
DNA sequence



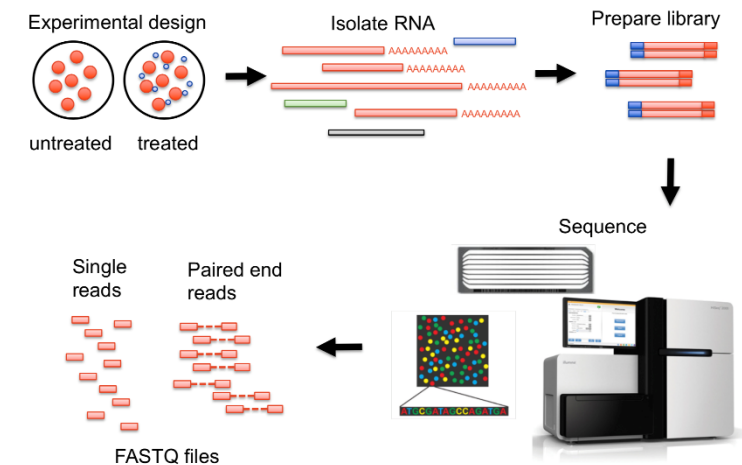
Copy number alteration

chromosomal structural change that
results in loss/gain of DNA



Transcriptome profile

Profile of RNA transcripts of a
biological specimen using RNA-seq



Why is visualization of genomics data important?

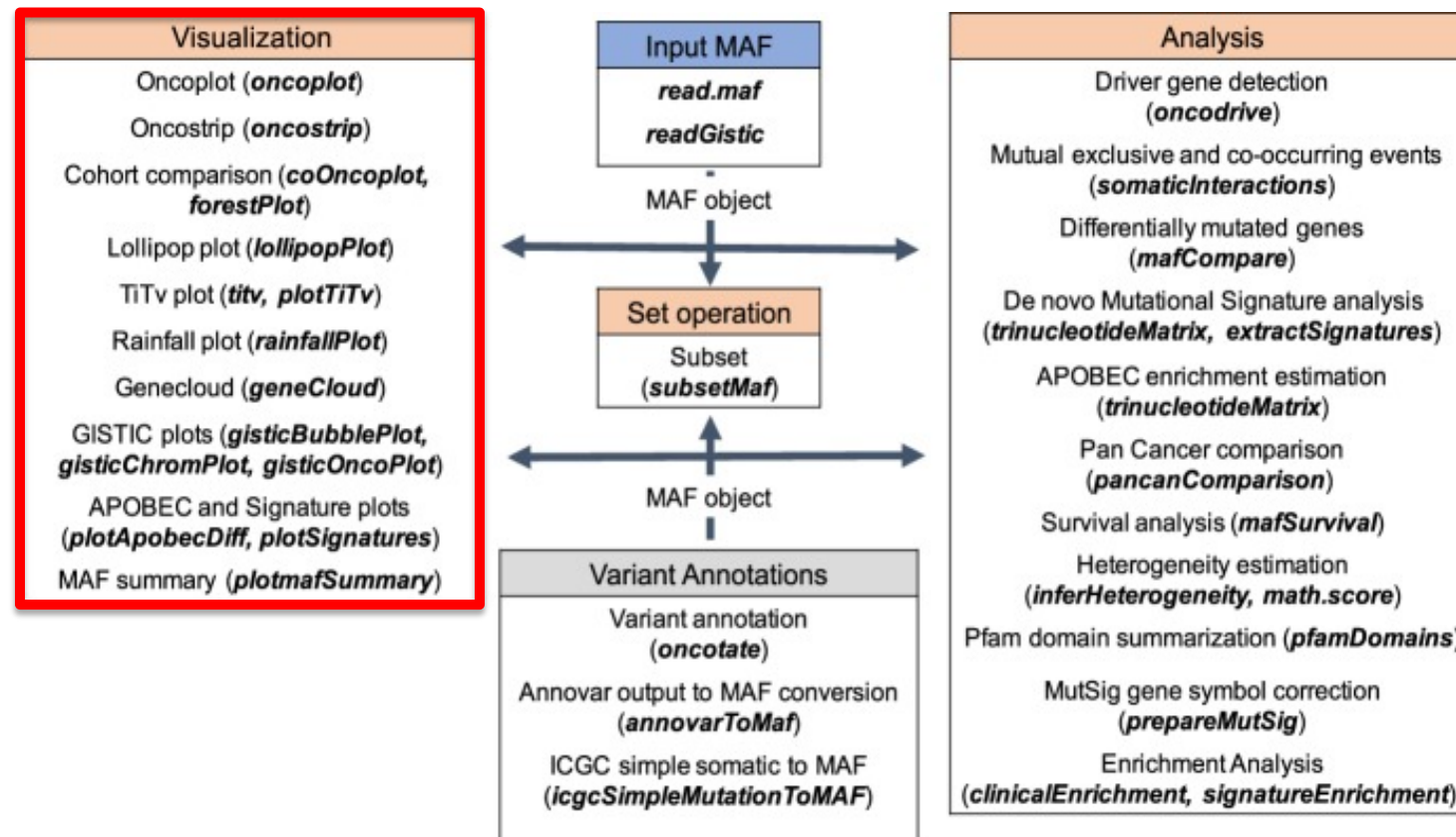
- Volume of data increasing drastically
- Mining and interpreting data in meaningful ways can be challenging
- Data visualization tools offer means to gain preliminary medical relevant insights from genomics datasets

Identifier	●	@SRR566546.970 HWUSI-EAS1673_11067_FC7070M:4:1:2299:1109 length=50
Sequence	●	TTGCCTGCCTATCATTTTAGTGCCTGTGAGGTGGAGATGTGAGGATCAGT
'+' sign	●	+
Quality scores	●	hhhhhhhhhhghhghhhhhfhhhhhfffffe'ee['X]b[d[ed'[Y[^Y
Identifier	●	@SRR566546.971 HWUSI-EAS1673_11067_FC7070M:4:1:2374:1108 length=50
Sequence	●	GATTTGTATGAAAGTATACAACATAAACTGCAGGTGGATCAGAGTAAGTC
'+' sign	●	+
Quality scores	●	hhhghfhhcghghggfcffdhfehhhhcehdchhdhahehffffde'bVd

Maftools (R package)

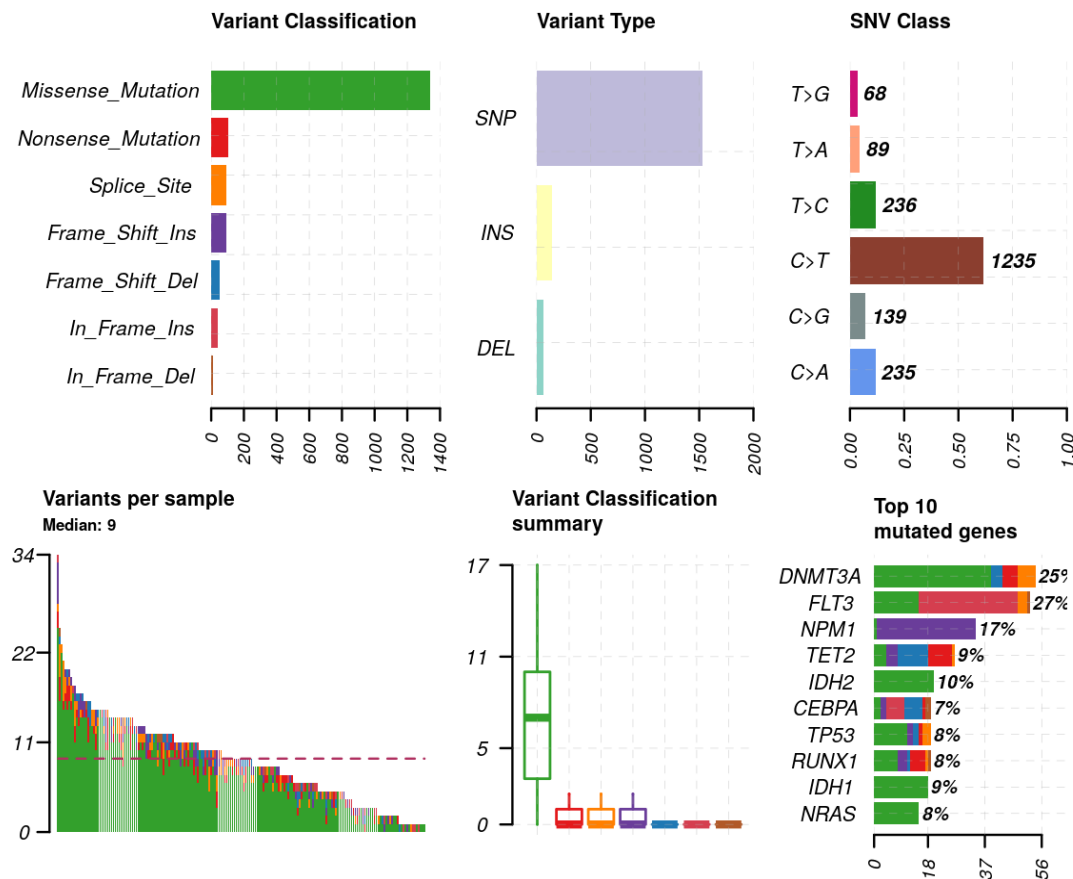
Overview of Maftools

- User-friendly R Bioconductor package
- Package includes tools for visualization and analysis of NGS data



Maftools Functions

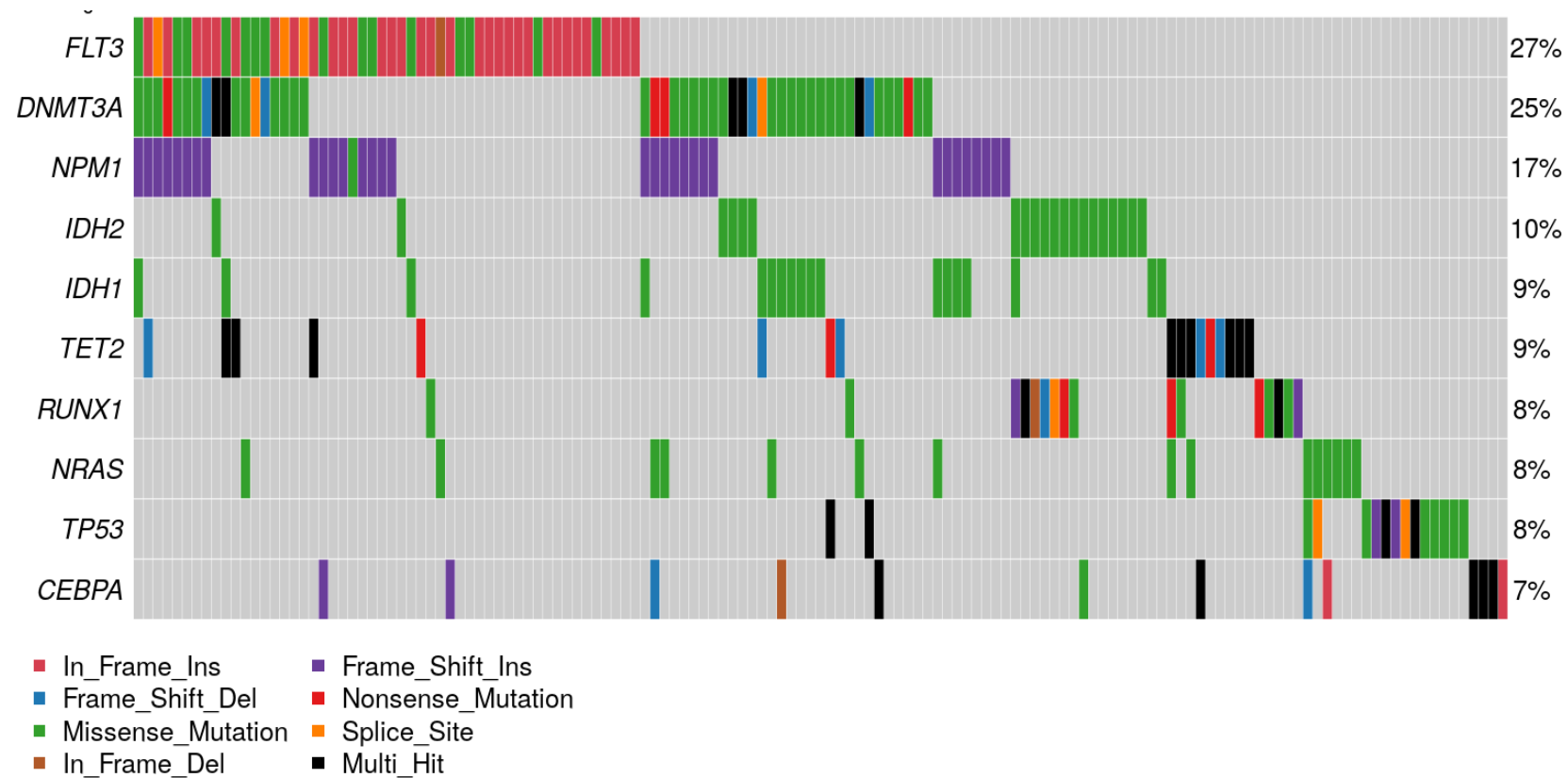
1) Summary visualization – *plotmafSummary* function provides a visual summary of the cohort



- Shows prevalence of variant classification/type
- Distribution of variants within sample
- Identifies top 10 mutated genes

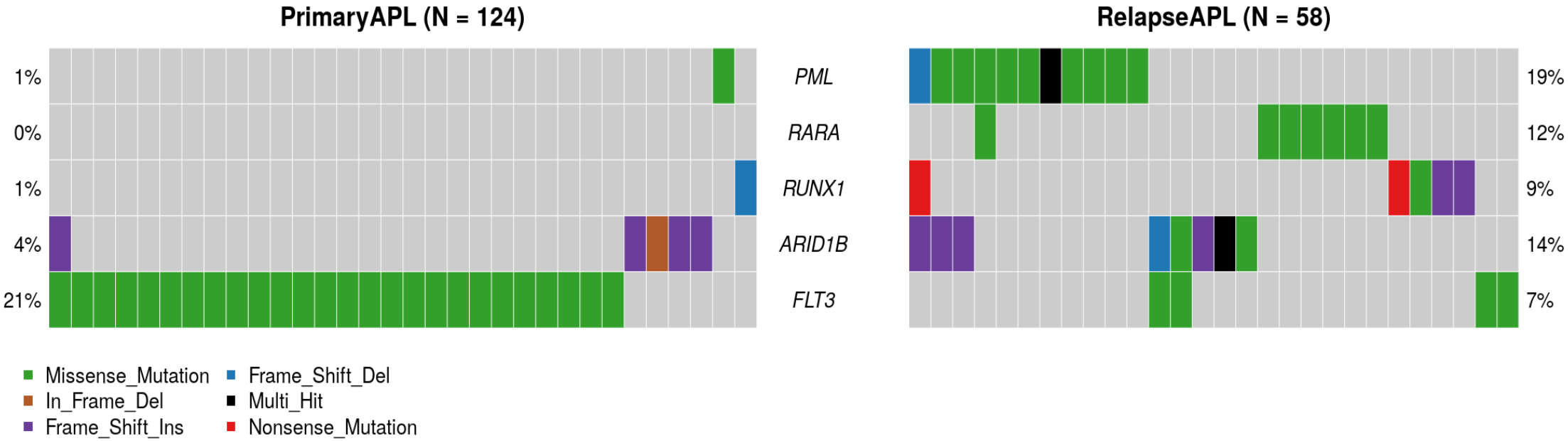
Maftools Functions

2) **Oncoplot** – *oncoplot* function used to visualize the mutational landscape of patients using a list of genes of interest



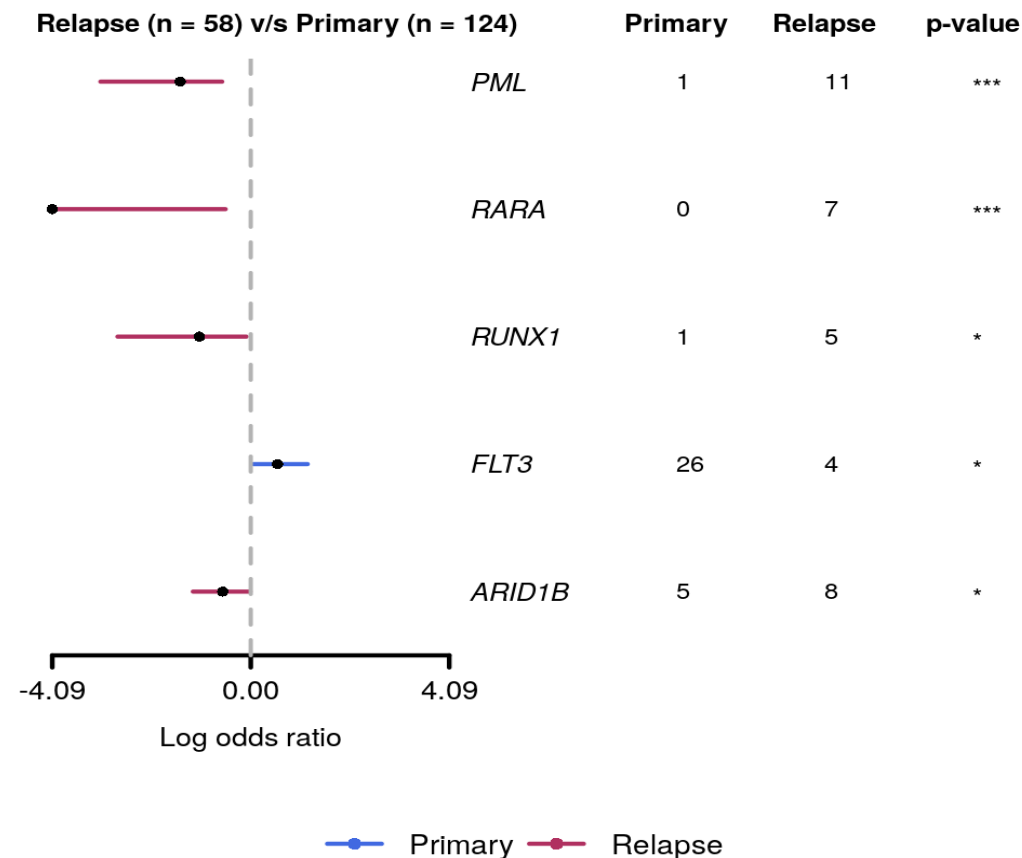
Maftools Functions

3) Co-onco plot – *coOncoplot* function used to compare two different oncoplots, eg. Primary vs. Relapse, Pre-treatment vs. Post-treatment, different arms of a study



Maftools Functions

4) Forest plot – *forestplot* function used to compare the frequency of mutations between two cohorts, while also taking into account the statistical significance of the results



ComplexHeatmaps (R package)

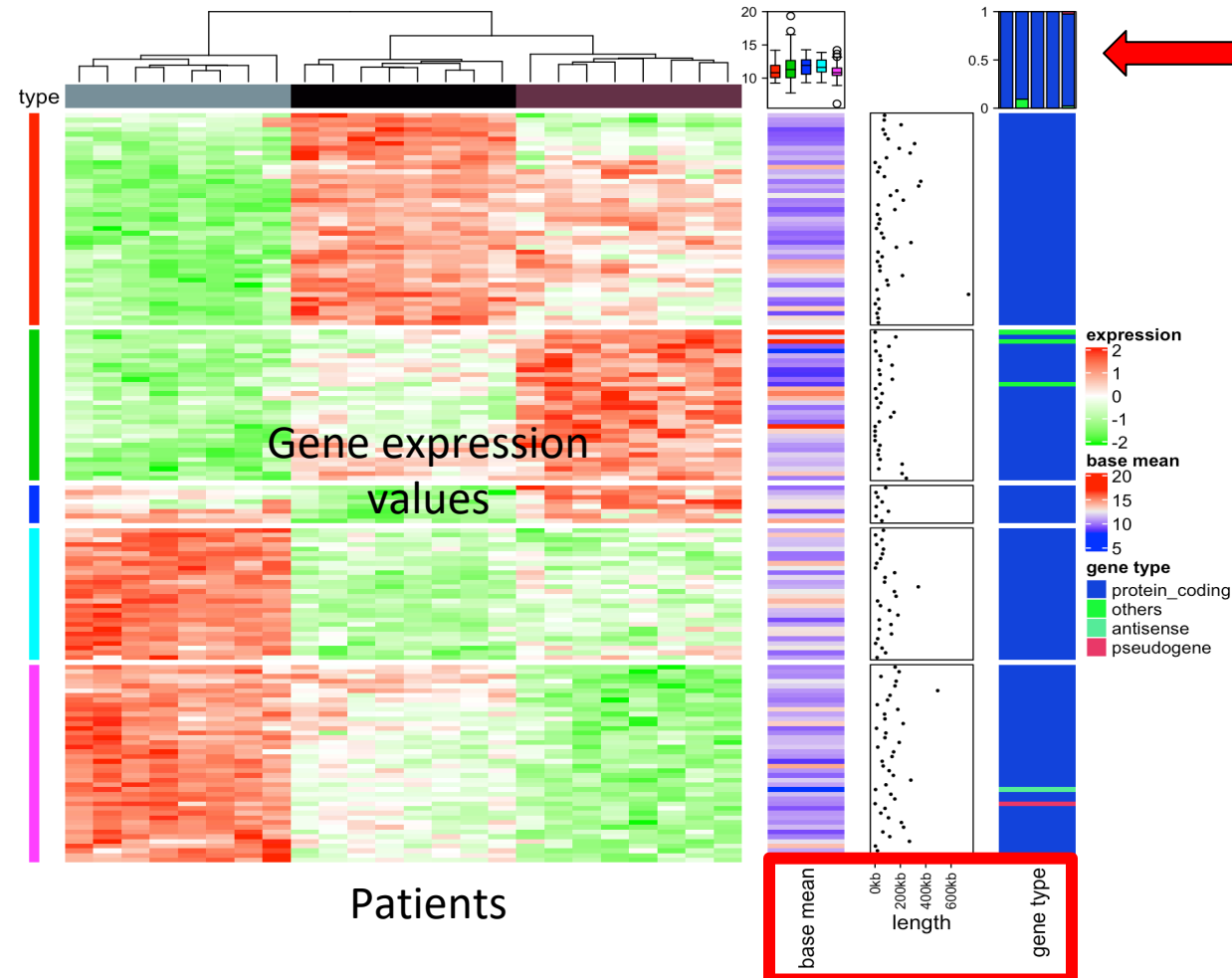
Overview of ComplexHeatmaps

- User-friendly R Bioconductor package
- Excellent tool to create multiple parallel heatmaps
- Can add multiple annotation graphics
- Examples of possible visualizations:
 - 1) Genomics alterations of entire cohort
 - 2) Expression profiles from RNA-seq data
 - 3) Relationships between methylation, gene expression, and gene-related information

ComplexHeatmaps Functions

1) Gene expression matrix – *Heatmap*, *HeatmapAnnotation*, *rowAnnotation* functions

Genes grouped into 5 clusters from k-mean clustering

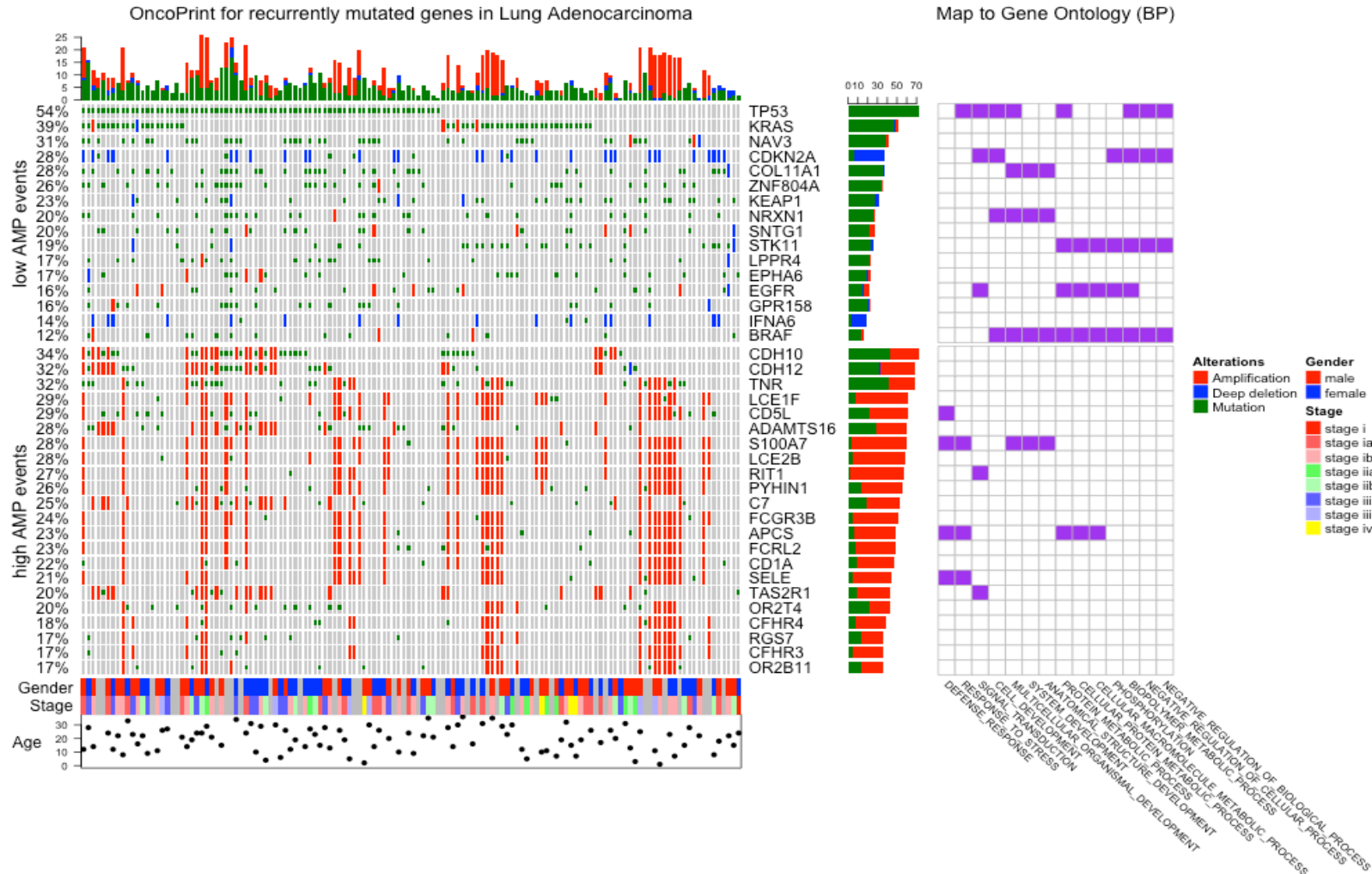


Distributions and statistics of data points in the 5 clusters

- Which cluster of genes to investigate?
- Which patients are clustered and why?
- Which genes have high or low base mean value across different patients?

ComplexHeatmaps Functions

2) Creating detailed oncoprints – *oncoPrint* function



- Which genes have highest frequency of alterations?
- Which pathways are affected by those genes?
- Are gender, stage or age enriched by certain alterations?
- Why are certain patients enriched in high AMP events?

**cBioPortal for Cancer Genomics (web
resource: *www.cbioportal.org*)**

Overview of cBioPortal

- Web resource for accessing, analyzing and viewing cancer genomics data from previously established studies
- Originally developed at Memorial Sloan Kettering Cancer Center (MSK)
- Examples of possible visualizations:
 - 1) Graphic summary of genomic alterations across patients
 - 2) Kaplan-Meier plots
 - 3) Visualization of networks that are altered in cancer
- Visualizations can be used to extrapolate information and compare to other studies

Getting started with a cBioPortal query

Query | **Download Data**

Select Cancer Study: Glioblastoma (TCGA, Nature 2008) 
The Cancer Genome Atlas (TCGA) Glioblastoma project. 206 primary glioblastoma samples.
Raw data via the TCGA Data Portal. TCGA, Nature 2008 [View details](#)

Select Genomic Profiles:

- ☒ Mutations 
- ☒ Copy Number data. Select one of the profiles below:
 - ☒ Putative copy-number alterations (RAE, 203 cases) 
 - ☐ Putative copy-number alterations (Consensus, GBM Pathways, 206 cases) 
- ☐ mRNA expression 

Select Patient/Case Set: Tumors with sequencing and aCGH data  [Build Case Set](#)

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

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

CDKN2A CDK4 RB1

Select from Example Gene Sets:
User-defined List 

[Submit](#)

1 Select a cancer study

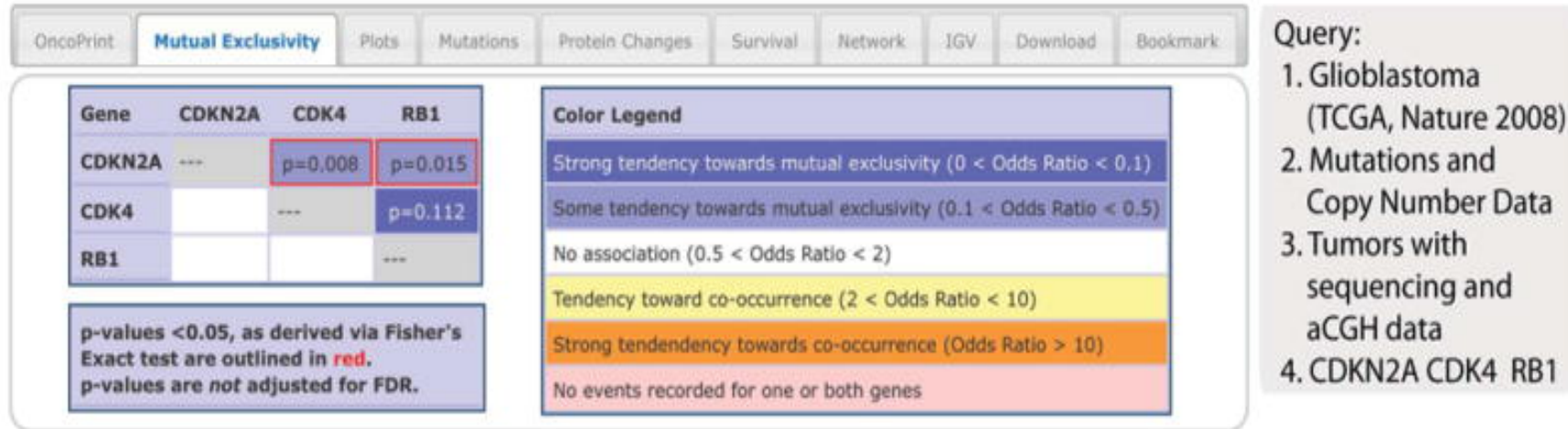
2 Select genomic profiles

3 Define a patient / case set

4 Enter genes of interest

cBioPortal Output Tabs

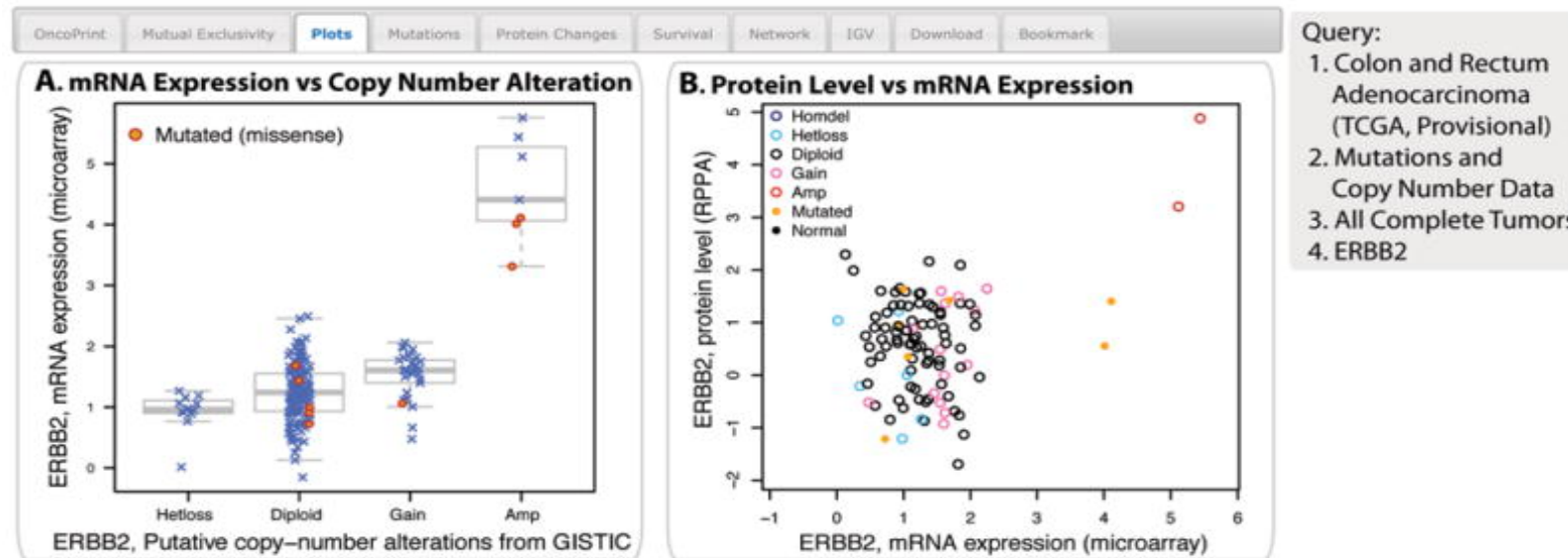
1) **Mutual Exclusivity** – provides information on mutual exclusivity of genes in the query population



- Genes that alter RB signal in glioblastoma have a tendency to be mutual exclusive
- Even though *CDK4* and *RB1* have the strongest tendency towards mutual exclusivity, it is not a statistically significant pair
- The other two pairs that have some tendency towards mutual exclusivity are statistically significant

cBioPortal Output Tabs

2) **Plots** - shows relationship between different genomic platforms in the query population

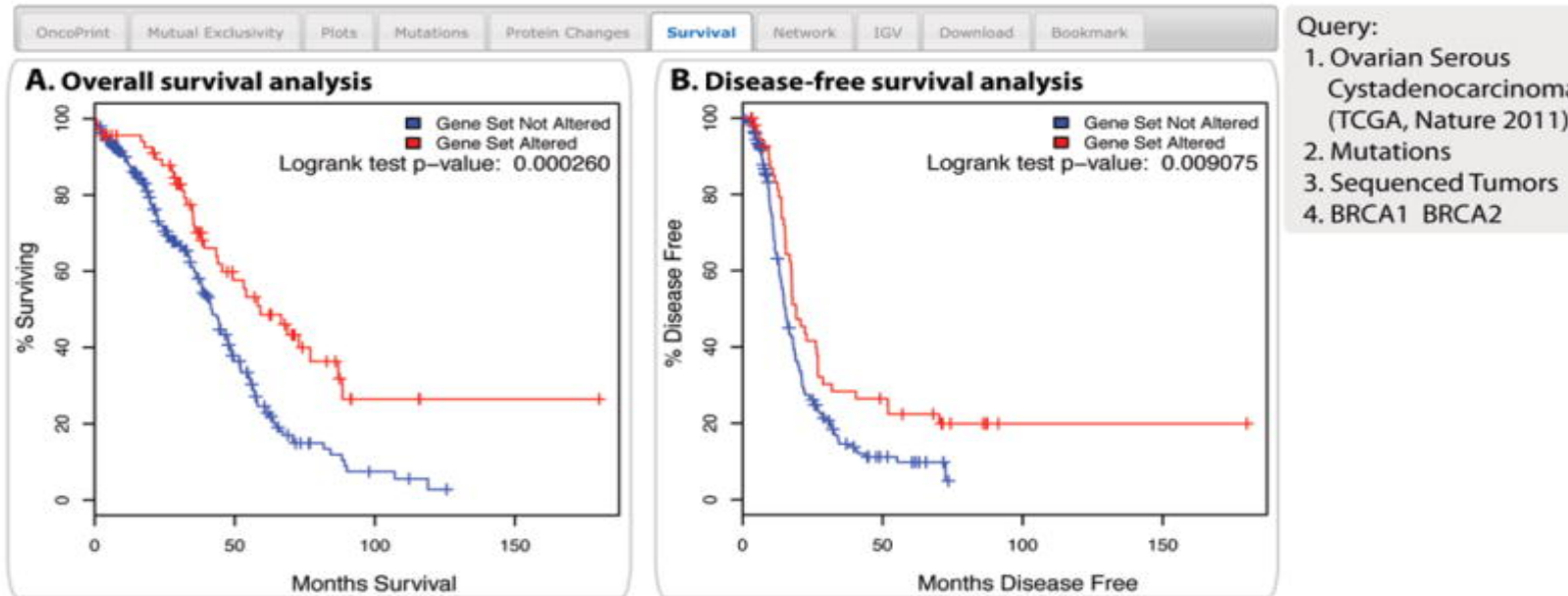


ERBB2 has high mRNA expression in patients who are also amplified for that gene

ERBB2 protein levels tend to be positively correlated with the gene expression levels

cBioPortal Output Tabs

3) **Survival** – plots Kaplan-Meier curves illustrating the results of survival analysis of the query population



Overall survival of
ovarian cancer patients
with and without *BRCA1*
or *BRCA2*

Disease-free survival of
ovarian cancer patients
with and without *BRCA1*
or *BRCA2*

Take-Home Messages

- Maftools, ComplexHeatmaps and cBioPortal provide excellent approaches to visualizing complex and multi-dimensional genomic datasets
- Used to gain preliminary insights that can be leveraged for more detailed downstream investigations
- User-friendly and easily accessible
- Helps us to continue to explore the world of genomics and change the future of therapeutics for the better



***Doing now what patients need
next***